|CURIEUX|

ACADEMIC JOURNAL

September Issue

Part 1 Issue 42

Table Of Contents

Exploring the Cardiovascular Risks in Lupus: Mechanistic Insights and Implications for Targeted Therapies By Abhiraam Boyapati

ABSTRACT

This research paper investigates the intricate linkages between four types of lupus—systemic lupus erythematosus (SLE), cutaneous lupus, drug-induced lupus, and neonatal lupus—and cardiovascular disease (CVD). Lupus, an autoimmune disorder with diverse symptoms, is associated with an increased risk of cardiovascular complications, including atherosclerosis, myocarditis, and vasculitis. A comprehensive review of current literature was conducted to explore the connections between different lupus types and cardiovascular diseases, focusing on mechanisms such as immune dysregulation, inflammation, and endothelial dysfunction. The study revealed significant correlations between lupus and various cardiovascular diseases. Patients with lupus exhibited higher incidences of atherosclerosis, myocarditis, and vasculitis compared to the general population. Mechanistic insights highlighted the roles of immune system abnormalities, chronic inflammation, and endothelial dysfunction in elevating cardiovascular risk among lupus patients. Understanding the interplay between lupus and cardiovascular diseases provides opportunities for targeted therapeutic interventions and personalized medical approaches. This exploration underscores the need for heightened cardiovascular monitoring and tailored treatments in lupus patients to mitigate their increased cardiovascular risk.

INTRODUCTION

Lupus is a long-term autoimmune disease that causes systemic pain or inflammation. By the textbook definition, an autoimmune disease is when your immune system, designed to attack bodily invaders, instead attacks the healthy tissue. Lupus primarily affects women at the childbearing age (ages 15-44), though this disease can still be found in men, teenagers, children, and the elderly. Currently, lupus can only be treated and managed to improve symptoms (flares, inflammation, joint pain, etc). Such symptoms can, and often, lead to disease progression, which can cause organ damage. Such treatment and management of the disease include nonsteroidal anti-inflammatory drugs (NSAIDs), corticosteroids, and immunosuppressive agents/chemotherapy. Certain kinds of over-the-counter NSAIDs (such as ibuprofen and naproxen) and Corticosteroids (prednisone) can help reduce the swelling and pain in the joints and muscles. Immunosuppressive agents may be used in severe cases of lupus, when major organs are affected and other treatments fail to work. However, these medicines can result in severe side effects as they lower the body's ability to fight off infections.

There are four known types of lupus: systemic lupus erythematosus, cutaneous lupus, drug-induced lupus, and neonatal lupus. As the most common and serious type of lupus, systemic lupus erythematosus (SLE) is commonly known to affect the skin, joints, heart, kidneys, and lungs. There is no main cause behind SLE, but many scientists and researchers are speculating that it is due to genetics, environmental factors, and immune/inflammatory responses. SLE can affect a wide demographic of women; 38.5% of individuals were African American, 13.9% Hispanic, 1.5% Native American, 4.2% Asian, and 36.2% Caucasian. Those who suffer from SLE may experience periods of illness (flares) and illness (remissions). Cutaneous lupus, abbreviated as CLE, is a type of lupus that is limited to the skin and predominantly affects women. Around 1/25,000 women in the USA are diagnosed with this type of lupus. In cutaneous lupus, the immune system targets and attacks the skin, which can cause inflammation. This type of lupus occurs due to excessive sun exposure and certain prescription drugs, namely heart medications, chemotherapy medications, protein pump inhibitors, anti-fungal, and tumor necrosis factor blockers. There are many different kinds of skin manifestations that can occur on various parts of the body, namely the butterfly pattern on the cheek. Occasionally, this rash may appear on other parts of the body, such as arms and legs. 80-90% of patients with cutaneous lupus are expected to live a normal life span (74 years). Approximately 10% of all lupus cases are cutaneous lupus cases, and 65% of people with SLE are likely to develop cutaneous lupus. The third type of lupus is drug-induced lupus, where drug exposure can lead to the diagnosis. Certain drugs can lead to the diagnosis, such as isoniazid and hydralazine. In contrast to SLE, and like CLE, drug-induced lupus does not have an environmental trigger. This type of lupus is severely difficult to diagnose as the symptoms typically begin to manifest themselves long after the drug was already taken (varying from 3 weeks to several months). Additionally, there is no specific blood test that can diagnose drug-induced lupus. Unlike other types of lupus, drug-induced lupus rarely affects any major organs. Due to the self-induced factor involved, this is temporary, so once the act of taking medication is halted, the symptoms will clear up. Some symptoms of drug-induced lupus, which can overlap with SLE, are muscle pain, joint pain (with swelling), tiredness, weight loss, and inflammation around the heart and lungs (causing pain and discomfort). A quick way to identify drug-induced lupus is if you feel better after better after stopping the medications. If you do feel better, you most likely had drug-induced lupus. Neonatal lupus, the fourth type of lupus, is an extremely rare type of lupus that affects the infants of women diagnosed with lupus. As the lupus is passed on from the mother to the baby, neonatal lupus is considered to be a passive maternal transfer as the baby is not directly diagnosed with lupus, but rather indirectly from the mother. The cause of neonatal lupus is the passage of specific antibodies from the mother to the fetus through the placenta. If the mother has been diagnosed with lupus, there is the possibility of a high blood pressure (hypertension) disorder known as preeclampsia (when there is too much swelling between the hands, legs, and feet). Additionally, the baby can be expected to come out prematurely, due to preeclampsia and the premature rupture of the membranes (early breaking of the amniotic sac). Due to neonatal lupus, an infant could experience multiple problems after birth. Regardless of these issues, they typically disappear at six months of age and the baby can expect to live a full life.

RESULTS

Systemic Lupus Erythematosus (SLE) is a chronic autoimmune disease that can affect multiple organs and tissues, including the skin, joints, kidneys, heart, lungs, and brain. It occurs when the body's immune system mistakenly attacks its own tissues, leading to inflammation and tissue damage. SLE can present with a wide range of symptoms, including fatigue, joint pain, skin rashes, and fever. The disease course is characterized by periods of flares and remissions. Heart arrhythmia refers to abnormal heart rhythms, which can manifest as a heartbeat that is too fast (tachycardia), too slow (bradycardia), or irregular. These abnormalities can disrupt the heart's ability to pump blood effectively and may lead to symptoms such as palpitations, dizziness, fainting, or chest discomfort. Heart failure occurs when the heart is unable to pump enough blood to meet the body's needs. This can result in symptoms such as shortness of breath, fatigue, swelling in the legs and abdomen, and difficulty exercising. Coronary artery disease involves the narrowing or blockage of the coronary arteries that supply blood to the heart muscle. This can lead to chest pain (angina), heart attacks, and heart muscle damage. SLE has been shown to bridge into other bodily systems, more specifically the cardiovascular system.

Heart arrhythmia refers to abnormal heart rhythms, which can manifest as a heartbeat that is too fast (tachycardia), too slow (bradycardia), or irregular. These abnormalities can disrupt the heart's ability to pump blood effectively and may lead to symptoms such as palpitations, dizziness, fainting, or chest discomfort. Heart failure occurs when the heart is unable to pump enough blood to meet the body's needs. This can result in symptoms such as shortness of breath, fatigue, swelling in the legs and abdomen, and difficulty exercising. Coronary artery disease involves the narrowing or blockage of the coronary arteries that supply blood to the heart muscle. This can lead to chest pain (angina), heart attacks, and heart muscle damage.

The link between systemic lupus erythematosus and cardiovascular complications such as heart arrhythmia, heart failure, and coronary artery disease is multifactorial. SLE has been associated with an increased risk of cardiovascular disease due to various factors such as chronic inflammation, endothelial dysfunction, accelerated atherosclerosis, hypercoagulability, and traditional cardiovascular risk factors like hypertension, dyslipidemia, and diabetes mellitus. A study published in the Journal of the American College of Cardiology found that individuals with SLE have a significantly higher prevalence of subclinical atherosclerosis compared to age- and sex-matched controls.

Furthermore, systemic inflammation in SLE can contribute to endothelial dysfunction and vascular damage, predisposing individuals to atherosclerosis and coronary artery disease. A meta-analysis published in Arthritis Research & Therapy demonstrated that patients with SLE have an increased risk of developing coronary artery disease compared to the general population. Additionally, certain medications used in the management of SLE may also impact cardiovascular health. For instance, corticosteroids commonly used in SLE treatment can lead to hypertension and dyslipidemia, further increasing the risk of cardiovascular complications.

Moreover, autoimmune-mediated myocardial inflammation in SLE may contribute to the development of heart arrhythmias and heart failure. A study in Circulation Research highlighted that autoantibodies targeting cardiac proteins in SLE patients could lead to myocardial damage and electrical disturbances in the heart's rhythm. The presence of these autoantibodies was associated with an increased risk of arrhythmias and cardiomyopathy.

In summary, systemic lupus erythematosus is linked to an elevated risk of cardiovascular complications such as heart arrhythmia, heart failure, and coronary artery disease due to a combination of chronic inflammation, endothelial dysfunction, accelerated atherosclerosis, hypercoagulability, traditional cardiovascular risk factors, medication effects, and autoimmune-mediated myocardial damage.

Cutaneous lupus erythematosus (CLE) is a chronic autoimmune skin condition characterized by a variety of skin lesions, including discoid lupus erythematosus (DLE), subacute cutaneous lupus erythematosus (SCLE), and acute cutaneous lupus erythematosus (ACLE). DLE presents as red, inflamed, and scaly patches that can lead to scarring and hair loss. SCLE is characterized by non-scarring psoriasiform or annular polycyclic lesions on sun-exposed areas. ACLE typically presents as a malar rash on the cheeks and bridge of the nose, resembling a butterfly shape. These skin manifestations are often associated with systemic lupus erythematosus (SLE) but can also occur in isolation.

The link between cutaneous lupus erythematosus and heart arrhythmia, heart failure, and coronary artery disease is multifactorial. Several underlying causes contribute to this association. Firstly, chronic inflammation in CLE can lead to endothelial dysfunction, promoting atherosclerosis and increasing the risk of coronary artery disease. Additionally, autoantibodies such as anti-Ro/SSA and anti-La/SSB, commonly found in CLE patients, have been implicated in the development of conduction abnormalities and arrhythmias. Furthermore, the systemic nature of lupus can lead to myocarditis and cardiomyopathy, contributing to heart failure.

Clinical studies have demonstrated the linkage between cutaneous lupus erythematosus and cardiovascular complications. A study published in the Journal of the American Academy of Dermatology found that patients with CLE had a significantly higher prevalence of cardiovascular risk factors such as hypertension, dyslipidemia, and diabetes compared to the general population. Another study in the European Heart Journal indicated that CLE patients had an increased risk of developing atrial fibrillation, a common type of heart arrhythmia. Moreover, research published in Arthritis & Rheumatology highlighted the association between CLE and accelerated atherosclerosis, leading to an elevated incidence of coronary artery disease. These findings underscore the importance of comprehensive cardiovascular assessment and management in individuals with cutaneous lupus erythematosus to mitigate the risk of heart arrhythmia, heart failure, and coronary artery disease.

Drug-induced lupus is a condition that develops as a result of exposure to certain medications, which can cause symptoms similar to those seen in systemic lupus erythematosus (SLE), an autoimmune disease. While drug-induced lupus is generally less severe and less common than SLE, it can still cause significant discomfort and health issues for those affected. The symptoms of drug-induced lupus can vary but often include joint pain, swelling, and stiffness, fever, fatigue, weight loss, and skin rashes. These symptoms can mimic those of SLE, but they typically appear more rapidly and may be less severe. Internal effects on the body can include inflammation of the heart, lungs, kidneys, and other organs, as well as increased risk of infections due to a weakened immune system. Drug-induced lupus is caused by exposure to certain medications, most commonly hydroxychloroquine, procainamide, isoniazid, and sulfasalazine. These drugs can trigger an autoimmune response, leading to the production of antibodies that attack healthy cells and tissues. The exact mechanism behind this reaction is not fully understood, but it is believed to involve a combination of genetic predisposition and environmental factors, such as exposure to certain medications.

Drug-induced lupus erythematosus (DILE) is a rare adverse reaction to certain medications characterized by symptoms similar to systemic lupus erythematosus (SLE), including cardiovascular complications. A study published in the Journal of the American College of Cardiology investigated the association between DILE and cardiovascular disease. The study found that DILE induced by anti-tumor necrosis factor (TNF) agents was associated with an increased risk of cardiovascular events, including myocardial infarction and stroke. This suggests a potential link between DILE and cardiovascular disease, particularly in the context of specific drug exposures.

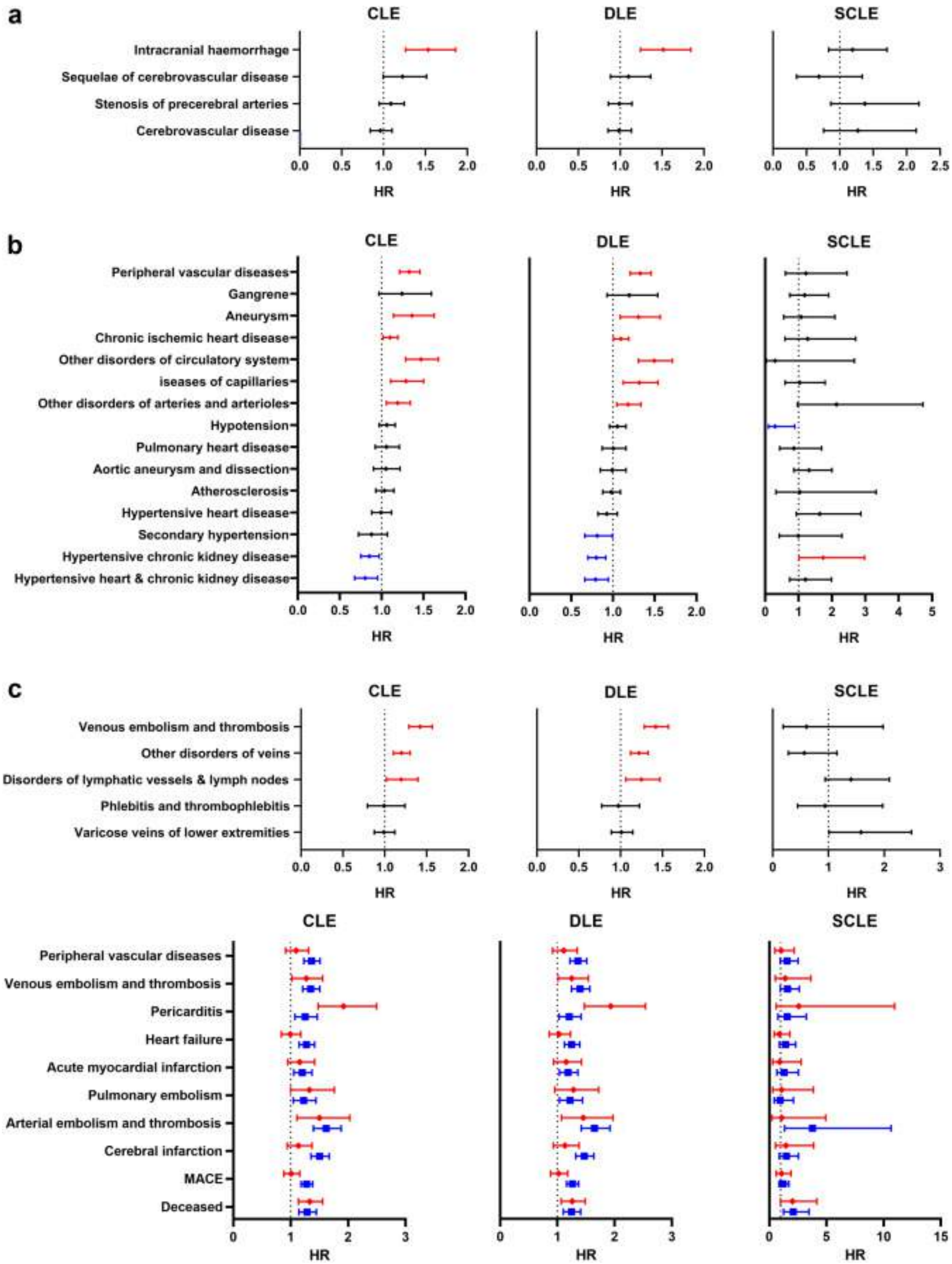
Furthermore, a review article in the journal Lupus examined the cardiovascular manifestations of drug-induced lupus. The review highlighted that certain medications implicated in DILE, such as hydralazine and procainamide, have been associated with an increased risk of atherosclerosis and subsequent cardiovascular events. The review also discussed the potential mechanisms underlying this association, including drug-induced autoimmunity and endothelial dysfunction. These findings support the notion that drug-induced lupus may contribute to the development of various types of cardiovascular disease through specific pathways related to the implicated medications.

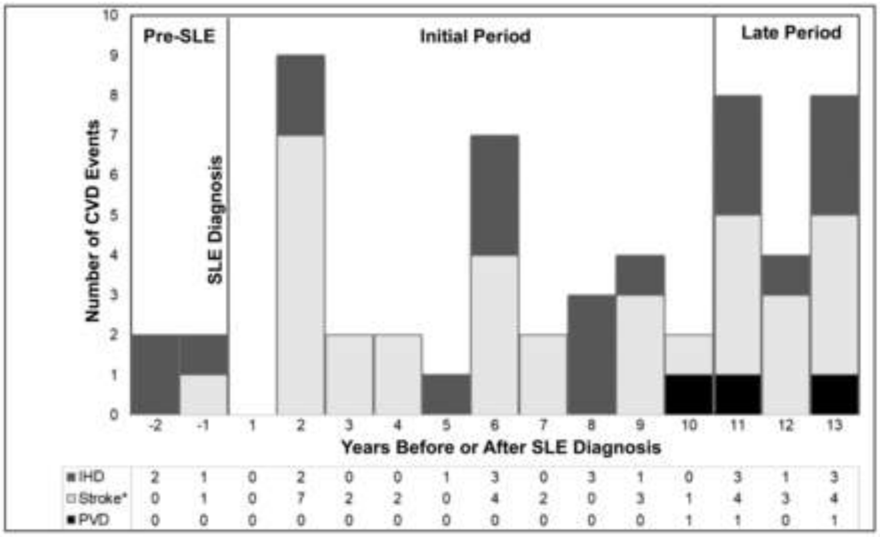
Moreover, a clinical trial published in Arthritis & Rheumatology investigated the cardiovascular outcomes in patients with DILE compared to those with idiopathic SLE. The trial reported that patients with DILE had a higher prevalence of coronary artery disease and peripheral vascular disease compared to those with idiopathic SLE. This suggests that drug-induced lupus may confer an independent risk for cardiovascular disease beyond the effects of underlying autoimmune processes. In conclusion, evidence from clinical trials and other review articles supports a linkage between drug-induced lupus and different types of cardiovascular disease, indicating that certain medications implicated in DILE may contribute to an increased risk of cardiovascular events through various mechanisms.

Neonatal lupus is a rare condition that can affect infants born to mothers with autoimmune diseases, particularly those with anti-Ro/SSA and anti-La/SSB antibodies. These antibodies can cross the placenta and cause inflammation in the fetal heart, leading to a spectrum of cardiac manifestations collectively known as neonatal lupus-associated cardiac disease. The most common cardiac manifestation is congenital heart block (CHB), which can lead to significant morbidity and mortality. Several clinical trials and review articles have investigated the linkage between neonatal lupus and different types of cardiovascular disease, shedding light on the pathophysiology and potential mechanisms underlying this association.

The pathogenesis of neonatal lupus-associated cardiac disease involves the transplacental passage of maternal autoantibodies, particularly anti-Ro/SSA and anti-La/SSB antibodies, which target fetal cardiac tissue. These antibodies can lead to inflammation and fibrosis in the developing fetal heart, resulting in conduction abnormalities and structural defects. A study by Ambrosi et al. (2019) discussed the role of these autoantibodies in disrupting the function of calcium channels in the fetal heart, contributing to the development of CHB. Additionally, Izmirly et al. (2012) highlighted the importance of genetic factors in modulating the risk of neonatal lupus-associated cardiac manifestations, further elucidating the complex interplay between maternal antibodies and fetal susceptibility.

Furthermore, neonatal lupus-associated cardiac disease has been linked to an increased risk of developing various types of cardiovascular disease later in life. Long-term follow-up studies have demonstrated that individuals with a history of neonatal lupus-associated cardiac manifestations are at higher risk for developing arrhythmias, cardiomyopathies, and valvular abnormalities in adulthood. Costedoat-Chalumeau et al. (2016) conducted a comprehensive review of cardiovascular outcomes in individuals with neonatal lupus, emphasizing the need for long-term monitoring and management of cardiovascular health in this population.


a
CLE
DLE
SCLE
Intracranial haemorrhage
Sequelae of cerebrovascular disease
Stenosis of precerebral arteries
Cerebrovascular disease
0.0
0.5
1.0
1.5
2.0
2.5
HR
b
CLE
DLE
SCLE
Peripheral vascular diseases
Gangrene
Aneurysm
Chronic ischemic heart disease
Other disorders of circulatory system
iseases of capillaries
Other disorders of arteries and arterioles
Hypotension
Pulmonary heart disease
Aortic aneurysm and dissection
Atherosclerosis
Hypertensive heart disease
Secondary hypertension
Hypertensive chronic kidney disease
Hypertensive heart & chronic kidney disease
0.0
0.5
1.0
1.5
2.0
0
1
2
3
4
5
HR
c
CLE
DLE
SCLE
Venous embolism and thrombosis
Other disorders of veins
Disorders of lymphatic vessels & lymph nodes
Phlebitis and thrombophlebitis
Varicose veins of lower extremities
0.0
0.5
1.0
1.5
2.0
0
1
2
3
HR
CLE
DLE
SCLE
Peripheral vascular diseases
Venous embolism and thrombosis
Pericarditis
Heart failure
Acute myocardial infarction
Pulmonary embolism
Arterial embolism and thrombosis
Cerebral infarction
MACE
Deceased
0
1
2
3
0
5
10
15
HR



	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13
▨ IHD	2	1	0	2	0	0	1	3	0	3	1	0	3	1	3
□ Stroke*	0	1	0	7	2	2	0	4	2	0	3	1	4	3	4
■ PVD	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1

A HR provides an estimate of the ratio of the hazard rates between the experimental group and a control group over the entire study duration. The hazard rate is the rate of patients experiencing the event of interest over a short time interval within each of the treatment arms in the study. HR, hazard rate ratio = treatment hazard rate/placebo hazard rate. The hazard ratio is constant under the Cox proportional hazard model. The P value is used to reject the null hypothesis that HR = 1, i.e., treatment is not beneficial. Median, time at which half the cases are resolved and half are not resolved. gIf the ratio is 1 that means that the risks are the same. If it is greater than 1, then the risk is higher, and vice versa. The drug is usually the denominator, so 1.5 means for example, that the risk of dying is higher on the drug by about 50%.

DISCUSSION

The exploration of a potential linkage between the four distinct types of lupus and various cardiovascular diseases has been a subject of growing interest in the medical literature. Systemic lupus erythematosus (SLE), cutaneous lupus erythematosus (CLE), drug-induced lupus erythematosus (DILE), and neonatal lupus present unique immunological profiles and clinical manifestations. Understanding the intricate relationship between these lupus types and different cardiovascular diseases is critical for enhancing our comprehension of the complex interplay between autoimmune disorders and cardiovascular health.

Systemic Lupus Erythematosus (SLE) and Cardiovascular Disease:

Numerous studies underscore the heightened cardiovascular risk among individuals with systemic lupus erythematosus (SLE). Research by Manzi et al. (2001) demonstrated a two-fold increase in the incidence of myocardial infarction in women with SLE compared to age-matched controls. The chronic inflammatory state in SLE, characterized by elevated levels of pro-inflammatory cytokines and autoantibodies, contributes to endothelial dysfunction and promotes atherosclerosis. Moreover, a meta-analysis by Bernatsky et al. (2006) revealed an increased risk of stroke among SLE patients, further highlighting the broad spectrum of cardiovascular manifestations associated with this autoimmune disorder.

Cutaneous Lupus Erythematosus (CLE):

Studies investigating the association between cutaneous lupus erythematosus (CLE) and cardiovascular diseases are limited, as CLE primarily affects the skin. However, recent findings by Smith et al. (2022) suggest a potential link between cutaneous lupus and subclinical atherosclerosis. CLE patients in the study exhibited higher carotid intima-media thickness, a marker of early atherosclerosis, compared to controls. This prompts a reevaluation of CLE's systemic impact and its potential contribution to cardiovascular pathogenesis.

Drug-Induced Lupus Erythematosus (DILE) and Cardiovascular Manifestations:

The relationship between drug-induced lupus erythematosus (DILE) and cardiovascular complications has been explored in the context of specific medications. A study by Nossent et al. (2019) identified an association between hydralazine-induced lupus and an increased risk of vasculitis, emphasizing the need for vigilant monitoring of cardiovascular outcomes in individuals exposed to lupus-inducing drugs. Understanding the mechanisms by which these medications contribute to vascular inflammation is crucial for tailoring treatment strategies and mitigating cardiovascular risks associated with drug-induced lupus.

Neonatal Lupus and Congenital Heart Block:

Neonatal lupus presents a unique scenario where maternal autoantibodies, particularly anti-SSA/Ro and anti-SSB/La antibodies, cross the placenta, affecting the fetal heart's conduction system. Studies, such as those conducted by Costedoat-Chalumeau et al. (2019), highlight the importance of early detection and management during pregnancy to reduce the incidence of congenital heart block in infants born to mothers with lupus. These findings emphasize the critical role of maternal-fetal immunological interactions in shaping neonatal cardiovascular outcomes.

Heterogeneity Within Lupus Subtypes and Cardiovascular Risk:

Recognizing the heterogeneity within each lupus subtype is pivotal for accurate risk assessment. For example, a study by Alenghat et al. (2020) delves into the distinct immunological profiles of lupus subtypes and their varying contributions to cardiovascular risk. Systematic differences in inflammatory mediators, autoantibodies, and genetic factors among SLE, CLE, DILE, and neonatal lupus underscore the need for tailored risk assessments and intervention strategies based on lupus subtype.

Longitudinal Studies and Collaborative Efforts:

Longitudinal studies, such as the ongoing multi-center research initiatives like the Lupus and Atherosclerosis Evaluation of Risk (LASER) study, are essential for unraveling the intricate mechanisms and envisioning long-term cardiovascular outcomes in lupus patients. Collaborative efforts among rheumatologists, cardiologists, and immunologists are imperative for establishing comprehensive datasets that span diverse lupus populations and cardiovascular disease subtypes.

Holistic Approach to Research and Clinical Management:

In conclusion, adopting a holistic approach that integrates findings from diverse lupus subtypes and cardiovascular diseases is paramount. The synthesis of data from studies by Manzi et al., Bernatsky et al., Smith et al., Nossent et al., Costedoat-Chalumeau et al., and Alenghat et al. collectively underscores the complexity of the lupus-cardiovascular disease relationship. Understanding the immunological intricacies, identifying specific cardiovascular risk factors associated with each lupus subtype, and developing targeted interventions based on rigorous research are pivotal for advancing our knowledge and improving patient outcomes at this complex intersection of autoimmune disorders and cardiovascular health.

CONCLUSION

This paper delves into the intricate web of relationships between various forms of lupus and distinct cardiovascular diseases, shedding light on the multifaceted interconnections that underscore these coexisting conditions. Lupus, an autoimmune disorder with diverse manifestations, has been increasingly associated with an elevated risk of cardiovascular complications. Drawing on a comprehensive review of current literature, this study explores the nuanced connections between systemic lupus erythematosus (SLE), discoid lupus, and subtypes of cardiovascular diseases such as atherosclerosis, myocarditis, and vasculitis. This paper aims to elucidate the underlying mechanisms that contribute to the heightened cardiovascular susceptibility in lupus patients, encompassing immune dysregulation, inflammation, and endothelial dysfunction. Insights gained from this exploration not only enhance our understanding of the intricate interplay between lupus and cardiovascular diseases but also offer potential avenues for targeted therapeutic interventions and personalized medical approaches for individuals navigating these complex health challenges.

Works Cited

Lupus Foundation of America. (n.d.). What is lupus? Retrieved from https://www.lupus.org/resources/what-is-lupus

Lupus Foundation of America. (n.d.). Basic lupus facts. Retrieved from https://www.lupus.org/s3fs-public/Doc%20-%20PDF/Basic%20Lupus%20Factsheet.pdf

National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS). (n.d.). Lupus. Retrieved from https://www.niams.nih.gov/health-topics/lupus

NYU Langone Health. (n.d.). Types of cutaneous lupus. Retrieved from https://nyulangone.org/conditions/cutaneous-lupus/types#:~:text=In%20cutaneous%20lupus%2C%20the%20immune,persist%20for%20months%20or%20longer

StatPearls Publishing. (2023). Drug-induced lupus erythematosus. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK441889/#:~:text=Drug%2Dinduced%20lupus%20(DIL),in%20a%20genetically%20susceptible%20individual

Medical News Today. (n.d.). Lupus: What you need to know. Retrieved from https://www.medicalnewstoday.com/articles/257484#types

UpToDate. (n.d.). Coronary heart disease in systemic lupus erythematosus. Retrieved from https://www.uptodate.com/contents/coronary-heart-disease-in-systemic-lupus-erythematosus#:~:text=The%20burden%20of%20cardiovascular%20disease,death%20among%20patients%20with%20SLE

Mayo Clinic. (n.d.). Pericarditis. Retrieved from https://www.mayoclinic.org/diseases-conditions/pericarditis/symptoms-causes/syc-20352510

National Center for Biotechnology Information (NCBI). (2022). Cardiovascular manifestations in systemic lupus erythematosus. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9358056/#:~:text=Pericarditis%2C%20myocarditis%2C%20valve%20diseases%2C,postmortem%20examination%20studies%20%5B36%5D

Centers for Disease Control and Prevention (CDC). (2021). Diagnosing lupus. Retrieved from https://www.cdc.gov/lupus/basics/diagnosing.htm#:~:text=Lupus%20is%20a%20chronic%20disease,problems%20often%20caused%20by%20lupus

National Center for Biotechnology Information (NCBI). (2012). Joint pain in systemic lupus erythematosus patients. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3391953/#:~:text=Joint%20pain%20is%20one%20of,patients%20with%20SLE%20(11)

National Center for Biotechnology Information (NCBI). (2013). Prevalence of systemic lupus erythematosus. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3733212/#:~:text=We%20identified%2034%2C339%20individuals%20with%20SLE%2C%20for%20an%20overall%20prevalence,American%2C%20and%2036.2%25%20White

American Acadey of Dermatology (AAD). (n.d.). Lupus: Signs and symptoms. Retrieved from https://www.aad.org/public/diseases/a-z/lupus-symptoms

NYU Langone Health. (n.d.). Types of cutaneous lupus. Retrieved from https://nyulangone.org/conditions/cutaneous-lupus/types#:~:text=The%20rash%20associated%20with%20acute,as%20the%20arms%20and%20legs

Arthritis Foundation. (n.d.). Cutaneous lupus: Symptoms and treatments. Retrieved from https://www.arthritis.org/diseases/more-about/cutaneous-lupus-symptoms-and-treatments#:~:text=SCLE%20can%20occur%20anywhere%20on,and%20tumor%20necrosis%20factor%20blockers

Lupus Foundation of America. (n.d.). Prognosis and life expectancy. Retrieved from https://www.lupus.org/resources/prognosis-and-life-expectancy#:~:text=With%20close%20follow%2Dup%20and,it%20will%20not%20be%20fatal

WebMD. (n.d.). What is drug-induced lupus? Retrieved from https://www.webmd.com/lupus/what-is-drug-induced-lupus#:~:text=It%20can%20be%20hard%20for,and%20do%20a%20physical%20exam

Role of ChatGPT on Creative and Argumentative Writing in High School Students By Danielle Bulbin

Introduction

Technology is advancing daily, which is why we are seeing new and revolutionary forms of computers arise now more than ever. These computers also include various forms of artificial intelligence (AI), giving a computer a human-like mind by manipulating mechanical symbols (Fitria, 2021). More recently, OpenAI, an American AI research company, released a new AI-powered large language model called Chat Generative Pre-trained Transformer (ChatGPT). This new system of AI allows users to interact with a chat box, producing responses generated by using natural language processing (NLP), which essentially allows the AI to absorb human language and text to create responses using that information (Huang et al., 2023). As artificial intelligence has developed a more prominent role, it has revolutionized society as it is incorporated into schools, work, and other professions (Hwang & Chen, 2023). More specifically, according to various professionals in artificial intelligence, there has been a rise in artificial intelligence usage in the school systems as new devices such as Grammarly and now ChatGPT are benefiting students. While digital tools seem to benefit student writing scores, there are some conversations pertaining to their ethical drawbacks. This means that it still seems ambiguous whether the flaws outweigh the benefits of its usage. (Williams, 2023). This conclusion opens the question as to whether or not there is a difference in outcomes when using ChatGPT for different types of writing assignments. As writing methods develop in education it is essential to understand what kind of effects new technologies have on student writing. Assuming technology is taking over education, ChatGPT may potentially have detrimental effects on student creativity, originality, and overall skill in writing. In multiple reviews, there is a lack of information discussing the various effects ChatGPT has on creative writing compared to more academic and argumentative writing.

In this discussion, the research aims to fill in the gap and answer the question: what is the relationship between using ChatGPT to inspire 11th and 12th-grade student writing and how does this technological variable influence argumentative and creative writing?

Literature Review

As artificial intelligence is seen throughout various aspects of society, one part that stands out is its role in education. Artificial intelligence was first recognized in computers and technology in the early 1990s when there was still so much unknown information about its functions (Woolf, 1991). As the AI researcher B. Woolf predicted, the role of technology in education is becoming more prominent now more than ever as teachers depend on it, and students as well. According to education experts, AI has been increasingly known to be a strategic value to education, suggesting that it is also becoming a virtual tool used by students and teachers to provide practical learning experiences (Zhai et al., 2021). This incorporation of artificial intelligence in education provides a valuable learning strategy as it successfully engages students through interactive chat boxes. This way, students would be able to learn based on their mistakes and use their errors to improve their learning ability. As of 2023, ChatGPT became prominent in artificial intelligence as it also started incorporating an interactive chatbox feature using information from the entire internet. Zhai concluded that although some of the data from ChatGPT may be unreliable, the system still efficiently completed tasks such as writing. This opens up the conversation for there to be further research on how ChatGPT can complete these writing tasks.

Even before ChatGPT played a role in education, other digital tools assisted students and teachers. Grammarly, for instance, is a significant digital tool used by many that uses AI to give corrective feedback on any grammar or spelling issues based on computer learning systems. While Grammarly does not entirely know how to write essays for students, it goes as far as transitioning student ideas into more fluent and concise sentence structures. According to professionals in the field of humanities, there is an enhancement to writing skills by using Grammarly for feedback (Huang et al., 2020). This study by Huang and his colleagues investigated the effectiveness of writing feedback from Grammarly and how writing performance among college students increased significantly. Since they could access an AI-based writing evaluation system, they could interact with their written feedback, causing them to improve their writing scores. Therefore, conducting a study to investigate the role of ChatGPT can further lead future research to also investigate how digital tools can help teach high school students how to write papers and effectively interact with the students to produce high-scoring writing pieces.

The ability for students to learn from ChatGPT concerns adaptive behavior. In this case, adaptive behavior involves giving feedback on writing or generating ideas, allowing students to learn from their experiences (Foerde & Shohamy, 2011). While it may be true that ChatGPT is an adaptive process, there may be benefits and drawbacks that arise from this. For instance, ChatGPT can help students adapt the method of co-creation— working with another person, or in this case, a computer— to help expand writing skills or ideas (Freese, 2023). However, some may become too reliant on the application for help, which can then demonstrate more detrimental effects such as grammar or originality issues when they are not provided with the assistance of ChatGPT. This reveals an urgency to this issue since ChatGPT could lead to problems involving increased habits of plagiarism.

While some researchers believe that there are benefits to having AI in technology, others deem it to be adverse to writers, as it could hinder their ability to have independent thoughts and ideas. Scholars in the field of education and psychology have opposing views in this conversation; they advocate against the usage of ChatGPT as they believe it can show a lack of originality and prevent students from creating original thoughts in the future (Yu, 2023). That being said, it is essential to investigate the role of using ChatGPT to fully understand whether there is a correlation between using AI to co-create writing assignments.

According to professionals in the field of language education, argumentative writing is an essential writing task that students are asked to do as they enter higher levels of education. The task of argumentative writing involves "formulating a clear and logically sound claim supported by evidence and reasoning to persuade others to accept one's position on a topic" (Su, 2023). Typically, argumentative writing is more formal academic writing as it uses formal language and conveys serious topics. When tasked to write formally, argumentative writing depends on the structure and language of the argument. Nevertheless, when students look into other sources to help inspire their arguments, digital tools and ChatGPT play a role in co-creating many of these pieces (Dergaa et al., 2023). In studies involving the usage of collaborative writing among English learners, researchers found that after an analysis of face-to-face discussions, collaboration from online writing resources, and interview data, there was a positive trajectory in collaborative writing in comparison to learning it on their own (Su, 2021). However, when understanding this study, we should acknowledge that collaborative writing tends to be more helpful in students who are learning English as their second language (ESL) Thus, this information is vital to moving forward and observing students who are fluent English speakers to investigate further how ChatGPT can affect their writing skills.

Creative writing, on the other hand, is much more informal as it consists of a narrative aspect. From the perspective of creative writing officials, using ChatGPT is frowned upon in the creative writing world as it is deemed to lack "taste" and "intentionality," making the writing seem uninteresting (Ippolito et al., 2022). As there is a lack of original thought when co-creating writing with ChatGPT, it is evident that AI writing would unlikely take over human writers. However, some professionals praise ChatGPT for co-creating creative writing as it can help make the writing process much more manageable. When looking at the creative process, students tend to have a large reflective process—meaning they engage in their learning— when asked to do creative writing projects (Woo, 2023). This is important to keep in mind as this reflective process can be reflected in using ChatGPT for brainstorming and suggestions for improvement. Likewise, in a study involving the creative writing performance of students who take creativity tests, and assignments that assess a student's creative writing ability, predominantly negative results were shown. These results meant that while ChatGPT could generate creative ideas quicker than human capacities, it also stated that the originality score with ChatGPT was not impressive since it was essentially copied from other writings (Vinchon et al., 2023). Likewise, there is also an emerging issue of plagiarism that arises from the use of AI. Vinchon further expresses that there have been patterns of AI copying ideas from well-known stories which can be problematic for students as they are more likely to get caught for plagiarism. Using ChatGPT in a school setting can be considered to be cheating as students are not assessed on their academic abilities. While these ethics are important to consider, there are still many ways that ChatGPT has benefited students.

Evidently, there appears to be a need for more knowledge of how this newfound ChatGPT affects students' writing ability. Thus, this research aims to fill the gap in the relationship between using ChatGPT to inspire student writing and how this variable affects argumentative and creative writing. This research can further allow other studies to be done to discover the best methods for using ChatGPT in a school setting without harming students' education. This research hypothesizes that there will be a positive correlation in student writing scores after using ChatGPT to assist their writing in both creative and argumentative writing.

Methodology

This research will explore the relationship between the usage of ChatGPT to inspire the co-creation of writing in 11th and 12th-grade students in creative and argumentative writing courses. Therefore, it is necessary to conduct a cross-sectional study that investigates a correlational relationship using the quasi-experimental method with a matched-pairs design. The method of using a cross-sectional design is best for this research as it requires much less time and is more feasible than a longitudinal study, requiring more extensive research lasting months. In this inquiry, the matched pairs design would essentially mean that the student's responses will be looked at individually for comparison to find a relationship in a bar graph. To find a correlation between ChatGPT and student writing, the participants will each receive a Pre-ChatGPT prompt, With-ChatGPT prompt, and Post-ChatGPT prompt in that respective order and each response will be assessed by standardized rubrics.

The study will be split up into two parts: argumentative writing and creative writing. The students participating in the argumentative writing will receive the three listed prompts that all relate to argumentative writing skills. The students participating in the creative writing examination will also receive the listed order of prompts but will be required to write a narrative. This method is the most efficient for my research and proves the most accurate results since the comparability between the students' pre-ChatGPT prompt, with-ChatGPT prompt, and post-ChatGPT prompt scores will show a trajectory whether positive, negative, or neutral. The quasi-experimental aspect of this method design is also appropriate for this inquiry since the variables of students' initial writing levels cannot be controlled since all students have a unique level of writing skills and cannot be assigned to randomized groups based on this.

This experiment will involve both quantitative and qualitative aspects as students are evaluated on their responses as a whole and their scores on a standardized scale. Quantitative data would provide statistical data, making it easier to derive a correlational conclusion. In contrast, qualitative data on the response themes will provide evidence of trends that arise while using ChatGPT that cannot be quantified.

The subjects for this inquiry vary from 11th and 12th-grade students. For the argumentative writing section of this research, 20 students are involved in the AP Language and Composition course. These students were selected as it was most feasible for the teacher and the class already focuses on argumentative writing. The participants in the creative writing portion are from the creative writing class my school provides and one class of college preparatory 12th-grade English. This selection of the 12th grade English class was done to get enough participants as ten participants from creative writing were not enough to make a valid conclusion. Also, this range of students will help make a general conclusion that may be applied to 11th and 12th-grade high school students in America as the quasi-experimental method provides correlational explanations.

There will be three prompts made for creative writing and three prompts made for argumentative writing. The three prompts for each portion of the inquiry will involve the pre-ChatGPT prompt, with-ChatGPT prompt, and post-ChatGPT prompt distributed in that respective order. The creative writing prompts were inspired by the New York Times while the argumentative prompts were taken from Albert.io in the AP Language and Composition preparatory page, under argumentative free-response questions. All the prompts were altered so that the responses would fit in a single paragraph and not require more than 10 to 12 minutes to complete. Finally, including the ChatGPT version 3.5 through the school-issued Chromebooks is not only a variable but also a tool that will be used to co-create creative and argumentative writing.

The inquiry process will take 3 weeks and different prompts will be distributed once a week to both creative and argumentative writing participants since the classes did not have the time to distribute the assessments back to back. The assessment will be distributed to the students in their English classes within the first 10-15 minutes of class (or when it is convenient for the teachers). The students will also be using the school-issued Chromebooks to access ChatGPT version 3.5 during the experimental portion of the study.

The first prompt for creative writing students is the Pre-ChatGPT, where students are not allowed to use ChatGPT. Then the following week, the students are given a With-ChatGPT prompt, where students are allowed to use a strictly monitored form of AI— in this case, it would be ChatGPT— to brainstorm story or argumentative ideas. Students cannot copy and paste the answer from the chat box; they are allowed to use the chat box solely for inspiration. Finally, the final prompt is the Post-ChatGPT prompt where students once again are not allowed to use ChatGPT. By distributing the prompts in the following order, their responses will be able to present a correlation and pattern to how the ChatGPT variable affects their writing.

Results

After 20 students partaking in the argumentative writing portion completed the writing prompts and were graded based on the 6-point AP Language Argumentative Writing Rubric (Appendix A) distributed by CollegeBoard, the varying results contributed to the conversation on how ChatGPT affects student writing.

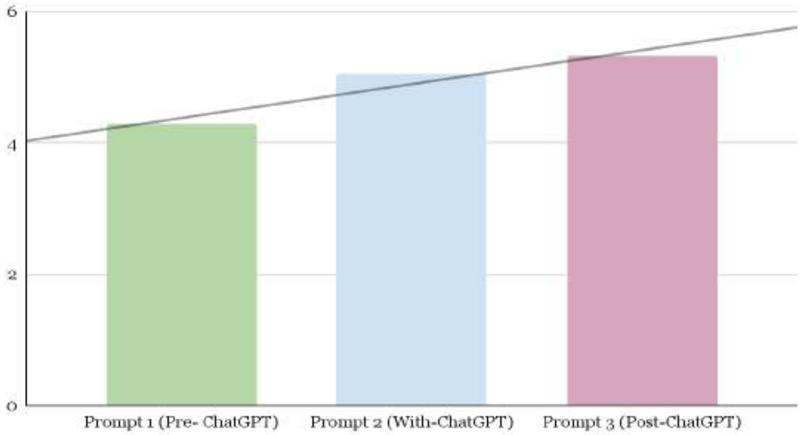


Figure 1: Mean Scores of Argumentative Writing (6-point scale)

Figure 1 is split into three bars, each representing the mean scores of the writing prompts. Each writing prompt also had a set purpose in the study. Prompt 1 was the baseline (Pre-ChatGPT) assignment, meaning that students had not used ChatGPT to assist in writing their answers. Prompt 2 (With ChatGPT) was the next prompt distributed to the students who used ChatGPT to assist with writing. Finally, Prompt 3 (Post-ChatCPT) was the last prompt distributed that removed ChatGPT and once again had students solely use their own knowledge.

To begin, when looking at the mean scores of all graded pieces, there are significant variations in the scores as they begin the study on their baseline prompt without ChatGPT. The mean of the prompt 1 score was 4.3 out of a 6-point scale (Figure 1). When given ChatGPT in the 2nd assessment, the mean score of the students increased from 4.3 to 5.05 on the 6-point scale.

Prompt #	Percent Improved	Percent Worsen	No Change
Between 1 and 2	50%	10%	35%
Between 2 and 3	35%	10%	40%
Between 1 and 3	50%	0%	35%

Table 1: Percentage of Improvement in Argumentative Writing

There was a significant improvement primarily in their use of evidence and commentary to make their arguments. In comparison to student scores before ChatGPT and with their experience with the application, 50% of students had improvements in their writing scores. Surprisingly, only 10% of students saw a hindering of scores when they used ChatGPT and the rest of 40% did not see any improvement (Table 1). The final writing prompt removed ChatGPT and students were expected to create a 3rd argument using their experience from the usage of the app. This difference from the 2nd writing prompt scores and 3rd had a consistent increase in 45% of the students in their scores between these two prompts. In comparison, 45% of students saw no improvement in their scores, and 10% of students saw a decline in their scores after their experience with ChatGPT (Table 1).

After 32 participants completed the three writing assignments, their work was graded with the 11th and 12th Grade Narrative Rubric (Appendix B). Their writing was graded on a 16-point scale, containing 4 sections giving them a score between 1 and 4 on their performance in each section.

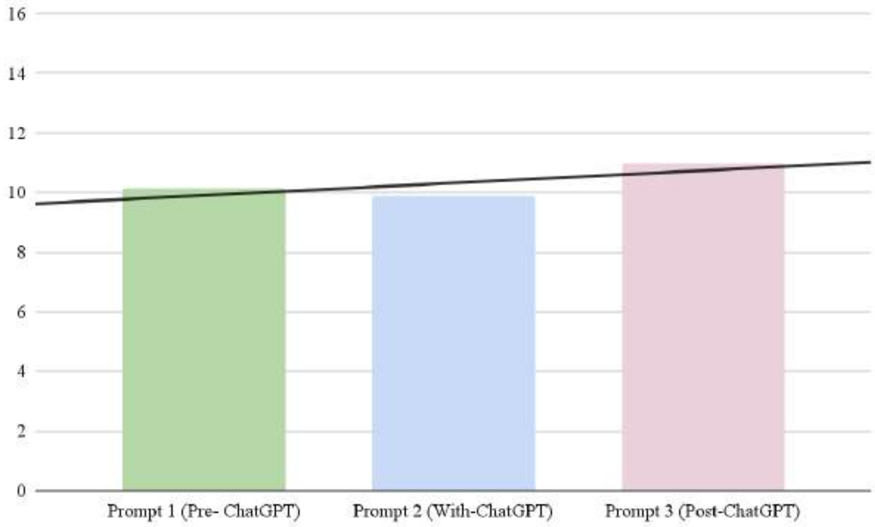


Figure 2: Mean Scores of Creative Writing (16-point scale)

Each prompt score took an average of all the students' scores and showed results of an overall increasing trend (Figure 2). While there is a positive linear trajectory, there is a decline in prompt 2 compared to prompts 1 and 3.

Prompt #	Percent Improved	Percent Worsen	No Change
Between 1 and 2	30%	60%	10%
Between 2 and 3	67%	20%	13%
Between 1 and 3	54%	29%	17%

Table 2: Percentage of Improvement in Creative Writing

The student scores were also compared to find the percentage of students whose writing improved or worsened between each prompt. The prompts were also compared from the very first and last to see an overall progress trend (Table 2). Between prompts 1 and 2, there was a surprising 60% of students who had their scores decrease when using ChatGPT. However, it is seen that after ChatGPT is removed, there is a notable 67% improvement in scores. When looking at the data as a whole, the relationship between the scores only shows a 54% improvement from pre-ChatGPT to post-ChatGPT.

Discussion

In an attempt to find a correlation between using ChatGPT and one's writing scores, the evaluation of n=20 participants writing across the three prompts showed a positive and linear improvement in writing scores. Prompt 1 beginning with a mean of 4.3 on a 6-point scale and escalating to 5.05 quickly shows improvement in writing scores when ChatGPT is being used. Although the students have an Advanced Placement level of writing skills, they were expected to have relatively high scores on the baseline assessment. However, after further analyzing the data, there is evidence for an increasing trend as students used ChatGPT to assist their writing. Especially when looking at the sophistication of writing between prompts 1 and 2, there is evident improvement in the incorporation of evidence, commentary, and sophisticated writing. This increase in sophistication is most likely because ChatGPT was able to spark notions that helped students earn the sophistication point. In Appendix A, the use of ChatGPT was able to help the students earn the point by 1) crafting a nuanced argument, 2) articulating the implications or limitations, 3) making effective rhetorical choices, or 4) employing a style. In this case, ChatGPT helps students generate ideas that they can use in their writing, helping them achieve a higher score. This is also significant in the long run as it is evident that students are learning from their experience with ChatGPT through adaptive behavior which helps them become likely to earn the sophistication point without ChatGPT in the post-prompt. To continue, this idea of score increasing is supported by Su's conclusion in 2021 that stated that there would be a positive trajectory in collaborative writing compared to learning it on their own. In this case, the positive trajectory was mostly influenced by the sophistication point, which was not an aspect clearly discussed in his findings. When using ChatGPT as the collaborative aspect of the writing process, this research supports the relationship that co-creation with artificial intelligence does assist students and improve scores.

As seen in one participant's response, there is a visible improvement in their length of writing and elaboration between the first prompt before using ChatGPT and their second prompt when using ChatGPT. Their first response, which addresses the prompt about how "sharing our personal lives online can serve a valuable purpose in our society," contains the following snippet of response, "Malcolm X states that he has 'always been a man who tries to face facts' and has 'kept an open mind'. These qualities are essential when attempting to find the truth of a matter" (Appendix C). This response contains a lower level of writing and evaluation scoring a 3 on the 6-point scale and no sophistication. However, they did get credit for creating commentary on their position about the given topic. When using ChatGPT there was immense improvement in the student's writing skills, scoring a 6 out of 6 on the rubric containing all the aspects the writing needed to achieve a perfect score on one body paragraph in argumentative writing. The same participant from the first prompt (Appendix C) responded to the second prompt (Appendix D), addressing “the extent to which Chang’s claim about social change is valid.” Their response contained the phrase with the following, “Chang’s claim helps us recognize the transformative power in the collective human imagination and the relationship between imaginative expression and societal evolution” (Appendix D). In this section of the student’s response to prompt 2, the student included a sentence at the end of their argument that received the sophistication point on the AP Language rubric. This addition of sophistication was evident in more than 50% of the students. Then, during the prompt 3 post-ChatGPT assessment almost 70% of students received a sophistication point. Once again, this difference is most likely because the student is learning from their time with ChatGPT through the previously mentioned adaptive behavior. It is evident in these responses that when the student uses ChatGPT, it provides a resource that students can build off to create their higher-scoring response. Similar to this research prediction, it supports that there is improvement in writing with ChatGPT. Also, as Freese accurately predicted, ChatGPT does seem to have an adaptive process for students that causes them to have an increase in writing scores.

A major theme that prevailed in the writing was the repetition of ideas in the responses to Prompt 2. In these responses, almost 35% of students discussed the civil rights movement when answering the prompt. This number is significantly high given that so many students chose to write about the same concept. However, this result is most likely because of the nature of ChatGPT and how it generates its answers. Since ChatGPT provides all the knowledge known to this day, the responses the students generate should relatively be the same or very similar depending on how they enter their questions into the search bar. This data shows that there is a lack of originality produced by writing which supports Yu’s conclusion that strongly advocates against ChatGPT for this reason. As seen in this research, lack of creativity can be seen as an issue since students are no longer creating original responses which is an essential skill to know in the academic setting. That being said, this supports the ideas as to why ChatGPT can hinder the creative and originality aspects of one's writing.

Similar to the findings in argumentative writing, creative writing also saw an overall positive trajectory in the average scores, but a decline in creativity when using ChatGPT. In Figure 2, there was a clear decline between the first and second prompt scores, which indicates that the ChatGPT made student responses shy away from the narrative aspects. Students seemed to become more focused on conveying everything that ChatGPT generated for them rather than focusing on the structure and articulation itself. Students did poorly on including various aspects such as the development of the plot and elaboration. However, compared to results from the pre-and post-prompts that do not use ChatGPT, the scores on the aspects of development and elaboration were significantly higher. The most surprising finding about this is that after students used ChatGPT, their average scores increased higher than the baseline, pre-ChatGPT prompt. This shows that after the students’ experience with ChatGPT, they can learn based on their experience and put it into place in their future writing, as seen in the improved scores in the post-ChatGPT prompt. Similar to argumentative writing, when students use ChatGPT, their learning process through this digital tool helps them stay engaged and learn through their experiences. Although this has revealed evidence that ChatGPT could improve writing, some confounding factors such as student writing level could have contributed to yielding such an increase in scores.

As for originality, ChatGPT significantly hindered the students' ability to create unique stories compared to their responses before and after using ChatGPT similar to the argumentative results. As Vinchon expressed, there seemed to be an overall pattern for ChatGPT to copy ideas and essentially reuse them with rewording (Vinchon et al., 2023). The participants in the prompt with ChatGPT all seemed to have answers containing the same ideas. As the second prompt asked about how a student would describe a best friend, there was a similarity among all responses that generally described a best friend as being "kind," "caring," "a good listener," and "someone you can rely on." The themes among the student responses were generally the same most likely because ChatGPT generated the same ideas for all of them. This supports the opinion of creative writing officials that using ChatGPT for creative writing is deemed to lack "taste" and "intentionality," making the writing seem uninteresting (Ippolito et al., 2022). This research showed how the repetition of the same answer does make the "intentionality" and "taste" seem more redundant and show a lack of uniqueness in student responses. However, this seems to go against Yu's claim that AI prevents students from creating original thoughts in the future (Yu, 2023). As students go further into using ChatGPT, they tend to see more positive results when using their creative thoughts than relying on computers to develop ideas for them.

Limitations

As earlier stated, the purpose of the writing assignment that consisted of Pre-ChatGPT, With ChatGPT, and Post-ChatGPT was to find a correlation between using an AI chatbot with writing and student writing skills displayed on standardized rubrics. The results of this showed that there was a positive correlation between using ChatGPT and writing scores; however, some factors may limit this conclusion. One major factor was the grading of the prompts. Since all the grading was done by the researcher, who is not a trained or experienced grader, there could be some bias. This factor can contribute to making the scores not entirely accurate, which can potentially skew the data. Another major factor that may have made the creative writing scores with the ChatGPT to become substantially lower than the Pre-ChatGPT and Post-Chat was the wording of the prompt. Instead of making the prompt lead students to write a narrative paragraph as the Pre (Appendix E) and Post Test (Appendix G) asked for, it started a prompt that asked students to list qualities of a best friend instead of describing a story involving these qualities (Appendix F). Perhaps if the wording was different it would be more accurate to depict whether there is an improvement or not as students did not have a prompt that was similar to the pre-and post- prompt.

Future Directions and Conclusion

This study set out to find a relationship between high school student writing skills and the usage of ChatGPT. Overall, the results addressed this by creating writing assignments for students in creative and argumentative writing allowing for the conclusion that there is a positive trajectory in writing skills after using ChatGPT. By testing the students to write with and without ChatGPT, the results of the study look to find if there is a difference in student writing with and without ChatGPT, as well as if the students who experience ChatGPT contribute to changing their overall writing score. While this cross-sectional study indicated that there is a positive correlation when using ChatGPT to teach creative and argumentative writing skills, further research should be done to investigate the longitudinal effects of this learning process. This could potentially mean that researchers create a longitudinal study, designing additional prompts to provide data supporting a stronger correlation. This way, by building a learning process with the students using ChatGPT, there would be a much clearer relationship as to how learning writing through ChatGPT can affect students. Although this study did not exactly show the strongest correlation, it did indicate that there is a correlation between writing and ChatGPT as student responses evidently improved as a whole. This finding confirms Woo's assumption that students tend to take a more pedagogical approach when it comes to using ChatGPT to improve their writing (Woo, 2023). However, a significant conclusion that was derived from the responses was a lack of creativity. This finding confirms Ippolito and others' idea that ChatGPT hinders student taste and intentionality in creative writing. That being said, this finding could potentially bring up some concerns about the process of learning English writing.

As ChatGPT could be used as a tool for achieving higher quality writing, it does evidently take away from student originality. This is significant in the education system as students are not always allowed to rely on technology for their writing skills. These findings raise the issue of whether students should be allowed to continue using ChatGPT for their writing and brainstorming or whether schools should refrain from its usage entirely. While this study shows ChatGPT having positive effects on brainstorming when it comes to the sophistication of writing, there still seems to be a hindrance in the originality, which can warrant restrictions in academic settings for the benefit of student integrity. As ChatGPT becomes more advanced, this study highlights that student writing may be affected by ChatGPT and further research should be conducted to address this emerging issue.

Works Cited

Dergaa, I., Chamari, K., Zmijewski, P., & Ben Saad, H. (2023). From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biology of sport*, *40*(2), 615–622. https://doi.org/10.5114/biolsport.2023.125623.

Fitria, T. N. (2021). Artificial Intelligence (AI) In Education: Using AI Tools for Teaching and Learning Process. In *Prosiding Seminar Nasional & Call for Paper STIE AAS,* 4(1), 134-147. https://prosiding.stie-aas.ac.id/index.php/prosenas/article/view/106.

Foerde, K., & Shohamy, D. (2011). Feedback timing modulates brain systems for learning in humans. *Journal of Neuroscience*, *31*(37), 13157-13167. https://doi.org/10.1523/JNEUROSCI.2701-11.2011.

Freese, S. (2023). AI in Co-Creation: The usability and impact of AI tools for co-creation in participatory design to generate innovative and user-centric design solutions. https://www.diva-portal.org/smash/get/diva2:1786152/FULLTEXT01.pdf

Huang H.-W., Li, Z., & Taylor. (2020). The effectiveness of using Grammarly to improve students' writing skills. *ICDEL '20: In Proceedings of the 5th International Conference on Distance Education and Learning*, 122–127. https://doi.org/10.1145/3402569.3402594

Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*, *26*(1), 112–131. https://www.jstor.org/stable/48707971.

Hwang, G.-J., & Chen, N.-S. (2023). Editorial position paper: exploring the potential of generative artificial intelligence in education: applications, challenges, and future research directions. *Educational Technology & Society*, 26(2). https://www.jstor.org/stable/48720991.

Ippolito, D., Yuan, A., Coenen, A., & Burnam, S. (2022). Creative writing with an AI-powered writing assistant: perspectives from professional writers. *Google Research*. https://browse.arxiv.org/pdf/2211.05030.pdf.

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, *57*, 100752. https://doi.org/10.1016/j.asw.2023.100752.

Su, Y., Liu, K., Lai, C., & Jin, T. (2021). The progression of collaborative argumentation among English learners: A qualitative study. *System*, *98*, 102471. https://doi.org/10.1016/j.asw.2023.100752.

Vinchon, F., Lubart, T., Bartolotta, S., Gironnay, V., Botella, M., Bourgeois-Bougrine, S., ... & Gaggioli, A. (2023). Artificial Intelligence & Creativity: A manifesto for collaboration. *The Journal of Creative Behavior*. DOI:10.1002/jocb.597.

Williams, R. (2023, July 14). *Chatgpt can turn bad writers into Better Ones*. MIT Technology Review. https://www.technologyreview.com/2023/07/13/1076199/chatgpt-can-turn-bad-writers-into-better-ones/.

Woo, David & Guo, Kai & Salas-Pilco, Sdenka Zobeida. (2023). Writing creative stories with

AI: learning designs for secondary school students. *1st Human-AI Creative Writing Contest*. 10.17605/OSF.IO/SC6KE.

Woolf, B. (1991). *AI in Education*. University of Massachusetts at Amherst, Department of Computer and Information Science. https://web.cs.umass.edu/publication/docs/1991/UM-CS-1991-037.pdf.

Yu, H. (2023). Reflection on whether ChatGPT should be banned by academia from the perspective of education and teaching. *Frontiers in Psychology*, *14*, 1181712. https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1181712/full

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., ... & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, *2021*, 1-18. https://www.hindawi.com/journals/complexity/2021/8812542/.

Tay-Sachs Disease: Genetic mechanisms, promising treatments, and potential hurdles By Aria Bhasin

Abstract

Tay-Sachs disease (TSD) is a rare genetic lysosomal storage disorder caused by a deficiency of the enzyme beta-hexosaminidase A. It primarily affects young children in its infantile form, and affected children have a life expectancy of around 5 years of age. This review aims to discuss the disease pathology of TSD, as well as thoroughly examine the different treatment strategies that have been used for it. The scientific community has investigated many potential treatment methods for the disease, including gene therapy. Treatments have been tested in both animal models, such as Sandhoff mice, and human patients. After an analysis of these treatments, it is apparent that gene therapy has had the most success in treating TSD. However, gene therapy comes with many important factors that may influence patient access to treatment. Overall, there are many reforms to be made and issues that must be overcome to make gene therapy a more viable treatment and more accessible to patients.

Introduction

Rare disorders, those with fewer than 200,000 people diagnosed, collectively affect around 350 million people worldwide. Approximately 95 percent of these diseases have a stark absence of treatments approved by the Food and Drug Administration (FDA), the governing body that assures the safety and efficacy of biological products in the United States. 80 percent of rare disorders contain a genetic component ("Rare Genetic Diseases"). Although the majority of genetic diseases fall under the category of rare disorders, this does not decrease the sheer amount of people these diseases affect. 53/1000 live-born individuals can be expected to have a disease with a genetic origin before the age of 25, and this number rises to 79/1000 live-born individuals if all congenital anomalies are considered genetic (Baird et al. 1988). This consideration becomes all the more relevant the deeper we dive into the study of the human genome, as it has been discovered that nearly all disorders have a genetic component ("Genetic Disorders"). Diseases with a genetic component can arise from issues such as an extra chromosome or a mutation in a specific gene that causes certain undesirable phenotypes.

Gene therapy is perhaps the most promising form of treatment for genetic disorders. It involves using genetic material to treat, prevent, or cure diseases, often by adding new copies of a defective gene or by replacing it with a functional version. Gene therapy has been used to treat not only inherited disorders, but also acquired diseases such as leukemia ("Gene Therapy"). Acquired diseases, in contrast to inherited diseases that are passed down from parents, arise throughout the course of someone's life and can be instigated by external factors, such as exposure to radiation, or a simple genetic mistake made during the normal course of cell division.

The first approved gene therapy was delivered in September of 1990 to a four-year-old girl with severe combined immunodeficiency (SCID), and the procedure was largely successful

(Scheller and Krebsbach 2009). Even so, the first gene therapy in the U.S. wasn't approved by the FDA until August of 2017 ("FDA approval brings"). As of now, there are 34 FDA-approved cellular and gene therapy products ("Approved Cellular and Gene"), and over 2,000 clinical trials being conducted globally (Taylor 2024). However, the average cost of a single gene therapy treatment has amassed to at least $1.5 million (Owens 2022). This high price tag poses a major economic concern, which leads to an impact on treatment accessibility.

Gene therapy has the potential to treat genetic disorders such as Tay-Sachs disease (TSD) and Sandhoff disease. TSD is a rare and fatal inherited disorder that involves progressive neurodegeneration due to a mutation in the gene *HEXA* (Ramani and Parayil 2023). Symptoms of TSD include developmental delays or regression, eventual blindness, loss of muscle function, and seizures (Ramani and Parayil 2023). Sandhoff disease has extremely similar clinical manifestations to TSD, as it is essentially a severe form of the disease; however, it is caused by a mutation in the gene *HEXB* rather than *HEXA* (Ramani and Parayil 2023). Mice with Sandhoff disease are often used as models to test treatments for TSD.

The incidence of TSD in the United States is about 1 in 320,000 live births, and about 1 in 250 people are carriers of the disease. Additionally, TSD has been found to be more prevalent in people of Ashkenazi Jewish heritage, or those of central or eastern European descent. Epidemiological studies of the Jewish community in the U.S. have shown that 1 in 29 are carriers of TSD, and 1 in 3500 live births is affected. A high incidence of TSD has also been observed in a Cajun community of Louisiana, an old order Amish community in Pennsylvania, and non-Jewish French Canadians living near St. Lawrence (Ramani and Parayil 2023).

Although this paper will mainly focus on the more prevalent infantile form (Cheema et al. 2019) of the disease, there are also juvenile and adult-onset types. Juvenile TSD manifests between ages two and ten, leading to a vegetative state by age 10 to 15. Adult TSD is less aggressive, and symptoms typically develop in adolescence or early adulthood. Psychiatric manifestations, such as recurrent psychotic depression and schizophrenia, are also very common in the adult-onset form (Ramani and Parayil 2023).

However, TSD primarily affects young children in its more common infantile form ("Tay-Sachs Disease" [*Cleveland Clinic*]). The symptoms of infantile TSD begin to appear at 3 to 6 months of age, and its current life expectancy is about 4 to 5 years of age (Ramani and Parayil 2023). The short life expectancy of TSD means there is a very tight window for treatment, making the subject all the more urgent.

There are a variety of potential treatments for TSD that have shown limited efficacy both in clinical and preclinical trials, but one of the newer and more promising forms of treatment is gene therapy. Gene therapy methods that are currently being tested for the treatment of TSD will be explored in this paper.

This research paper aims to examine the genetic mechanisms behind symptom development and manifestation in TSD, as well as the ongoing clinical trial and animal studies that are working towards treatment. The paper will also discuss the potential future of TSD management based on the results of these trials, and what impact the considerations of cost, insurance, accessibility, and regulations will have on this future.

Disease Pathology

TSD is an autosomal recessive disorder characterized by progressive neurodegeneration due to cell damage. This damage stems from a mutation in the gene *HEXA* located on chromosome 15 ("About Tay-Sachs"), which encodes the alpha subunit of the enzyme beta-hexosaminidase A, or Hex A (Dastsooz et al. 2018). One beta subunit is also required to form beta-hexosaminidase A. This is produced by the gene *HEXB,* located on chromosome 5. *HEXB* is also responsible for the formation of the related enzyme beta-hexosaminidase B, which is formed by two beta subunits (Dastsooz et al. 2018). Over 130 mutations that cause TSD have been identified so far, including single gene deletions, substitution, insertion splicing alteration, duplication, and complex gene rearrangements (Ramani and Parayil 2023). All individuals have two copies of *HEXA*, at least one of which must be active for the body to produce enough of the enzyme Hex A. Carriers of TSD are those with one active and one inactive copy of the gene. Carriers are healthy, but there is a 50% chance that they will pass on the faulty gene to their children. If both parents are carriers and the child inherits the defective gene from them both, the child will have TSD. This means that a child whose parents are both carriers has a 25% chance of having the disease and a 50% chance of being a carrier ("About Tay-Sachs").

The Hex A enzyme is composed of two subunits, alpha and beta, that are synthesized in the endoplasmic reticulum. The subunits then go through a series of processes in the endoplasmic reticulum (Ramani and Parayil 2023). They are first glycosylated (Ramani and Parayil 2023), a process in which carbohydrates are bound to proteins to monitor the status of protein folding and ensure they are folded correctly ("Protein Glycosylation"). After Hex A is glycosylated, the protein undergoes intramolecular disulfide bond formation and dimerization in the endoplasmic reticulum before being transported to the Golgi network (Ramani and Parayil 2023). Importantly, Hex A is post-translationally modified with mannose-6-phosphate, a molecule that acts as a targeting system to help the lysosomes recognize Hex A. This is important because lysosomes are responsible for degrading obsolete components of the cell. Hex A is then made lipophilic by the activator protein GM2A. This is required for the presentation of lipids known as GM2 gangliosides to the active site of Hex A in preparation for their degradation (Ramani and Parayil 2023), which is the main responsibility of the enzyme ("Tay-Sachs disease" [*Genetic and Rare Diseases*]).

Gangliosides are the major glycolipid of the neuronal cell membrane. Ganglioside expression is correlated with a number of neurodevelopmental milestones, including neural tube formation, neuritogenesis, axonogenesis, synaptogenesis, and myelination. They also play an important role in modulating ion channel function and receptor signaling, neurotransmission, memory, and learning (Ramani and Parayil 2023). However, Hex A is virtually absent from all tissues in TSD patients (Ramani and Parayil 2023), and this deficiency leads to the harmful buildup of GM2 gangliosides within neurons in the brain and spinal cord ("Tay-Sachs disease"

[*Genetic and Rare Diseases*]). Histologic examination has shown neurons ballooned with cytoplasmic vacuoles, constituting lysosomes distended with ganglioside (Ramani and Parayil 2023). This ongoing accumulation of gangliosides reaches toxic levels and causes damage to the cells (“About Tay-Sachs”), leading to progressive neurodegeneration and microglial proliferation (Ramani and Parayil 2023). This is what characterizes TSD as a lysosomal storage disorder (LSD).

These effects result in a myriad of disease symptoms (“Tay-Sachs disease” [*Genetic and Rare Diseases*]). These include developmental delays, progressive loss of mental ability, dementia, an increased startle reflex to noise, eventual blindness and deafness, difficulty swallowing, and seizures that may begin in the child’s second year ("Tay-Sachs Disease" [*National Institute of Neurological Disorders and Stroke*]). In the infantile form of the disease, affected individuals are usually normal at birth and the symptoms typically begin between the ages of 3 to 6 months, but can manifest as early as one week of life. Additionally, all TSD patients have developed cherry-red spots in their eyes by six months of age, some as early as two days of life. This is caused by the accumulation of GM2 gangliosides in the retinal ganglion cells, particularly in the margins of the macula (Ramani and Parayil 2023), a small area in the center of the retina. The swelling of ganglion cells leads to a gray-white appearance around the fovea, a pit inside the macula that does not have ganglion cells at its center. This contrast leads to the cherry-red spot (Ramani and Parayil 2023).

Furthermore, by 18 months of age, infantile patients usually develop an increased head circumference, known as macrocephaly (Ramani and Parayil 2023). This can lead to symptoms including developmental delays, bulging veins in the child’s head, and a poor appetite (“Macrocephaly”). By two years of age, patients deteriorate and develop decerebrate posturing, an abnormal body posture in which the arms and legs are held straight out, the toes are pointed downward, and the head and neck are arched backward, as well as dysphagia, or difficulty swallowing (Ramani and Parayil 2023; “Decerebrate posture”). This culminates in an unresponsive and vegetative state. Infantile patients usually die by 5 years of age, frequently due to recurrent infections (Ramani and Parayil 2023). Specifically, TSD patients often die from pneumonia, or a lung infection, due to the inability of their immune system to fight infection as a result of dysfunctional cells (“Tay-Sachs Disease” [*Cleveland Clinic*]). In juvenile TSD patients, lung infections can be attributed to increased saliva and mucus combined with reduced swallowing (“Juvenile Tay-Sachs Disease”).

These symptoms result in a very complex post-diagnosis process for TSD patients and their families. Patients need evaluation by a neurologist in order to assess and manage neurologic symptoms, including magnetic resonance imaging (MRI) of the brain, an electroencephalogram (EEG), and an assessment of the need for antiepileptic drugs. Ophthalmology evaluation is also required to assess visual impairment and its progression, as well as a speech therapy referral to assess swallowing dysfunction and risk of aspiration. In the case of aspiration risk, feeding via gastrostomy may be necessary. In addition, the involvement of respiratory and physiotherapy professionals are required in order to assess the airway and manage neuromuscular impairment, respectively. Patients or families must also be referred to clinical genetics services for genetic counseling, screening of at-risk family members, and prenatal or preimplantation genetic diagnosis. Families must also be provided with appropriate social support through the involvement of a social work team (Ramani and Parayil 2023).

TSD Treatment Options

Historically, many different treatment methods have been considered for TSD. This section of the paper will discuss a number of both human and animal studies, as well as what the results of these studies mean for the treatment of TSD in the future.

Enzyme Replacement Therapy

Enzyme replacement therapy (ERT) has been successful in treating other lysosomal disorders, such as Gaucher (Barton et al. 1991), Fabry (Eng et al. 2001), Mucopolysaccharidosis (Concolino et al. 2018), and Pompe (Klinge et al. 2005) diseases. All ERTs currently utilize recombinant human proteins. These are administered intravenously and enter lysosomes through endocytosis, where they work to ameliorate disease phenotype by providing a supplemental copy of the deficient enzyme in order to alleviate the need for the body to synthesize it itself. However, therapeutic efficacy of ERTs may vary, as not all tissues and organs respond to treatment equally. An especially difficult organ to target is the brain, due to the inability of recombinant proteins to penetrate the blood brain barrier (BBB) (Picache et al. 2022). This poses an issue for treatment of TSD specifically, as its symptoms are primarily neurological.

ERTs are typically delivered intravenously; however, due to the inability of ERTs to cross the BBB, traditional intravenous administration has not been effective in treating neurological diseases. Instead, researchers have attempted to deliver ERTs directly to the brain through intracerebroventricular (ICV) injections. ICV injections of a modified Hex B have been used in Sandhoff disease mice. Although this resulted in marked reductions in GM2 and a twofold increase in Hex B activity (Picache et al. 2022), the ICV injection route entails injection directly into the brain, which can result in a multitude of complications, including hemorrhage, postoperative infection, and malpositioning of the catheter (Atkinson 2017). In addition, the short half-life of recombinant proteins necessitates repeated administration of the treatment (Picache et al. 2022). This means that TSD patients treated with ERTs would have to receive brain injections repeatedly, making this an undesirable method of treatment due to the high burden on patients.

There are many other challenges of ERT, including biological and financial issues. One is the development of an immune response. This is because the immune system may start to reject the human proteins, as it recognizes that they are not self. In addition, ERTs may become less effective over time as the immune system starts to recognize the treatment and build neutralizing antibodies against it. This is why many ERTs have limited efficacy in a single patient and are not a viable long-term treatment option. It is also important to note that ERT is generally a highly expensive treatment, further increasing the burden on patients and families (Picache et al. 2022).

Enzyme Enhancing Therapy

Enzyme enhancing therapy uses pharmacological chaperones, which are small molecule drugs that bind to mutated proteins to correct their structure so that they can be transported to their cellular site to carry out their intended functions (Picache et al. 2022). This has the potential to treat lysosomal storage disorders, such as TSD, because in many LSDs, mutant enzymes are synthesized but cannot be trafficked to lysosomes due to misfolding and retention in the endoplasmic reticulum (Picache et al. 2022).

Pharmacological chaperones have a multitude of advantages over ERTs. These include the capability to penetrate the BBB to treat neuronal symptoms, lower manufacturing costs, lack of immunogenicity issues, and perhaps the most significant, the potential for convenient oral administration (Picache et al. 2022). Oral administration, compared to the repeated brain injections necessary for ERT in TSD, is a much more desirable method of delivery because of its non-invasiveness, patient compliance, convenience of administration, cost-effectiveness, and ease of large-scale manufacturing (Alqahtani et al. 2021).

However, pharmacological chaperones do not come without risks. In some cases of TSD, the body recognizes the misfolded Hex A enzymes relatively quickly, which causes them to degrade before they can move to the lysosome (Picache et al. 2022). This poses a potential need for the combination of different drugs in order to achieve successful results. For example, the chaperone compound pyrimethamine (PYR) has been found to inhibit Hex A degradation, but only showed temporary increases in Hex A activity during clinical trials (Picache et al. 2022). This could indicate that adding another chaperone that can correct protein structure may have efficacy in combination with PYR. However, this creates the possibility of harmful interactions between drugs, which can lead to undesirable side effects in patients. These risks must be thoroughly considered when developing treatments involving multiple chaperones. Overall, enzyme enhancing therapy does not seem to be among the most promising treatment options for TSD, but it cannot be discounted completely due its potential for convenient oral administration and ability to combine with other treatments.

Substrate Reduction Therapy (SRT)

Substrate reduction therapy (SRT), rather than directly working to fix the mutant enzyme itself, inhibits the formation of specific substrates of the enzyme to reduce the need for it. In lysosomal storage disorders, this translates to a decrease in substrate accumulation (Picache et al. 2022). Many SRTs, like enzyme enhancing therapies, are in the form of a drug that can be administered orally.

SRT has been tested in animal models of TSD as well as in clinical trials. The drug miglustat is an immunosugar inhibitor of glucosylceramide synthase (Picache et al. 2022), an enzyme that catalyzes the synthesis of the glycolipid glucosylceramide. This reduces the accumulation of glucosylceramide in patients with Gaucher's disease. Gaucher's disease is caused by a deficiency of the enzyme glucocerebrosidase, which catalyzes the conversion of glucocerebroside back to glucose and ceramide ("Glucocerebrosidase"). Miglustat was found to be partially efficacious in TSD mice, and due to these results, as well as other notable attributes such as the ability to pass through the BBB, miglustat was then tested in multiple TSD patients (Picache et al. 2022)..

One clinical trial involving the use of miglustat in the infantile form of TSD included two TSD patients, aged 11 months (Patient 1) and 14 months (Patient 2). Both patients initially received an orally administered, daily miglustat dose of 100 mg divided into 3 doses (Bembi et al. 2006). This again calls to attention one of the main advantages of SRT: an administration method that is convenient for the patient and reduces the heavy burden upon them. After 3 months of 100 mg/day, the total daily dose was doubled and maintained for another 3 months. However, this led to the appearance of persistent diarrhea, and the dosage was reverted back to the original amount (Bembi et al. 2006). At baseline, both patients exhibited many classical TSD symptoms, including the cherry red spot and psychomotor retardation. Despite this, head circumference was normal at baseline and remained that way.

In Patient 1, heat intolerance and fever that were present during the first few months following baseline disappeared during therapy. In addition, neither patient developed a pathologic increase of head circumference, referred to as macrocephaly, in the follow-up period, as is normally expected in the course of TSD. Aside from the persistent diarrhea due to an increase in dosage, no adverse effects were observed (Bembi et al. 2006).

However, these are perhaps the only successes of this trial. During the follow-up period, progressive neurodegeneration was evident, along with an increase in severity of symptoms. Nuclear magnetic resonance (NMR) imaging uses radio waves, a powerful magnet, and a computer to make a series of detailed images ("Nuclear magnetic resonance"). NMR scans of both the patients' brains showed development of severe atrophy and worsening of the myelination pattern, indicating critically damaged neurons. Thus, despite its success in a mouse model of TSD, the use of miglustat was unable to modify the clinical course of the disease (Bembi et al. 2006). Cases like this one serve as a reminder that the results of animal studies do not always translate perfectly into humans. This is an important factor to consider when treating rare diseases.

Although the results of this trial were overall unsuccessful, it is still important to analyze their potential significance in future treatments for TSD. The fact remains that while miglustat alone did not affect the course of the disease and is therefore not yet a viable treatment, that does not render the drug entirely futile. The authors of the study believe that their result of neither patient developing macrocephaly may be due to a secondary anti-inflammatory effect of miglustat, reducing the substrate accumulation and consequently slowing the rate of inflammation in the CNS (Bembi et al. 2006). This is significant because it raises the possibility that miglustat may have more success in combination with another treatment; for instance, with a therapy that is found to be successful in other aspects but lacking in macrocephaly treatment. Although no such treatments have been found, miglustat is still an important compound to keep in mind for future research.

Another SRT, Genz-529468, which is 250-fold more potent than miglustat, has also been tested in mice. Despite an increase in intracellular GM2, the treatment resulted in a delayed loss of motor function and longer lifespan. Furthermore, anti-inflammatory responses such as lowered microglial activation, lowered astrogliosis, and delayed neuronal apoptosis have been observed in mice treated with both miglustat and Genz-529468. This indicates that there is still potential for the use of anti-inflammatory SRTs as treatment for GM2 gangliosidosis disorder such as TSD (Picache et al. 2022).

Hematopoietic Stem Cell Transplantation

Hematopoietic stem cell transplantation (HSCT) is mainly achieved through transplantation of stem cells from peripheral blood, bone marrow, or umbilical cord blood, which is made possible by the ability of Hex enzymes to pass from cell to cell. HSCT has proven successful in several lysosomal storage disorders, including Mucopolysaccharidosis (MPS) I and II and Gaucher disease type 1 and 2 (Picache et al. 2022). Bone marrow transplantation (BMT), which falls under the umbrella of HSCT, has been tested in murine Sandhoff disease, which is closely related to TSD in humans. Therefore, the results can be extrapolated to indicate what may happen if the treatment were to be translated to TSD. The following trial in Sandhoff mice examines the effect of BMT on microglial activation preceding neuronal death (Wada et al. 2000).

Before treatment, symptomatic Sandhoff mice had prominent apoptotic neuronal cell death in the caudal regions of the brain. Based on gene expression and histologic analysis, activated microglial expansion was found to precede the neuronal death characterized by the disease. In Sandhoff mice ranging from 1-4 months of age, an expanded population of activated microglial cells in the functional tissue of the spinal cord, brainstem, and thalamus was observed, compared to the resting microglia in the brainstems of control mice (Wada et al. 2000).

The Sandhoff disease mice were then treated using bone marrow transplantation. After BMT, no difference in glycolipid levels between treated and untreated mice was observed, demonstrating that the treatment was not effective in decreasing ganglioside storage. However, apoptotic death was virtually absent in the spinal cord, brainstem, and thalamus of BMT-treated Sandhoff mice at 4 months of age compared to the abundant apoptosis in untreated mice, indicating that the treatment was successful in reducing neuronal death. Activated, amoeboid microglia were present at a much lower density compared to untreated mice, with a 78% decrease in the spinal cord and a 91% decrease in the brainstem. This demonstrates yet another success of the study, as the presence of activated microglia can trigger death in neurons already compromised by excessive ganglioside storage (Wada et al. 2000).

This study suggests that activated microglia were not simply a reaction to massive cell death, but rather were present throughout the course of the disease in the same regions that eventually demonstrated a high degree of cell death. The activated microglia produced a proinflammatory cytokine known as TNF-α, which has been shown to have both neurotoxic and neuroprotective effects (Wada et al. 2000). Its potential neurotoxic effects may explain why neuronal death in Sandhoff disease was preceded by activated microglia, as the TNF-α produced may have increased the amount of cell death.

Additionally, brain samples from a Sandhoff disease patient were histologically and biochemically examined to determine if an inflammatory reaction similar to that in the untreated Sandhoff mice was present in the human disorder. Neurons distended with ganglioside storage were found throughout the patient's brain, along with neuronal loss. Apoptotic death was detected in the cortex, brainstem, thalamus, and cerebellum. Activated microglia with a plump, amoeboid shape were also observed in the patient (Wada et al. 2000). The similarities between the brains of the human patient and Sandhoff mice indicate that this treatment could potentially be translated to human Sandhoff and TSD patients and show similar results.

Overall, BMT in Sandhoff disease mice has had both successes and shortcomings. The treatment largely suppressed microglial activation and neuronal cell death, but did not result in beta-hexosaminidase-positive neurons, nor a decrease in neuronal GM2 ganglioside storage in the spinal cord and brainstem (Wada et al. 2000). However, perhaps the most significant result was that beta-hexosaminidase-positive microglia could be detected. This is significant because this demonstrates the relationship between the inflammatory response of activated microglia and acute neurodegeneration. In the mouse model, glycolipid accumulation is the primary cause of neuronal dysfunction and damage. Microglia recognize damaged and dying neurons and remove them through phagocytosis, a process that can be expected to elicit an inflammatory reaction. However, this is exacerbated by the inability of the microglia to degrade the endocytosed glycolipids due to their own Hex A deficiency. This issue is resolved by the introduction of normal microglia into the CNS utilizing BMT. These Hex A-positive microglia are able to carry out their function of removing damaged storage neurons. Therefore, the expansion of activated microglia is suppressed and the neurodegenerative process is slowed by reducing the neuronal insult from the excessive inflammatory response (Wada et al. 2000).

However, anti-inflammatory agents alone are not expected to halt disease progression in infantile GM2 gangliosidosis diseases such as Sandhoff and TSD, as neuronal glycolipid storage will continue to accumulate. This introduces the potential for combination therapies involving BMT that could lower both the inflammatory process and glycolipid storage (Wada et al. 2000).

Subsequently, allogeneic BMT followed by substrate reduction therapy was tested in a child with TSD. It is important to note that the patient was asymptomatic at the time of transplantation aside from two episodes of seizures, as it has been found that BMT is not effective when severe neurological symptoms are already present (Jacobs et al. 2005). This brings to light the importance of early diagnosis and treatment, especially in a disorder like infantile TSD.

BMT was performed at 3 years and 10 months, and was followed by the experimental drug Zavesca ®, or miglustat, which is an SRT that has been shown to reduce ganglioside accumulation and prevent neuropathology in TSD mice, as discussed previously. This was administered to the patient when they were 5½ years old (Jacobs et al. 2005). However, the treatment was not as successful in the patient. Although Hex A activity levels in leucocytes and plasma did increase to normal values directly after BMT, they decreased after 3 months. BMT was unable to prevent seizures, deterioration of motor skills and speech, or cerebral atrophy. The introduction of Zavesca ® after BMT did not influence the deterioration either (Jacobs et al. 2005). A very slight increase in Hex A levels 5 months after starting Zavesca ® can be observed, which could potentially indicate a difference after adding the drug to the treatment, but the amount appears statistically insignificant and therefore does not amount to a promising result.

Overall, this therapeutic regimen was unsuccessful in treating the disease. As earlier stated, both SRTs and BMT have the potential for success in combination with other therapies despite their individual shortcomings, but even combination was inefficacious in the case of BMT followed by Zavesca ®. This is why many different treatment options must be explored in order to determine what works and what does not.

BMT has also been tested in a 15-year-old late-onset TSD patient. This was more successful, resulting in Hex A activity close to normal levels 8 years after the transplant. Although the treatment did not entirely alleviate the patient's symptoms, it did prevent neurological degeneration such that they were "tolerable for daily life" (Picache et al. 2022). The authors of the study believe that poor outcomes of HSCT in forms of TSD other than late-onset may be attributable to an insufficient enzyme dose or the treatment being performed too late. This raises the question of whether newborn screening for TSD should be considered. The authors also note that this patient is the first case of late-onset TSD who has had a positive disease outcome with HSCT (Stepien et al. 2018).

This effort shows that despite the shortcomings of HSCT in the previously discussed trials, the approach still has some promise. It is possible that it will generally have more efficacy in late-onset TSD patients than in the infantile form, especially considering that the infantile patient treated with BMT was asymptomatic, while the late-onset patient had already been exhibiting a variety of symptoms (Stepien et al. 2018). This demonstrates that not only the point in the timeline of disease progression at which the treatment is delivered has an effect on its viability, but also the form of the disease.

In addition, inefficacy thus far in infantile TSD is not the only hurdle pertaining to HSCT. The challenge remains of finding a proper donor match, as well as managing the negative immunogenic responses of an improper match. To combat this impediment, there have been recent efforts in other diseases to transplant a patient's own gene therapy-corrected stem cells into them (Picache et al. 2022), referred to as autologous HSCT, that could potentially be used for TSD. This can be preferable over the use of donor cells, as the patient's body may recognize that these donor cells are not self and reject them, resulting in an immune response. Although patient autologous stem cell replacement therapy is a promising treatment option for TSD, it is an extremely long and complicated process. Firstly, the patient must be healthy enough and produce enough stem cells to be utilized in the treatment. Secondly, not only is there a need to harvest these stem cells from the patient, but also to correct them using gene therapy and then administer them back into the patient. In addition, patients often have to go through leukodepletion through radiation and or chemotherapy so that their immune system will be more willing to accept transfused cells (Khaddour et al. 2024). Therefore, this is yet another treatment with a very high patient burden that decreases feasibility in offering a solution to TSD.

Antisense Oligonucleotide (ASO) Therapies

Therapies with mRNA and antisense oligonucleotide (ASO) technology use a single-stranded oligonucleotide, a short nucleic acid polymer, that binds to ribonucleic acids (RNA) in order to modify gene expression or gene splicing, and in turn enzymatic expression and activity (Picache et al. 2022). In total, there are currently nine ASO drugs that have been approved by the FDA and the European Medicines Agency (EMA), and the predominant mechanisms of these approved ASOs are knockdown of target genes and correction of mutations through exon skipping (Picache et al. 2022). This is a form of RNA splicing used to cause cells to “skip” over misaligned or faulty sections, or exons, of genetic code, leading to a functional protein despite the genetic mutation (Kyriakopoulou et al. 2023). However, the drug nusinersen, which has been approved for the treatment of spinal muscular atrophy, another rare disease, works through a different mechanism. The drug corrects splicing of a redundant gene (Picache et al. 2022), one that performs the same function as another, the inactivation of which has little to no effect on biological phenotype (Nowak et al. 1997). The approval for nusinersen, which is administered via intrathecal injection into the cerebrospinal fluid of the spine, provides precedence for direct CNS delivery, despite the fact that efficient delivery of ASOs to target cells remains a challenge *in vivo* (Picache et al. 2022). This makes ASOs a potential viable therapeutic option for TSD, for which treatment must target cells in the nervous system.

The exon skipping strategy is unlikely to be successful in TSD, due to the nature of Hex A. Specifically, Hex A has multiple regions that are critical to form the catalytic pocket, dimerization surface, GM2 activator binding side, disulfide bonds, and post-translational modifications necessary for trafficking (Picache et al. 2022). However, it is possible that a different approach using ASOs for genetic SRT may be efficacious. For example, ASOs have been successfully used to decrease glycogen synthase 1 activity and levels in mice with Pompe disease, another LSD. Additionally, it stands to be experimentally determined which of the 43 out of over 200 *HEXA* mutations along chromosome 15q23 that cause splicing defects would be amenable to ASO correction (Picache et al. 2022). This warrants further investigation into therapies utilizing ASOs for the treatment of TSD, as a treatment that is able to both modify enzyme expression and reduce substrates may be possible using this method. A combination of the two strategies may have a greater chance of success at treating the disease. However, as discussed previously, intrathecal administration generates a high patient burden, which must be considered for all therapeutic methods.

Gene Therapy

Gene therapy is perhaps the most promising therapeutic modality for treatment of TSD. Although the research discussed involving other therapeutic methods is slightly dated, it establishes an important background on why gene therapy is the most viable option. Many strategies that are perhaps more straightforward than gene therapy have been tested, but the results have not been as promising. This justifies gene therapy as the method with the most potential to treat TSD, despite the many unknowns associated with it.

Gene therapy primarily employs the use of viral vectors for the delivery of a functional gene to correct a genetic mutation (Picache et al. 2022). Gene therapy is an especially promising treatment for TSD, since it is caused by a monogenic mutation. One method of gene therapy administration is *ex vivo,* where a patient's own cells are isolated, transduced with viral vectors such as HIV-1 derived lentivirus or gamma-retrovirus, then reintroduced into the body (Picache et al. 2022). The other method is *in vivo,* in which a gene therapy is directly administered to the patient through injection of a genetic package-carrying viral vector such as adeno-associated virus (AAV). This gene therapy modality has been administered in about 250-300 clinical trials and has had a good overall safety record in patients (Picache et al. 2022).

There are currently several gene therapies that have been approved by the FDA and EMA, one of which is for an LSD. Libmeldy has been approved by the EMA as an *ex vivo* gene therapy treatment for metachromatic leukodystrophy (MLD) and is currently in clinical trials in the U.S. The preferred *ex vivo* gene therapy strategy for an LSD is to transduce a patient's own hematopoietic stem cells with the gene of interest as a modified HSCT therapy (Picache et al. 2022). The primary *in vivo* strategy to treat LSDs, since it is difficult for most AAVs to cross the BBB, is to administer the treatment locally via intracerebral or intrathecal routes (Picache et al. 2022). Due to the predominantly neurological nature of TSD, this is a promising method.

However, the unique challenge of gene therapy for TSD remains that the gene *HEXA* encodes only the alpha subunit of the heterodimer Hex A enzyme, and needs to complex with the beta subunit, encoded by *HEXB,* in order to be functionally active. Therefore, gene therapy for TSD requires the delivery of not one, but two subunits to form a functional Hex A enzyme. However, this strategy has an advantage as well: a single therapy can be utilized to treat two GM2 gangliosidosis diseases, TSD and Sandhoff disease. To do so, several strategies have been tested in animal models (Picache et al. 2022).

One animal model that has been used to test gene therapy for TSD are Sandhoff mice, as they manifest many signs of classical human TSD. With the acute course of the disease, these mice die before 20 weeks of age (Cachón-González et al. 2006). In one study, adult Sandhoff mice were treated by stereotaxic intracranial inoculation of recombinant AAV vectors encoding complementing human Hex A alpha and beta subunit genes and elements, which are similar to mouse endogenous genes, including an HIV tat sequence to enhance protein expression and distribution (Cachón-González et al. 2006). The mice treated with the alpha subunit will be referred to as alpha-transduced, and those treated with the beta subunit will be referred to as beta-transduced. This study, like the previously discussed BMT trial, mentions the connection between inflammatory responses involving microglia and bone marrow-derived macrophages to neurological decline in GM2 gangliosidoses (Cachón-González et al. 2006). The results of the study are overall very promising. The animals transduced with both the alpha and beta subunit survived for over a year, with sustained, widespread, and abundant Hex A delivery in the nervous system. Onset of the disease was delayed, motor function was preserved, and inflammation and GM2 ganglioside storage in the brain and spinal cord was reduced (Cachón-González et al. 2006). Histological examinations of the brain and spinal cord show a significant difference between Hex A activity in treated and untreated mice. Hex A levels in treated mice appear much closer to disease-free wild type mice, and are also widespread throughout different parts of the brain and spinal cord. In addition, GA2 and GM2 storage levels in the cerebrum and cerebellum trend downwards in treated mice (Cachón-González et al. 2006). These data demonstrate that the treatment was effective. The results also show a reduction of microglial immune responses in a treated sample as compared to untreated, also appearing closer to the wild type mice with very few cells of microglia/macrophage lineage observed (Cachón-González et al. 2006). As activated microglia are also a marker indicating neurological injury, these results demonstrate the reduction of this marker and therefore effective treatment of the disease.

In addition, the weight of the mice over time was monitored. In TSD, as the body degenerates, subjects lose weight as they rapidly decline. This is why weight is an important marker for treatment efficacy. Out of all groups of treated mice, the beta-transduced mice have the closest pattern of weight gain to the wild type. In contrast, the weight of the alpha-transduced mice increased at first, but then quickly declined (Cachón-González et al. 2006). This indicates that treatment using both the alpha and beta subunits of Hex A was most effective in this study at maintaining weight gain. Hind limb performance, which is a common measure of motor function in animal models, was also recorded in order to examine the effect of the treatment on symptoms. The treated mice once again appear much more similar to the wild type than the untreated group (Cachón-González et al. 2006). This means that the study was overall successful in treating the loss of motor function characterized by Sandhoff disease. The authors of the study also concluded that gene delivery of Hex A through AAV vectors has realistic potential for treating TSD in human patients (Cachón-González et al. 2006).

Although mice are a good starting point, it is also important to understand how therapeutics work in larger animal models that may better reflect the normal course of the human disease. For this reason, a sheep model of TSD, the experimental model with the most similar clinical signs of the disease to human patients, has also been used to test potential strategies for treatment. One study utilized two variations of intracranial gene therapy to treat the TSD sheep: AAVrh8 monocistronic vectors encoding the α-subunit of hexosaminidase (TSD α) and a mixture of two vectors encoding both α and □ subunits separately. These were both injected at either a high or low dose. The results of this study were promising as well. There was a delay in symptom onset and/or reduction of acquired symptoms in all AAV-treated sheep, with the greatest mean lifespan being 13.6±2.5 months in the TSD α+□ low-dose cohort, an increase of about 50% compared to untreated sheep. However, superior levels of Hex A formation and vector genome distribution were observed in the brain of TSD α+□ sheep compared to TSD α sheep. Ganglioside clearance was also the most widespread in TSD α+□ high dose sheep. In addition, histological images show that many regions of the brain had similar morphology to normal sheep after treatment compared to untreated TSD sheep. Microglial activation and proliferation lessened after gene therapy, indicating a positive outcome. However, vector genome and Hex A distribution in the spinal cord was low in all groups (Gray-Edwards et al. 2018), which is perhaps the only significant shortcoming of this study.

To conclude, the α+□ high dose group of sheep had the most promising results. The major issue of this study that must be resolved is the lack of Hex A distribution to the spinal cord. This is important because an ideal treatment would reduce the toxic levels of GM2 gangliosides in the spinal cord in addition to the brain in order to preserve nerve cells in both areas primarily affected by the disease. The notable differences between the results of high and low dose of treatment must be considered, as they demonstrate a relationship between dosing and subsequently efficacy of the therapy. Despite the successes of high dosage, it may not be feasible to immediately treat patients with high doses from the start, as it is important to find a balance and to try to avoid negative immune responses to the treatment. The results of this study demonstrate therapeutic efficacy for TSD in a sheep brain, which is on the same order of magnitude as a child's brain (Gray-Edwards et al. 2018), meaning their overall size and capacity are comparable.

The findings of this study have now been applied to a clinical trial involving two TSD patients. The first patient, referred to as TSD-001, demonstrated developmental delays at 5-6 months of age, an exaggerated startle reflex at 8 months, and macrocephaly, severe seizures, and abnormal myelination by 1 year of age (Flotte et al. 2022). At the time of treatment, the patient had advanced cortical atrophy, mild ventricular enlargement, and diffusely affected white matter. The patient was treated at 30 months using an equimolar mix of AAVrh8-HEXA and AAVrh8-HEXB, administered via intrathecal injection. 75% of the dose was delivered to the cisterna magna in the posterior of the brain, and 25% to the thoraco-lumbar junction. The patient was hospitalized 9 months after treatment and 3 months after immune suppression, due to viral pneumonia that was not definitively linked to the treatment (Flotte et al. 2022). This may indicate a potential adverse effect. Hex A was detectable in TSD-001 at 6 months post-treatment, and there was no further deterioration of the patient. At the time of publication, the patient was 4.5 years old and remained seizure-free (Flotte et al. 2022).

The second patient of the study, TSD-002, had two preceding siblings with TSD, one of which died before 3 years of age. TSD-002 was treated at 7 months of age with a combination of bilateral thalamic and intrathecal infusion of AAVrh8-HEXA and AAVrh8-HEXB. This patient was also hospitalized 4 months post-treatment with a febrile urinary tract infection. Hex A was detectable at 3 months post-treatment, but the mild dysmyelination present at baseline remained unchanged. An MRI also at 3 months post-injection showed stabilization of the disease, but this was only a temporary deviation from the natural history of infantile TSD. The patient declined again 6 months post-treatment, despite the areas of increased myelination present at this time. The patient also developed seizures by 2 years of age, between months 13 and 17 of treatment. However, since this patient was treated at an early symptomatic stage, Hex A activity increased in the cerebrospinal fluid (CSF) and ongoing myelination was apparent (Flotte et al. 2022).

Both patients also underwent immunosuppression with sirolimus, corticosteroids, and rituximab in order to be more receptive to the treatment (Flotte et al. 2022), illustrating the complexity of the gene therapy process. The injection procedures were well tolerated, with no vector-related adverse events to date. Overall, CSF Hex A activity nearly doubled from baseline, increasing from ~0.3 nmol/hr/mL to 0.5-0.6 nmol/hr/mL after treatment, and remained stable. Serum Hex A activity increased slightly at all time points except at 6 months post-treatment in TSD-002. In addition, a western blot of CSF showed faint bands corresponding to Hex A and Hex B proteins, indicating low level expressions of these enzymes and demonstrating a positive result of the study on a molecular level. Neither patient showed an increase in anti-capsid IgG from baseline to 6 months post-treatment (Flotte et al. 2022), indicating no negative immune responses to the viral vector, which contrasts with the marked increase caused by other intrathecal AAV gene therapies in the past (Flotte et al. 2022).

The data in the study shows an increase in Hex A activity in the CSF of both patients, but levels in serum remained relatively constant from baseline. The western blots for TSD-002 show more improvement than TSD-001, and increased myelination is apparent in TSD-002 (Flotte et al. 2022). This implies that the treatment, before the effects began to dissipate in TSD-002 as indicated by the patient's eventual decline, had better results in TSD-002 than TSD-001. This may be due to the age difference at the time of treatment, meaning it does make a difference to treat the disease at an earlier age. Although there are some potential marginal improvements that can be made, the study was overall very effective. This was only the first in-human AAV gene therapy in TSD, and these are only the early stages of clinical trials. The results are promising enough to use the information and build on it; for example, by increasing doses, enrolling patients earlier, or other minor adjustments to this therapy that has already proven itself to be safe and effective. To conclude, the data found in this study is quite promising and warrants further exploration into this gene therapy strategy.

Discussion

As TSD has a very short window for treatment, it is important to identify the most promising path forward. In addition, any potential challenges associated with this path must be discussed.

Gene therapy viability as a treatment for TSD

Out of all methods discussed, gene therapy appears to be the most promising for the treatment of TSD. Animal and clinical trials employing the use of gene therapy have had the most overall success in both increasing Hex A levels and arresting the natural disease course. However, there are many important factors surrounding gene therapy that must be taken into consideration.

Gene therapies are known to come attached with excessively high price tags. Although this can be justified by the high cost of development and manufacture of the therapy (Carvalho et al. 2021), it poses a major affordability issue for many patients and families. In addition, there may also be other costs associated with gene therapy, such as the need for increased medical follow-up for surveillance of adverse effects (Carvalho et al. 2021). Although clinical trials attempt to be thorough in their anticipation of side effects, issues may still arise, especially in a larger patient population than the clinical trial numbers initially treated. It may not be provable that these issues are directly tied to the therapy, which may result in difficulty for the patient to get coverage or care.

Furthermore, many gene therapy products have been developed using government or public funds, but this is not taken into account when deciding on the final price that patients must pay (Carvalho et al. 2021). Ideally, the investment of government funds should be reflected in the pricing of the treatment, and it should be made more accessible to those with government healthcare or insurance.

Many health insurers may also be unwilling to reimburse the costs of gene therapy (Carvalho et al. 2021), which further exacerbates the affordability issue. It is unclear what reimbursement strategies should be used, as well as how to appropriately discount gene therapies (Carvalho et al. 2021). One recently approved gene therapy for sickle cell disease, Casgevy, hit the market with a price tag of $2.2 million (*Vertex Pharmaceuticals*). It is unclear how many insurance companies are going to cover or reimburse this amount (Coombs 2024).

This hefty price tag is not unique to any one disease. UMass Medical School was given $1.4 million in order to advance the previously discussed gene therapy clinical trial for TSD ("Blu Genes Foundation"), demonstrating how high the cost of the treatment may be if it hits the market in the future, as this number is likely to climb. As the role of insurance in this payment remains unclear, affordability will likely be a major issue in the future of TSD treatment.

In addition, there are variations in how health technology assessment (HTA) agencies recommend payment coverage across different countries, which leads to different levels of patient access to therapy. HTA systems may not account for relevant factors such as the ability of patients to return to work, work productivity, or impact on caregiver burden (Carvalho et al. 2021). These elements must be considered in order to achieve proper treatment reimbursement.

There also seems to be a misalignment between the needs of payers and regulators. This can lead to a therapy being approved for commercialization by health authorities but not reimbursed by insurers (Carvalho et al. 2021), which may render the treatment inaccessible and therefore unable to realize its full impact on patients. This demonstrates the need for coordination and synergy between the different parties involved in making gene therapies accessible to patients. If a gene therapy for TSD similar to the one delivered in the clinical trial is approved by the FDA but not reimbursed by insurance companies, this poses a serious issue for patients. There is an extremely tight window for the treatment of TSD, as infantile patients have a life expectancy of only 5 years of age (Ramani and Parayil 2023). These patients and their families cannot afford to have such restrictions to treatment access, as they have so little time. In addition, the results of the gene therapy clinical trial for TSD as described by Flotte et. al suggest that treating the disease earlier in its course has an effect on treatment efficacy (Flotte et al. 2022). This further demonstrates that any unnecessary delay between the approval of treatments and the administration to patients must be eliminated. Regulators and payers must be in coordination so that TSD patients are able to receive treatment as soon as possible after it is approved.

Ethical considerations related to TSD gene therapy access

There are also a number of ethical factors that may influence patient access to gene therapy, including individual values and beliefs. For instance, some patients may be unwilling to receive gene therapy due to religious beliefs or simply being intrinsically against genetic therapy. It is also important to ensure that patients are given the right of informed consent. Many patients are apprehensive that they will not receive all the information about the treatment. Some may incorrectly assume that gene therapy can alter features such as identity and personality (Carvalho et al. 2021). This is why it is important that patients thoroughly understand their rights and have access to all necessary information about the treatment they are receiving.

This is related to a general lack of genetic literacy from both patients and caregivers, which in turn could lead to an inaccurate perception of gene therapy. Many also fear that gene therapies will be used irresponsibly in order to improve intellect or physical ability, therefore enhancing social inequality. Patients may also be unwilling to receive gene therapy due to psychological challenges, such as not being in an appropriate mental state to go through treatment after receiving the burdensome news of testing positive for a genetic disease (Carvalho et al. 2021). This burden on patients can be lessened by doctors and other professionals who are properly equipped to explain treatments and reassure patients and families.

This relates to autonomy, the first principle of biomedical ethics. Autonomy ensures that all healthcare procedures must represent the patient's wishes, and that their decision making process should be free of manipulation, coercion, and deceit ("The Four Principles"). In order to assure the autonomy of gene therapy patients, it must be made certain that they are aware of exactly what the treatment entails. Any doubts that patients may have must be truthfully answered and properly explained.

Patients must also be assured of the second and third principles of biomedical ethics, non-maleficence and beneficence. These ensure that treatment does not result in any undue harm to patients without the promise of substantial benefit to their welfare ("The Four Principles"). Patient burden can be further alleviated if they believe that their treatment meets these standards.

Additionally, patient access to gene therapy may also be restricted by socioeconomical, cultural, and geographical factors. Differences in cost and availability of therapies based on geographical region can cause further discrepancies in patient access (Carvalho et al. 2021). This brings to mind the fourth principle of biomedical ethics, justice, which is defined as the "fair, equitable, and appropriate distribution of benefits and norms ("The Four Principles")." It must be considered how to fairly distribute treatments such as gene therapy and how to make them accessible to those who need them, especially by those who have control over access to new therapeutics, such as insurance companies and governments.

For TSD, there are copious amounts of information that must be delivered to patients and families just about the disease itself. They should be informed about all expected and potential outcomes of the disease, including progressive neurodegeneration, seizures, and risk of recurrent infections. Those who are carriers or at risk of being carriers must receive appropriate genetic counseling4 from professionals who are properly prepared and trained to handle these conversations.

In addition, patients must be fully informed about the process and mechanisms of the treatment they are receiving. The novelty of gene therapy compared to classic treatments means it requires formal healthcare professional training (Carvalho et al. 2021). Medical professionals should be properly equipped to explain the delivery of alpha and beta subunits of Hex A to those who may not have an education background in the sciences, as well as the use of AAV vectors. Families may fear that their children may contract adenoviruses from the treatment. Any emotions and doubts like these that families may have must not be dismissed, and should all be properly addressed. Physicians should also be adequately trained to explain the benefits and risks of gene therapy to patients and families (Carvalho et al. 2021).

In addition, medical professionals must do everything in their power to lessen the burden on patients and families. Gene therapy administration generally involves an especially high patient burden. Patients need to be hospitalized for variable periods of time (Carvalho et al. 2021), as the administration of intrathecal and intracerebral injections must be in the hospital setting. As discussed previously, injections into the brain are inconvenient and pose a number of risks, further contributing to burden on TSD patients. Patients may also have to travel in order to receive treatments that are only available in specific locations, which poses another hurdle for patient access (Carvalho et al. 2021).

Finding easily measured patient-centered outcomes to assess efficacy of treatment has been identified as another hurdle that may affect patient access. Many trials also have limited data and time to test therapies. Surrogate endpoints are also often used (Carvalho et al. 2021). These are results of a clinical trial that indicate certain successes of a treatment, such as a shrinking tumor or lower levels of a disease biomarker. They can be measured earlier than clinical endpoints such as longer survival or improved quality of life, but they are not always true indicators of treatment efficacy ("Surrogate endpoint"). Therefore, these may not allow capturing the combined benefit-risk profile of a therapy and may not ultimately translate to long-term benefits for a patient (Carvalho et al. 2021), as surrogate endpoints often draw premature conclusions. For example, patient TSD-002 in the gene therapy clinical trial conducted by Flotte et. al showed an increase in Hex A activity and myelination, but ultimately declined and developed seizures (Flotte et al. 2022). This demonstrates how common surrogate endpoints for TSD may not always translate to lasting clinical success.

Another factor that may restrict patient access is the process of getting patients diagnosed and recruited into clinical trials, as well as encouraging adherence to medical follow-up (Carvalho et al. 2021). The process of consultations following the initial diagnosis of TSD is very complex and involves multiple professionals. The complexity and longevity of this process not only increases the burden on patients and families, but also prolongs the period of time in which patients are not receiving treatment. Certain hurdles may also arise during this process, such as reluctance to MRIs and EEGs or complications involving gastrostomy feeding. The more complex the process is, the more potential there is for restrictions on the patient's access to gene therapy. As stated earlier, it is important for infantile TSD especially that patients receive treatment as soon as possible. The post-diagnosis process may also create more affordability issues for families, as it can be assumed that the cost would be excessive in line with other approved therapies, which again poses a hurdle for TSD patients to receive treatment.

Yet another hurdle that may restrict patient access to gene therapy is the uncertain long-term efficacy of gene therapy products. Most clinical trials for gene therapy are conducted in limited patient populations in which the main endpoints for clinical efficacy are unfamiliar to the scientific community, and it may be unclear how to measure and evaluate them. Whether or not these endpoints are the best choice only becomes clear after time (Carvalho et al. 2021).

Conclusion

Tay-Sachs disease is an extremely severe neurodegenerative disorder that presents itself in 1 in 320,000 live births in the U.S. (Ramani and Parayil 2023). Historically, treatment options for TSD are very limited, and the scientific community has been exploring novel therapeutic approaches such as gene therapy. However, gene therapy has had limited success in arresting the disease course in human patients, although scientists have made significant progress in animal models. Gene therapy, although poised to benefit a great number of people suffering from genetic disorders, is still in the early stages of development. There are several important considerations that come along with it, from financial aspects to questions of ethics. However, many of these therapies have shown promise to make a substantial impact on the lives of patients who have previously had no options whatsoever. Therefore, it is important that scientists continue to work towards developing gene therapies for severe diseases like TSD, and that governments continue to invest funding in research that explores the mechanisms of TSD and its potential treatments.

Works Cited

"About Tay-Sachs Disease." National Human Genome Research Institute. https://www.genome.gov/Genetic-Disorders/Tay-Sachs-Disease.

Akhtar F, Bokhari SRA. Down Syndrome. [Updated 2023 Aug 8]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK526016/

Alqahtani, Mohammed S et al. "Advances in Oral Drug Delivery." *Frontiers in pharmacology* vol. 12 618411. 19 Feb. 2021, doi:10.3389/fphar.2021.618411

"Approved Cellular and Gene Therapy Products." Food and Drug Administration. https://www.fda.gov/vaccines-blood-biologics/cellular-gene-therapy-products/approved-cellular-and-gene-therapy-products.

Atkinson, Arthur J Jr. "Intracerebroventricular drug administration." *Translational and clinical pharmacology* vol. 25,3 (2017): 117-124. doi:10.12793/tcp.2017.25.3.117

Baird, P A et al. "Genetic disorders in children and young adults: a population study." *American journal of human genetics* vol. 42,5 (1988): 677-93.

Barton, N W et al. "Replacement therapy for inherited enzyme deficiency--macrophage-targeted glucocerebrosidase for Gaucher's disease." *The New England journal of medicine* vol. 324,21 (1991): 1464-70. doi:10.1056/NEJM199105233242104

Bembi, B et al. "Substrate reduction therapy in the infantile form of Tay-Sachs disease." *Neurology* vol. 66,2 (2006): 278-80. doi:10.1212/01.wnl.0000194225.78917.de

"Blu Genes Foundation gives UMMS $1.4M to bring Tay-Sachs gene therapy to clinical trial." *UMass Chan News,* UMass Chan Medical School. 06 Nov. 2018. https://www.umassmed.edu/news/news-archives/2018/11/blu-genes-foundation-gives-umms-$1.4m-to-bring-tay-sachs-gene-therapy-clinical-trial.

Cachón-González, M Begoña et al. "Effective gene therapy in an authentic model of Tay-Sachs-related diseases." *Proceedings of the National Academy of Sciences of the United States of America* vol. 103,27 (2006): 10373-10378. doi:10.1073/pnas.0603765103

Carvalho M, Sepodes B, Martins AP. Patient access to gene therapy medicinal products: a comprehensive review. *BMJ Innovations,* 2021;**7:**123-134.

Cheema, Huma A et al. "Unusual case of Juvenile Tay-Sachs disease." *BMJ case reports* vol. 12,9 e230140. 12 Sep. 2019, doi:10.1136/bcr-2019-230140

Concolino, Daniela et al. "Enzyme replacement therapy: efficacy and limitations." *Italian journal of pediatrics* vol. 44,Suppl 2 120. 16 Nov. 2018, doi:10.1186/s13052-018-0562-1

Coombs, Bertha. "New sickle cell gene therapies are a breakthrough, but solving how to pay their high prices is a struggle." CNBC. 23 Feb. 2024. https://www.cnbc.com/2024/02/23/sickle-cell-disease-gene-therapies-casgevy-lyfgenia-insurance-cost-issues.html.

Dastsooz, Hassan et al. “Identification of mutations in *HEXA* and *HEXB* in Sandhoff and Tay-Sachs diseases: a new large deletion caused by *Alu* elements in *HEXA*.” *Human genome variation* vol. 5 18003. 15 Mar. 2018, doi:10.1038/hgv.2018.3

“Decerebrate posture.” Medline Plus. https://medlineplus.gov/ency/article/003299.htm.

Eng, C M et al. “Safety and efficacy of recombinant human alpha-galactosidase A replacement therapy in Fabry's disease.” *The New England journal of medicine* vol. 345,1 (2001): 9-16. doi:10.1056/NEJM200107053450102

“FDA approval brings first gene therapy to the United States.” Food and Drug Administration. 30 Aug. 2017. https://www.fda.gov/news-events/press-announcements/fda-approval-brings-first-gene-therapy-united-states.

Flotte, Terence R et al. “AAV gene therapy for Tay-Sachs disease.” *Nature medicine* vol. 28,2 (2022): 251-259. doi:10.1038/s41591-021-01664-4

“Gene Therapy.” National Human Genome Research Institute. https://www.genome.gov/genetics-glossary/Gene-Therapy

“Genetic Disorders.” National Human Genome Research Institute. https://www.genome.gov/For-Patients-and-Families/Genetic-Disorders

“Glucocerebrosidase.” LiverTox: Clinical and Research Information on Drug-Induced Liver Injury [Internet]. Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases; 2012-. [Updated 2018 Mar 5]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK548625/

Gray-Edwards, Heather L et al. “Adeno-Associated Virus Gene Therapy in a Sheep Model of Tay-Sachs Disease.” *Human gene therapy* vol. 29,3 (2018): 312-326. doi:10.1089/hum.2017.163

Jacobs, J F M et al. “Allogeneic BMT followed by substrate reduction therapy in a child with subacute Tay-Sachs disease.” *Bone marrow transplantation* vol. 36,10 (2005): 925-6. doi:10.1038/sj.bmt.1705155

“Juvenile Tay-Sachs Disease.” National Tay-Sachs & Allied Disease Association. https://ntsad.org/diseases/tay-sachs-disease/juvenile-tay-sachs-disease/.

Khaddour K, Hana CK, Mewawalla P. Hematopoietic Stem Cell Transplantation. [Updated 2023 May 6]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK536951/

Klinge, L et al. “Enzyme replacement therapy in classical infantile pompe disease: results of a ten-month follow-up study.” *Neuropediatrics* vol. 36,1 (2005): 6-11. doi:10.1055/s-2005-837543

Kyriakopoulou, Eirini et al. “Gene editing innovations and their applications in cardiomyopathy research.” *Disease models & mechanisms* vol. 16,5 (2023): dmm050088. doi:10.1242/dmm.050088

“Macrocephaly.” Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/22685-macrocephaly.

Nowak, M A et al. “Evolution of genetic redundancy.” *Nature* vol. 388,6638 (1997): 167-71. doi:10.1038/40618

“Nuclear magnetic resonance imaging.” National Cancer Institute. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/nuclear-magnetic-resonance-imaging.

Owens, Caitlin. “Multimillion-dollar gene therapies offer hope and huge cost concerns.” *Axios,* 26 Sep. 2022. https://www.axios.com/2022/09/26/gene-therapies-drug-prices-cures.

Picache, Jaqueline A et al. “Therapeutic Strategies For Tay-Sachs Disease.” *Frontiers in pharmacology* vol. 13 906647. 5 Jul. 2022, doi:10.3389/fphar.2022.906647

“Protein Glycosylation.” ThermoFisher Scientific. https://www.thermofisher.com/us/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/protein-glycosylation.html.

Ramani PK, Parayil Sankaran B. Tay-Sachs Disease. [Updated 2023 Jan 25]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK564432/

“Rare Genetic Diseases.” National Human Genome Research Institute. https://www.genome.gov/dna-day/15-ways/rare-genetic-diseases.

Scheller, E L, and P H Krebsbach. “Gene therapy: design and prospects for craniofacial regeneration.” *Journal of dental research* vol. 88,7 (2009): 585-96. doi:10.1177/0022034509337480

Stepien, Karolina M et al. “Haematopoietic Stem Cell Transplantation Arrests the Progression of Neurodegenerative Disease in Late-Onset Tay-Sachs Disease.” *JIMD reports* vol. 41 (2018): 17-23. doi:10.1007/8904_2017_76

“Surrogate endpoint.” National Cancer Institute. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/surrogate-endpoint.

Taylor, Michelle. “Outlook on Cell and Gene Therapy: 2024 and Beyond.” *labcompare,* 23 Jan. 2024. https://www.labcompare.com/10-Featured-Articles/610359-Outlook-on-Cell-and-Gene-Therapy-2024-and-Beyond/.

“Tay-Sachs Disease.” Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/14348-tay-sachs-disease.

“Tay-Sachs disease.” *Genetic and Rare Diseases Information Center,* National Center for Advancing Translational Sciences. https://rarediseases.info.nih.gov/diseases/7737/tay-sachs-disease.

“Tay-Sachs Disease.” National Institute of Neurological Disorders and Stroke. https://www.ninds.nih.gov/health-information/disorders/tay-sachs-disease.

“The Four Principles of Biomedical Ethics.” Healthcare Ethics and Law. https://www.healthcareethicsandlaw.co.uk/intro-healthcare-ethics-law/principlesofbiomedethics.

United States, Security and Exchange Commission. *Vertex Pharmaceuticals Incorporated.* 8 Dec. 2023. https://investors.vrtx.com/node/31276/html.

Wada, R et al. “Microglial activation precedes acute neurodegeneration in Sandhoff disease and is suppressed by bone marrow transplantation.” *Proceedings of the National Academy of Sciences of the United States of America* vol. 97,20 (2000): 10954-9. doi:10.1073/pnas.97.20.10954

European Football: An Analysis of Player Market Value By Saahil Ahuja

Abstract

The market value of a football player has become extremely complex in recent years due to increasing competition, focus on factors other than performance, and massive investments. In order to understand the complexity, this paper carries out data analysis to determine the factors affecting the market value of a player, with specific focus on the European football leagues. The present study aims to identify factors aside from performance and their importance in understanding the causality of player market value. The study also aims to evaluate the FIFA Rating and Career Mode Potential in their valuation of players because these can be important factors and are seldom used.

A sample of 150 players (50 most valuable attackers, defenders, and midfielders) was taken as the dataset from the source ‘Transfermarkt’ and the players’ characteristics were analyzed with the help of correlation and regression analysis. It was found that FIFA Rating, Career Mode Potential, Goals/Assists+Goals, and Instagram followers were positively correlated with the market value of the players. The predictive effect of all these factors was confirmed by the regression analysis. The most important finding was the significant effect of FIFA Rating and social media presence on the market value. The findings can be used by football analysts, player agents, and footballers themselves to understand the key factors that go into being a valuable footballer aside from the performance component.

Introduction

Football is one of the most profit-oriented sports in the sports industry. Data collected from the football sector demonstrates a continuous increase in revenue over the past 20 years (Bács and András, 2016). Aside from being a sport, European football is a huge business involving several stakeholders that requires strategic decisions to be managed effectively. One of the most crucial managerial decisions concerns player transfers (Al-Asadi & Tasdemir, 2022), stirring public interest in the field of market value determination.

The market value of a player is an estimate of the player’s transfer fee; i.e., how much the player’s current team would sell them for and, consequently, how much the buying team would buy them for (Herm, Callsen-Bracker, and Henning Kreis 2016). The market value is the key factor during transfer negotiations. In the last decade, an influx of private investors has been observed in the sport; these investors have superior resources as compared to other investors, which are becoming increasingly important to stay competitive in European football (Kuper 2014). This influx of investment has also had a tremendous impact on the salaries of players.

Deloitte’s Annual Review of Football Finance suggests that the average salary paid by domestic investors (63 million euros) is less than half the average salary paid by foreign investors (137 million euros) (Rohde and Breuer 2016). Since annual salary and market value are positively correlated with the influx of private/foreign investors in football, the impact on the market value of players has also been positive. The average transfer fee of a player has increased from 3.07 million euros in the 2013/14 season to 4.02 million euros in the 2022/23 season (Besson, Ravenel, and Poli, 2023).

The football industry lags in involving data analytics to determine market values (Weinmann et al., 2017). Data analysis is used in training sessions and decisions regarding line-ups; however, its full potential has not yet been realized. The most well-known example of using statistics to analyze players and accordingly build a team was seen in the movie "Moneyball" for the baseball industry. Billy Beane and Peter Brand scout and select a team for the Oakland Athletics purely based on statistics, go against all odds and make the playoffs in Major League Baseball every second season for two decades despite having one of the lowest budgets in the league. This method, used by almost every team in Major League Baseball (Weinmann et al., 2017), has not been used extensively in analyzing football players yet; however, the practice is gaining traction.

Based on previous studies of the player market values and changes in the football industry, numerous factors aside from the performance of the players have been identified to affect the valuation:

- ***Age:*** Age is considered as an important parameter because it indicates the stage of career a player is in. Players aged 18-22 are extremely fit, fresh but require experience. Players between the ages of 22-25 have the highest market values as this is the usual age for the "prime" of the player. Above 25, players have lots of experience but their performance starts declining. So, from an investment point of view, players aged 22-25 are the best because they are likely to play for a long time, they are experienced and they perform well and hence they have the highest market values. (Poli et al., 2023)

- ***Position:*** Attackers have higher market values as they are the ones who score goals for the team and lead to winning matches, eventually this means success for the club. (Salvo et al., 2006) According to Transfermarkt values, among the 20 most valuable players 10 players are attackers, 8 midfielders and 2 defenders.

- ***Key Performance Indicator (KPI):*** Key performance indicators (KPIs) are quantifiable metrics that measure a football player's performance over time to achieve a specific goal. Due to this factor, market values of defenders, goalkeepers, midfielders, and attackers should be analyzed differently (Metelski,2021). The KPI of a defender is the number of duels, saves for a goalkeeper, assists for a midfielder, and goals for an attacker. The KPI for a center back and full back should also be taken differently because full backs are responsible for assists and passes as well.

- ***Brand Value:*** An athlete's brand worth extends beyond their athletic abilities, including their sponsorships, marketability, and general impact. Popularity of a player is an extremely important factor while considering the market value of a player. Popularity is measured by non-performance related criteria, for instance, social media presence that positively influences a player's market value (Franck and Nuesch 2010). It indicates how much the fans like the player and also indicates the social influence of the player. To demonstrate the social potential, in 2012 David Beckham earned more money from advertising than playing football (Kiefer, 2012).

The factors mentioned above have been used when valuing players; however, past studies have not highlighted the importance and effect of the FIFA Rating and the Career Mode potential of the player.

- ***FIFA Rating:*** The FIFA Rating summarizes all the performance indicators of the player. It summarizes the performance of a player according to his/her position while taking all factors into account.

- ***Career Mode Potential:*** The Career Mode Potential predicts an accurate potential of the player using data-driven estimations. It is the maximum FIFA rating the player is expected to have in the future.

These factors should be utilized when valuing players. This paper contributes to the body of literature on football player valuation by including these two additional factors in the effect analysis.

Methodology

The aim of this research paper is to analyze the factors affecting the market values of football players, how important each factor is, and most importantly to use data-driven estimations to improve accuracy in determining market value.

Independent Variables	**Dependent Variable**
Age	Market Value of Players
KPI	
Rating of the Player in the game FIFA	
Social Media Followers	
Career Mode Potential in the game FIFA	

Hypothesis

Ho: There would be no significant impact of FIFA Rating and Career Mode Potential on the economic value of football players.

Ha: There would be a significant impact of FIFA Rating and Career Mode Potential on the economic value of football players.

Sample:
The sample involves 50 of the most valuable players for every position according to transfermarkt.com (attackers, midfielders, defenders) so a total of 150 players who will be analyzed separately according to their position.

Data analysis strategy:
The data of the 150 football players, including factors like age, brand value, KPI, playing time, FIFA Rating, and Career Mode Potential will be analyzed using correlation analysis and Regression Models. This is done to understand the relationships between the various factors and their correlation to the market value.

Results
The following section focuses on the regression and correlation analysis of FIFA Rating, Age, Goals, Instagram Followers, Career Mode Potential and Market Value.

Table 1: *Correlation Analysis of Fifa Rating, Age, Assists + Goals, Instagram Followers, and Career Mode potential for midfielders (N=50)*

		Fifa Rating	Career Mode Potential	Age	Goals	Instagram Followers (Millions)	Market value (Million euros)
Fifa Rating	Correlation	1	0.57**	0.56**	0.39**	0.59**	0.68**
Career Mode Potential	Correlation	0.57**	1	-0.26	0.27	0.53**	0.78**

Age	Correlation	0.56**	-0.26	1	0.32*	0.26	0.02
Goals	Correlation	0.39**	0.27	0.32*	1	0.52**	0.46**
Instagram Followers (Millions)	Correlation	0.59**	0.53**	0.26	0.52**	1	0.68**
Market value (Million euros)	Correlation	0.68**	0.78**	0.02	0.46**	0.68**	1

Note: p<.05, p<.01***

A positive correlation can be seen between all the variables beside age of a footballer and career mode potential r(48)=-0.26, p>.05. Hence, FIFA rating is positively correlated with Age r(48)= 0.56, p<.001, Goals r(48)= 0.39, p<.01, Instagram Followers r(48)= 0.59, p<.001, and **Market Value r(48)= 0.68, p<.001.** Similarly, Career mode rating is positively correlated with Instagram Followers r(48)= 0.53, p<.001 and **Market Value r(48)= 0.78, p<.001.** Goals are positively correlated with Age r(48)=0.32, p<.05, Instagram Followers r(48)=0.52, p<.001, and **Market Value r(48)=0.46, p<.001**. Lastly, **Instagram Followers and Market Value were positively correlated, r(48)=0.68, p<.001** [Table 1]**.**

Table 2: *Regression Analysis of Fifa Rating, Age, Assists + Goals, Instagram Followers, and Career Mode potential for midfielders (N=50)*

Source	B	SE B	t	p
(Constant)	-337.96	111.13	-3.04	.004
Fifa Rating	5.32	1.62	3.28	.002

Career Mode Potential	0.98	1.91	0.51	.611
Age	-5.3	2.12	-2.5	.016
Goals	1.03	0.49	2.13	.039
Instagram Followers (Millions)	0.43	0.17	2.56	.014
F	39.32			<.001
R2	0.78			

The regression analysis was carried out with market value as the only independent variable. **The predictive effect of FIFA rating, Age, Goals, and Instagram followers was confirmed.** FIFA rating (b=5.32,t(48)=3.28, p<.01), Age (b=-5.3, t(48)=-4.16, p<.05), Goals (b=1.03, t(48)=2.13, p<.05), Instagram Followers (b=0.43, t(48)=2.56, p<.05), significantly predicted market value of football players. **FIFA Rating, Age, Goals, and Instagram followers also explained a significant proportion of variance in market value of players,** $R^2 = .78, F(4) = 39.32, p < .001$ [Table 2].

Table 3: *Correlation Analysis of Fifa Rating, Age, Assists + Goals, Instagram Followers, and Career Mode potential for midfielders (N=50)*

	Fifa Rating	Career Mode Potential	Age	Assists+Goals	Instagram Followers (Millions)	Market Value (Millions)
Fifa Rating	1	0.68**	0.55**	0.09	0.51**	0.57**
Career Mode Potential	0.68**	1	-0.11	0.17	0.56**	0.73**
Age	0.55**	-0.11	1	-0.07	0.11	-0.12

Assists+Goals	0.09	0.17	-0.07	1	0.09	0.36**
Instagram Followers Millions)	0.5**	0.56**	0.11	0.09	1	0.55**
Market Value (Millions)	0.57**	0.73**	-0.12	0.36**	0.55**	1

Note: _p<.05=*, p<.01=**_

A positive correlation has been found between all the variables beside age of a footballer and career mode potential r(48)=-0.11, p>.05, age of a footballer and assists+goals r(48)=-0.07, p>.05, and age and market value of a player r(48)=-0.12, p>.05. Hence, **FIFA rating is positively correlated** with age r(48)= 0.55, p<.001, career mode potential r(48)= 0.68, p<.001, instagram followers r(48)= 0.51, p<.001, and **market value r(48)= 0.57, p<.001.** Similarly, **career mode rating is positively correlated** with instagram followers r(48)= 0.56, p<.001 and **market value r(48)= 0.73, p<.001. Assists+goals are positively correlated** with age r(48)=0.32, p<.05 and **Market Value r(48)=0.36, p<.01**. Lastly, **Instagram followers and market value were positively correlated, r(48)=0.55, p<.001** [Table 3]**.**

Table 4: *Regression Analysis of Fifa Rating, Age, Assists + Goals, Instagram Followers, and Career Mode potential for midfielders   (N=50)*

Model	B	SEB	t	p
(Constant)	-233.9	97.2	-2.41	.02
Fifa Rating	4.09	1.38	2.97	.005

Career Mode Potential	0.65		1.75	0.37	.005
Age	-4.29		1.45	-2.96	.715
Assists+Goals	1.19		0.42	2.81	.007
Instagram Followers (Millions)	0.97		0.46	2.13	.039
R2	.69				
F	19.56				<.001

The regression analysis was carried out with market value as the only independent variable. The predictive effect of FIFA rating, Career Mode Potential Age, Assists+Goals, and Instagram followers was confirmed. FIFA rating (b=4.09,t(48)=2.97, p<.05), Career Mode Potential (b=0.65, t(48)=0.37,p<.01), Age (b=-4.29, t(48)=-2.96, p<.01), Assists + goals (b=1.19, t(48)=2.81, p<.01) , Instagram Followers (b=0.97, t(48)=2.13, p<.05), **significantly predicted market value of football players**. FIFA Rating, Career Mode Potential, Age, Assists+Goals, and Instagram followers also explained a significant proportion of variance in market value of players, ***R^2 = .69, F(5) =19.56 , p < .001 [Table 4].***

Table 5: *Correlation Analysis of Fifa Rating, Age, Instagram Followers, and Career Mode potential for defenders (N=50)*

	Fifa Rating	Career Mode Potential	Age	Instagram followers (millions)	Market value (million euros)
Fifa Rating	1	0.49**	0.63**	0.49**	0.56**
		<.001	<.001	<.001	<.001

Career Mode Potential	0.49**	1	-0.28	0.28	0.65**
Age	0.63**	-0.28*	1	0.24	-0.06
Instagram followers (millions)	0.49**	0.28	0.24	1	0.44**
Market value (million euros)	0.56**	0.65**	-0.06	0.44**	1

Note: p<.05=*, p<.01=**

A positive correlation has been found between all the variables besides age of a footballer and career mode potential r(48)=-0.28, p<.05. Hence, FIFA rating is positively correlated with Age r(48)= 0.63, p<.001, Career Mode Potential r(48)= 0.49, p<.001, Instagram Followers r(48)= 0.49, p<.001, and **Market Value r(48)= 0.56, p<.001.** Similarly, Career mode rating is positively correlated with **Market Value r(48)= 0.65, p<.001.** Lastly, Instagram Followers and **Market Value were positively correlated, r(48)=0.44, p<.001**[Table 5]**.**

Table 6: *Regression Analysis of Fifa Rating, Age, Assists + Goals, Instagram Followers, and Career Mode potential for defenders   (N=50)*

Model	B	SEB	t	p
(Constant)	-156.88	63.23	-2.48	.017
Fifa Rating	3.76	1.07	3.5	.001
Career Mode Potential	-0.27	1.21	-0.22	.826

Age	-3.26	1.1	-2.95	.005
Instagram followers (millions)	0.55	0.36	1.52	.136
F	20.01			<.001
R2	.65			

The regression analysis was carried out with market value as the only independent variable. The predictive effect of FIFA rating and Age was confirmed. FIFA rating (b=3.76,t(48)=3.5, p<.001) and Age (b=-3.26, t(48)=-2.95, p < .01) significantly predicted market value of football players. Fifa Rating and Age also explained a significant proportion of variance in market value of players, ***R^2 = .65, F(4) =20.01, p < .001[Table 6].***

Discussion

The following section presents the findings:

A positive correlation was found between all factors besides age and market value of attackers, and age and career mode potential of midfielders and defenders. Some football players who are over the age of 30 years old can still play although not all of them have good performance. In football, it is considered that there is a productive age, which is until the player still meets the fitness criteria while playing determined by the club. Players who are older will have declining performance and lower market value. This was further supported by a study which analyzed the transfer fees invested over the past ten seasons by clubs worldwide and concluded that football players find their best performance at the age of 22 to 25 years old (Poli et al., 2023).

Moreover, it was found that FIFA rating, career mode potential, goals/assists+goals, and Instagram followers were positively correlated with the market value of the player. Similarly, a study analyzing data from transfermarkt for the 150 most valuable football players found that the goals, assists, club value and FIFA rating are closely related to the valuation process of the footballer. (Majewski, 2021)

FIFA rating gives a good summary of a player's important performance characteristics in a way that can be compared to other players as well. Each player in FIFA has an overall rating as well as six scores for the key stats; Pace, Shooting, Passing, Dribbling, Defending, and Physical.

These stats are combined with a player's international recognition to calculate the player's overall rating.

A study of different performance ratings for 17,000 football players found that a player's international reputation is the second most important feature in determining a player's value, after the player's performance. (Al-Asadi & Tasdemir, 2022). Players who have a good brand value can be fun to watch, popular with fans, popular on social media and a draw for advertisers and sponsors who want to be associated with specific players (Valentini, 2020). All of these off-field factors contribute positively to a team's profit. Consequently, Instagram followers were positively correlated with the market value of the player. To support those findings, a study conducted by Flores which studied players that were part of an English Premier League (EPL) team in the 2015/2016 season found that the number of users that click on the follow button of each players' account, and the number of assists a player has has a positive relationship with the market value.

Therefore, the predictive effect of FIFA rating, Age, Goals/Assists+Goals, and Instagram followers was also confirmed by the regression analysis. FIFA Rating, Age, Goals, and Instagram followers also explained a significant proportion of variance in market value of attackers.

Conclusion

This research paper analyzes the various factors affecting player market value, the relationship between them and their impact on the final economic value. The paper also contributes to the existing body of research by including two new factors for analysis: FIFA rating and Career Mode Potential.

The findings from this paper can be used by football analysts to understand the parameters that should be looked at while analyzing a footballer and use data driven estimations to determine the actual market value of the player. Moreover, the findings can be used by player agents to understand the strengths of their player. Lastly, the findings can be used by footballers themselves to understand what are the key factors that go into being a valuable footballer apart from performance.

Limitations

Since we have only taken a sample of the 50 most valuable players from each position, the results cannot be generalized for the entire set of professional players. Secondly, information about the market values was taken from transfermarkt: users are not aware of all the factors that influence the market values (major ones are known but minor ones tend to be overlooked) so there could be a lack of accuracy. Lastly, quantitative variables were taken as determinants for the market value so other qualitative factors that can have an impact like attitude, team spirit and aggression were not taken into consideration.

The paper can aid in raising awareness about the different factors like age and social media that impact the value of football players aside from their sport performance. Various stakeholders within the industry, spectators, and researchers can understand the economic movements and impact of the subject, leading to more transparency in the valuation. More awareness of the subject could pave the way for controls and accountability from the perspective of investments, salaries, and market value. Further research could potentially use this paper as a jumping off point to look into the economic shifts in the industry caused by changing methods of valuation.

Works Cited

1. Al-Asadi, M. A., & Tasdemir, S. (2022, March 4). Predict the Value of Football Players Using FIFA Video Game Data and Machine Learning Techniques. *IEEE Access*, *10*(1). 10.1109/ACCESS.2022.3154767
2. Bacs, B. A. (2019, September 27). Situatión Repórt óf Európean Club Fóótball (2017 – 2019). *International Journal of Engineering and Management Sciences*, *6*(2). 10.21791/IJEMS.2021.2.6.
3. Franck, E., & Nüesch, S. (2010, December 22). TALENT AND/OR POPULARITY: WHAT DOES IT TAKE TO BE A SUPERSTAR? *Economic Inquiry*, *50*(1), 202-216. https://doi.org/10.1111/j.1465-7295.2010.00360.x
4. Herm, S., & Callsen-Bracker, H.-M. (2014, January 29). When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community. *Sports Management Review*, *17*(4), 484-492. https://doi.org/10.1016/j.smr.2013.12.006
5. Kiefer, S. (2021, November). *The impact of the Euro 2012 on popularity and market value of football players PDF Logo*. EconStor. Retrieved March 4, 2024, from https://hdl.handle.net/10419/67719
6. Majewski, S. (2021, June 30). Football players' brand as a factor in performance rights valuation. *Journal of Physical Education and Sport*, *21*(4), 1751-1760. 10.7752/jpes.2021.04222
7. Metelski, A. (2021, April 30). Factors affecting the value of football players in the transfer market. *Journal of Physical Education and Sport*, *21*(2), 1150-1155. 10.7752/jpes.2021.s2145
8. Muller, O., Simmons, A., & Weinmann, M. (2017, December 1). Beyond crowd judgments: Data-driven estimation of market value in association football. *European Journal of Operational Research*, *263*(2), 611-624. https://doi.org/10.1016/j.ejor.2017.05.005
9. Poli, R., Ravenel, L., & Besson, R. (2023, February 1). *Inflation in the football players' transfer market (2013/14-2022/23)*. CIES Football Observatory. Retrieved March 5, 2024, from https://football-observatory.com/IMG/pdf/mr82en.pdf
10. Rohde, M., & Breuer, C. (2016, March 7). *The Financial Impact of (Foreign) Private Investors on Team Investments and Profits in Professional Football: Empirical Evidence from the Premier League | Rohde | Applied Economics and Finance*. Redfame Publishing. Retrieved May 5, 2024, from https://redfame.com/journal/index.php/aef/article/view/1366
11. Rohde, M., & Breuer, C. (2016, June 2). Europe's Elite Football: Financial Growth, Sporting Success, Transfer Investment, and Private Majority Investors. *International Journal of Financial Studies*, *4*(2). 10.1055/s-2006-924294

12. Salvo, V. D., Baron, R., Tschan, H., Montero, F. J. C., Bachl, N., & Pigozzi, F. (2006, October). Performance Characteristics According to Playing Position in Elite Soccer. *International Journal of Sports Medicine*, *28*(3), 222-227. 10.1055/s-2006-924294
13. Valentini, K. (2020, May). *TRANSFER PRICING: AN ANALYSIS OF THE IMPACT OF PLAYER BRAND VALUE ON TRANSFER FEES IN EUROPEAN FOOTBALL*. Scholarly Commons. Retrieved March 4, 2024, from https://repository.upenn.edu/server/api/core/bitstreams/064ddf9f-a14f-49a5-91c5-96cbcbbc5fb5/content

Finerenone: Pharmacokinetics, Pharmacodynamics, and Applications By Shubhay Mishra

Abstract

Finerenone, a nonsteroidal mineralocorticoid receptor (MR) antagonist, is a novel treatment option for chronic kidney disease (CKD) in patients with type 2 diabetes. It metabolizes into 14 inactive metabolites, resulting in a bioavailability of 43.5%. By inhibiting aldosterone binding, finerenone reduces MR overactivation, outperforming spironolactone and eplerenone in potency and selectivity. Approved for CKD in type 2 diabetes, finerenone's potential extends to non-diabetic CKD and MR-related conditions. Ongoing research explores its combined use with SGLT2 inhibitors to enhance efficacy and address hyperkalemia. This review examines its pharmacokinetics, pharmacodynamics, and clinical applications.

Keywords: Finerenone, Pharmacokinetics, Pharmacodynamics

Definitions

Extraction Ratio: The fraction of a drug that permeates into a given organ.
Hypervolemia: A condition where the liquid portion of blood plasma is too high.
MR IC50: A quantitative measure indicating the concentration of drug (nM) that is required to inhibit the binding of mineralocorticoid receptors by 0.5 nM aldosterone by 50%.
Bioavailability: The concentration of a drug that remains unmetabolized.

Methods

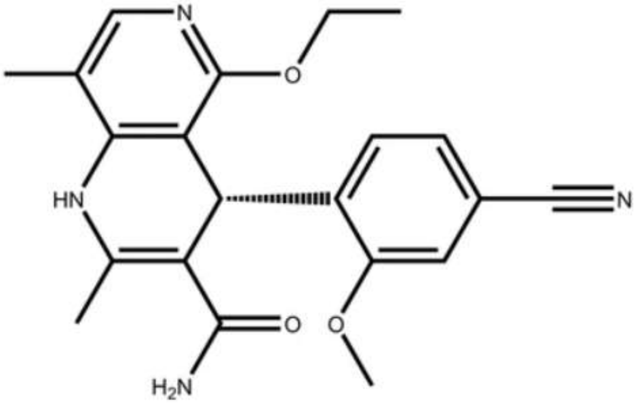
The databases ScienceDirect and ProQuest were used to obtain information. Articles were searched for by entering keywords, including but not limited to finerenone, pharmacokinetics, pharmacodynamics, metabolites, mineralocorticoid receptor antagonists, and aldosterone. To ensure that the information was up to date, the range of publication years of the articles was limited to 2014 to 2024. In terms of inclusion criteria, studies that performed analyses on the FIGARO-DKD and FIDELIO-DKD clinical trials, as well as those that provided detailed pharmacokinetic and pharmacodynamic profiles of finerenone, were included. Emphasis was put on research sponsored by researchers affiliated with Bayer Pharmaceuticals. Studies that did not focus on finerenone or were not peer-reviewed were excluded from the review. The selected articles were then analyzed to extract relevant data on the drug's mechanism of action, therapeutic effects, adverse effects, and comparison with other mineralocorticoid receptor antagonists. The gathered information was synthesized by topic: Pharmacokinetics, pharmacodynamics, and applications.

Background

Mineralocorticoid receptors (MRs) are crucial in regulating blood pressure and maintaining electrolyte and fluid balance. (Gomez-Sanchez & Gomez-Sanchez, 2014). They are expressed in cells across the body, especially in the kidneys. These include epithelial cells, smooth muscle cells, and fibroblasts (González-Juanatey et al., 2023). Overactivation of MRs can lead to adverse effects. In particular, the binding of aldosterone, an MR agonist, leads to the transcription of certain genes, some negative, such as pro-inflammatory and pro-fibrotic genes (National Center for Biotechnology Information [NCBI], 2024). Diabetic chronic kidney disease (CKD) symptomizes as lowered glomerular filtration rate, albuminuria, and hypertension (Anders et al., 2018). Treatments for this disease, namely angiotensin-converting enzyme (ACE) inhibitors, can result in hypersecretion of aldosterone.

Finerenone is the first approved nonsteroidal MR antagonist which has been shown to alleviate the above symptoms in clinical trials (Kintscher & Edelmann, 2023). It is even more potent and selective toward MRs as compared to other MR antagonists, such as spironolactone or eplerenone (Yang & Young, 2016). Previous MR antagonists, being steroidal, can have antiandrogenic effects. Finerenone, however, avoids this issue (Rico-Mesa et al., 2020). Marketed under the name Kerendia and developed by Bayer HealthCare Pharmaceuticals, finerenone was approved in July 2021, primarily for treating CKD. (Frampton, 2021). This literature review aims to provide an overview of the pharmacokinetic and pharmacodynamic aspects of finerenone and its applications.

Figure 1. Finerenone



1 Pharmacokinetics of Finerenone

1.1 Absorption and Distribution

Finerenone is orally administered, usually as a tablet (Heinig & Eissing, 2023). There is no clinically significant effect on finerenone area-under-the-curve concentration following administration with high-calorie food (Food and Drug Administration [FDA], 2021). It is completely first-pass metabolized in the gut wall and liver. Preclinical trials revealed a quantitatively larger extraction ratio in the gut wall of 0.425 as compared to the liver, with an extraction ratio of 0.244, suggesting metabolism in the gut wall was more important to the overall bioavailability (Heinig et al., 2018). The plasma half-life of finerenone is roughly 2 hours (Yang et al., 2016).

1.2 Metabolism

Finerenone undergoes extensive metabolic oxidative biotransformation into 14 metabolites, primarily by the Cytochrome P450 3A4 (CYP3A4) isoenzymes (~90%) and the Cytochrome P450 2C8 (CYP2C8) isoenzymes (~10%) (Heinig et al. 2018). Notably, none of these metabolites possess pharmacological activity (Singh et al., 2022). Firstly, finerenone is aromatized to metabolite M1 by CYP3A4 and CYP2C8 (Gerisch et al., 2018), and later hydroxylated to the metabolite M2 by CYP3A4. From here, two paths emerge: (1) M2 can be further oxidized by CYP3A4 to form M3; (2) M2 can be epoxidized by CYP3A4 and hydrolyzed by CYP2C8 to form M4. M4 is hydroxylated again by CYP3A4 to form M5 and oxidized again by CYP3A4 to form M8. Alternatively, finerenone can be hydroxylated by CYP2C8 to the M7 metabolite. The M10 metabolite is produced when finerenone undergoes demethylation, oxidation, and ring-opening polymerization. The M13 metabolite forms through the de-ethylation of finerenone by CYP1A1. The metabolic pathway for the M14 metabolite is not yet fully understood, but it involves CYP2C8 and CYP3A4 (NCBI, 2024). Approximately 56.5% of the original dose is biotransformed (Heinig et al., 2018).

Though finerenone possesses no active metabolites, they are important to note for several reasons. These inactive metabolites are non-toxic, bolstering the clinical application of finerenone. Furthermore, the drug-drug interactions of these metabolites with other substances have not fully been investigated, which may be an area of future study. Also, this eliminates the possibility of finerenone as a prodrug, i.e., a drug that becomes pharmacologically active after being metabolized, unlike other MR antagonists such as spironolactone.

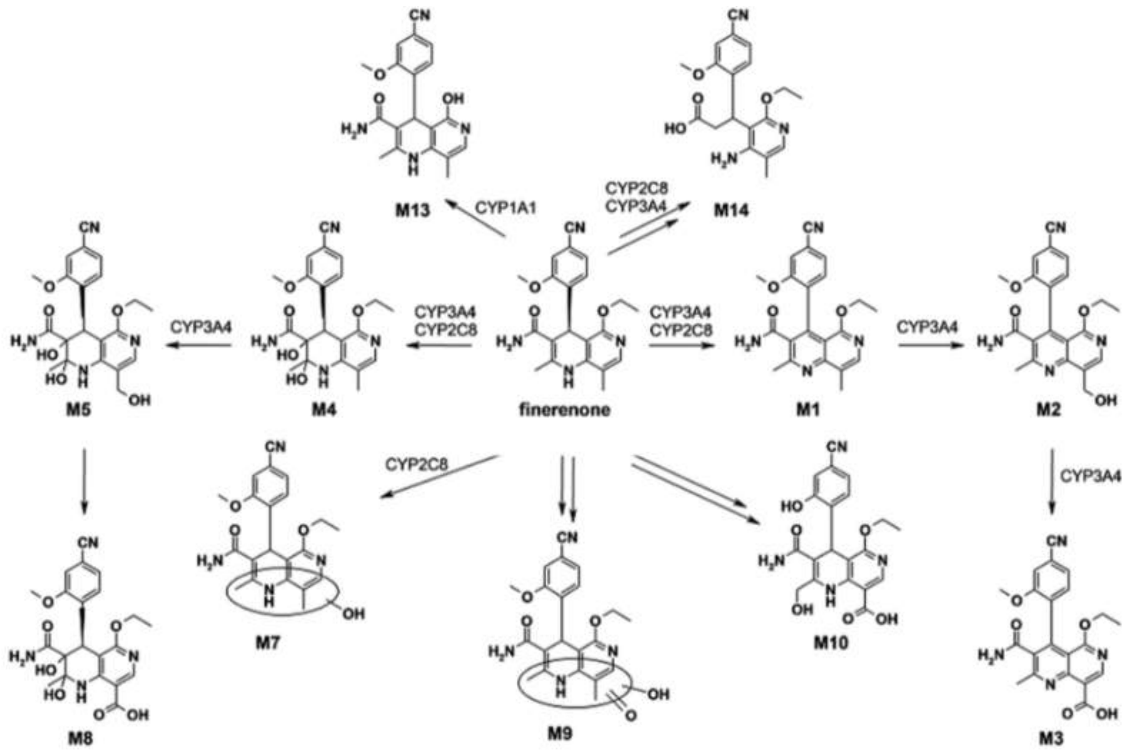
1.3 Excretion

Finerenone typically has an elimination half-life of between 2 and 3 hours in the dose range of up to 20 mg (Heinig & Eissing, 2023). Approximately 80% is excreted in urine, while the remaining 20% is excreted in feces (Lerma & Wilson, 2022). Approximately 1% of finerenone remains unchanged and is excreted in feces (Gerisch et al., 2018).

(Gerisch et al., 2023) conducted a metabolic profile study, revealing that the M2, M3, M4, and M5 metabolites made up the largest portion of excreted metabolites. Among these, M3 was the most prevalent, accounting for 47.8%.

1.4 Other Pharmacokinetic Features

The pharmacokinetics of Finerenone are unaffected by age, sex, race, or weight. They are also unaffected by renal impairment (FDA, 2021). Although a study found that hepatitis impairment had no significant effect on the pharmacokinetics of Finerenone (Heinig et al., 2019), administration in patients with severe hepatitis impairment should be avoided (Frampton, 2021).



Metabolites of finerenone. Adapted from "Biotransformation of Finerenone, a Novel Nonsteroidal Mineralocorticoid Receptor Antagonist, in Dogs, Rats, and Humans, In Vivo and In Vitro," by M. Gerisch, Drug Metabolism and Disposition, *46*(11). Copyright 2018 by The Authors. Adapted with permission.

2 Pharmacodynamics of Finerenone

2.1 Physiological Activity

Aldosterone is a hormone that binds to mineralocorticoid receptors, especially in renal cells (MRs). It is normally secreted in response to hyponatremia, hypernatremia, or hypovolemia (Khan, 2021). Upon binding, aldosterone stimulates transcription and insertion of sodium channels, sodium-potassium pumps, and potassium channels into the cell membrane to restore homeostasis (Khan, 2021). In CKD, aldosterone is hypersecreted (Verma et al., 2022). Finerenone binds to mineralocorticoid receptors (MRs) with high potency and selectivity, inhibiting MR-mediated sodium reabsorption and overactivation and blocking aldosterone. Not only does this reduce the transcription of channel proteins in general, but it also decreases MR expression (FDA, 2021). This slows the decline of the glomerular filtration rate and prevents fibrosis (Sridhar et al., 2021).

2.2 Selectivity and Potency

Finerenone displays a relatively low MR IC50 value of 18 nM, which is especially significant when put in the context of previously used MR antagonists, namely spironolactone and eplerenone, which demonstrate IC50 values of 24 nM and 990 nM, respectively (Yang & Young, 2016). It is more than 500 times more selective for MRs than other steroid receptors (Frampton, 2021). This bolsters the applicability compared to other MR antagonists, especially spironolactone, which can act as antiandrogens (Khan, 2021). Pharmacometric models have revealed body weight and original eGFR as the covariates influencing finerenone pharmacodynamics, and therefore effectiveness, aside from initial dose volume (Eissing et al., 2023). Measured glomerular filtration rate is therefore used to determine starting dose (Heinig & Eissing, 2023).

3 Applications

3.1 Current Applications

The FIGARO-DKD study revealed finenerone's application in reducing CKD progression and CV events in diabetic patients (Frampton, 2021). Aldosterone, which is excessively high in patients with cardiac disease or renal disorders, can cause excessive water retention or hypervolemia due to its action at the MRs, as specified above (Khan, 2021). By preventing aldosterone from binding, finerenone effectively is a diuretic. It is classified as a potassium-sparing diuretic, as is spironolactone. Ergo, finerenone can also treat albuminuria symptoms (Haller et al., 2016).

Finerenone was initially developed as a treatment for diastolic heart failure. Although promising results were shown, the investigation was eventually converted to prevent renal and cardiovascular complications in diabetic patients due to reducing inflammation and fibrosis (Azizi et al., 2019). Further studies may be useful to identify whether finerenone can be further developed to address heart failure. Being an MR antagonist, it is likely that finerenone can be used as a modest antihypertensive. Hypervolemia caused by excess aldosterone binding leads to higher blood flow and pressure, which seriatim causes hypertension (Khan, 2021). This would be counteracted by finerenone.

3.2 Future Prospects

Research is ongoing to explore further therapeutic uses of finerenone. One preclinical study, performed in induced-hypertension rats, discovered a significant reduction in albuminuria when finerenone was taken in conjunction with the sodium-glucose cotransporter-2 (SGLT2) inhibitor empagliflozin (Kolkhof et al., 2021). This combination has also been shown to alleviate hyperkalemia symptoms, the only identified clinical risk of finerenone. Currently, a large-scale study (n=807), known as CONFIDENCE, is being conducted, examining the relationship between concurrent finerenone and empagliflozin treatment on the change in eGFR, a proxy for CKD progression (Theodorakopoulou & Sarafidis, 2023). The study began in June 2022 and is estimated to end in February 2025 (Study Details, 2023).

The main clinical trials for finerenone, FIGARO-DKD and FIDELIO-DKD, focused on the treatment potential of finerenone in diabetic CKD. However, some believe it may have applications in the treatment of non-diabetic CKD due to its unique pharmacological profile (Theodorakopoulou & Sarafidis, 2023). A recent clinical trial (n=18) displayed that finerenone significantly reduced albuminuria symptoms in non-diabetic CKD patients in an eight-week period. This effect was increased with the addition of the SGLT2 dapagliflozin. The researchers also noted a decrease in systolic blood pressure of 10 mmHg (Mårup et al., 2023). Though these results are promising, further clinical trials in patients with non-diabetic CKD with larger sample sizes must be performed.

The only identified clinical risk of finerenone is hyperkalemia (Heinig & Gerisch, 2018). A post-hoc analysis of the FIDELIO-DKD trial revealed that approximately 21.4% of the patients experienced mild hyperkalemia, while 4.5% of the patients experienced moderate hyperkalemia. Since the patients in the study already suffered from predispositions to hyperkalemia, as indicated by the fact that 9.2% of the patients in the placebo arm exhibited mild hyperkalemia symptoms, however, use of SGLT2 inhibitors minimized the risk (Agarwal et al., 2024). Notably, finerenone tends to cause less hyperkalemia than other MR antagonists because it accumulates in the kidneys to a lesser extent due to its short half-life (Kim et al., 2023).

Conclusion

Finerenone represents a significant advancement in the treatment of diabetic CKD. Its pharmacological profile is characterized by high bioavailability, high selectivity and potency, and prospective versatility underscore its therapeutic potential. Current applications have shown promising results in reducing CKD progression and cardiovascular events, and ongoing research suggests potential benefits in non-diabetic CKD and in combination therapies to mitigate hyperkalemia. As research progresses, finerenone may become a cornerstone in the management of CKD and other MR-related conditions.

Acknowledgements

Thanks to Professor Sanjay Mishra at the University of Pittsburgh for mentorship and peer review.

Works Cited

Agarwal, R., Ruilope, L. M., Ruiz-Hurtado, G., Haller, H., Schmieder, R. E., Anker, S. D., Filippatos, G., Pitt, B., Rossing, P., Lambelet, M., Nowack, C., Kolkhof, P., Joseph, A., & Bakris, G. L. (2023). Effect of finerenone on ambulatory blood pressure in chronic kidney disease in type 2 diabetes. *Journal of Hypertension*, *41*(2), 295–302.

Anders, H.-J., Huber, T. B., Isermann, B., & Schiffer, M. (2018). CKD in diabetes: Diabetic kidney disease versus nondiabetic kidney disease. *Nature Reviews Nephrology*, *14*(6), 361-377. https://doi.org/10.1038/s41581-018-0001-y

Azizi, M., Rossignol, P., & Hulot, J.-S. (2019). Emerging drug classes and their potential use in hypertension. *Hypertension*, *74*(5), 1075-1083. https://doi.org/10.1161/hypertensionaha.119.12676

Chemical structures of finerenone (BAY 94-8862) and metabolites identified in vitro and in vivo as well as involved human P450 isoforms [Illustration]. (n.d.). https://dmd.aspetjournals.org/content/dmd/46/11/1546/ F3.large.jpg?width=800&height=600&carousel=1

Eissing, T., Goulooze, S. C., van den Berg, P., van Noort, M., Ruppert, M., Snelder, N., Garmann, D., Lippert, J., Heinig, R., Brinker, M., & Heerspink, H. J. L. (2023). Pharmacokinetics and pharmacodynamics of finerenone in patients with chronic kidney disease and type 2 diabetes: Insights based on FIGARO‐DKD and FIDELIO‐DKD. *Diabetes, Obesity and Metabolism*, *26*(3), 924-936. https://doi.org/10.1111/dom.15387

Filippatos, G, Anker, S, Pitt, B. et al. Finerenone and Heart Failure Outcomes by Kidney Function/Albuminuria in Chronic Kidney Disease and Diabetes. J Am Coll Cardiol HF. 2022 Nov, 10 (11) 860–870. https://doi.org/10.1016/j.jchf.2022.07.013

Food and Drug Administration. KERENDIA (finerenone) prescribing information. 2021. Retrieved July 2, 2024 from https://www. accessdata.fda.gov/drugsatfda_docs/label/2021/215341s000lbl.pdf

Frampton, J. E. (2021). Finerenone: First approval. *Drugs*, *81*(15), 1787-1794. https://doi.org/10.1007/s40265-021-01599-7

Gerisch, M., Heinig, R., Engelen, A., Lang, D., Kolkhof, P., Radtke, M., Platzek, J., Lovis, K., Rohde, G., & Schwarz, T. (2018). Biotransformation of finerenone, a novel nonsteroidal mineralocorticoid receptor antagonist, in dogs, rats, and humans, in vivo and in vitro. *Drug Metabolism and Disposition*, *46*(11), 1546-1555. https://doi.org/10.1124/dmd.118.083337

Gomez-Sanchez, E., & Gomez-Sanchez, C. E. (2014). The multifaceted mineralocorticoid receptor. *Comprehensive Physiology*, *4*(3). https://doi.org/10.1002/cphy.c130044

González-Juanatey, J. R., Górriz, J. L., Ortiz, A., Valle, A., Soler, M. J., & Facila, L. (2023). Cardiorenal benefits of finerenone: Protecting kidney and heart. *Annals of Medicine*, *55*(1), 502-513. https://doi.org/10.1080/07853890.2023.2171110

Haller, H., Bertram, A., Stahl, K., & Menne, J. (2016). Finerenone: A new mineralocorticoid receptor antagonist without hyperkalemia: An opportunity in patients with CKD? *Current Hypertension Reports*, *18*(41). https://doi.org/10.1007/s11906-016-0649-2

Heinig, R., & Eissing, T. (2023). The pharmacokinetics of the nonsteroidal mineralocorticoid receptor antagonist finerenone. *Clinical Pharmacokinetics*, *62*(12), 1673-1693. https://doi.org/10.1007/s40262-023-01312-9

Heinig, R., Gerisch, M., Engelen, A., Nagelschmitz, J., & Loewen, S. (2018). Pharmacokinetics of the novel, selective, non-steroidal mineralocorticoid receptor antagonist finerenone in healthy volunteers: Results from an absolute bioavailability study and drug–drug interaction studies in vitro and in vivo. *European Journal of Drug Metabolism and Pharmacokinetics*, *43*(6), 715-727. https://doi.org/10.1007/s13318-018-0483-9

Heinig, R., Lambelet, M., Nagelschmitz, J., Alatrach, A., & Halabi, A. (2019). Pharmacokinetics of the novel nonsteroidal mineralocorticoid receptor antagonist finerenone (BAY 94-8862) in individuals with mild or moderate hepatic impairment. *European Journal of Drug Metabolism and Pharmacokinetics*, *44*. https://doi.org/10.1007/s13318-019-00547-x

Khan, J. A. (2021). A literature review of the pharmacokinetics, pharmacodynamics, and possible uses of spironolactone. *Journal of Natural Sciences*, *2*(1). https://doi.org/10.33137/jns.v2i1.35920

Kim, D.-L., Lee, S.-E., & Kim, N. H. (2023). Renal protection of mineralocorticoid receptor antagonist, finerenone, in diabetic kidney disease. *Endocrinology, Diabetes & Diabetes*, *38*(1), 43-55. https://doi.org/10.3803/EnM.2022.1629

Kintscher, U., & Edelmann, F. (2023). The non-steroidal mineralocorticoid receptor antagonist finerenone and heart failure with preserved ejection fraction. *Cardiovascular Diabetology*, *22*(1). https://doi.org/10.1186/s12933-023-01899-0

Kolkhof, P., Hartmann, E., Freyberger, A., Pavkovic, M., Mathar, I., Sandner, P., Droebner, K., Joseph, A., Hüser, J., & Eitner, F. (2021). Effects of finerenone combined with empagliflozin in a model of hypertension-induced end-organ damage. *American Journal of Nephrology*, *52*(8), 642-652. https://doi.org/10.1159/000516213

Lerma, E. V., & Wilson, D. J. (2022). Finerenone: A mineralocorticoid receptor antagonist for the treatment of chronic kidney disease associated with type 2 diabetes. *Expert Review of Clinical Pharmacology*, *15*(5), 501-513. https://doi.org/10.1080/17512433.2022.2094770

Mårup, F. H., Thomsen, M. B., & Birn, H. (2023). Additive effects of dapagliflozin and finerenone on albuminuria in non-diabetic ckd: An open-label randomized clinical trial. *Clinical Kidney Journal*, *17*(1). https://doi.org/10.1093/ckj/sfad249

National Center for Biotechnology Information (2024). PubChem Compound Summary for CID 60150535, Finerenone. Retrieved July 2, 2024 from https://pubchem.ncbi.nlm.nih.gov/compound/Finerenone.

Palanisamy, S., Funes Hernandez, M., Chang, T. I., & Mahaffey, K. W. (2022). Cardiovascular and renal outcomes with finerenone, a selective mineralocorticoid receptor antagonist. *Cardiology and Therapy*, *11*(3), 337-354. https://doi.org/10.1007/s40119-022-00269-3

Rico-Mesa, J. S., White, A., Ahmadian-Tehrani, A., & Anderson, A. S. (2020). Mineralocorticoid receptor antagonists: A comprehensive review of finerenone. *Current Cardiology Reports*, *22*(140). https://doi.org/10.1007/s11886-020-01399-7

Singh, A. K., Singh, A., Singh, R., & Misra, A. (2022). Finerenone in diabetic kidney disease: A systematic review and critical appraisal. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *16*(10), 102638. https://doi.org/10.1016/j.dsx.2022.102638

Sridhar, V. S., Liu, H., & Cherney, D. Z. (2021). Finerenone—A new frontier in renin-angiotensin-aldosterone system inhibition in diabetic kidney disease. *American Journal of Kidney Diseases*, *78*(2), 309-311. https://doi.org/10.1053/j.ajkd.2021.02.324

Study details: A study to learn how well the treatment combination of finerenone and empagliflozin works and how safe it is compared to each treatment alone in adult participants with long-term kidney disease (chronic kidney disease) and type 2 diabetes (CONFIDENCE). (2023, July 3). Bayer Pharmaceuticals. https://clinicaltrials.gov/study/NCT05254002

Theodorakopoulou, M. P., & Sarafidis, P. (2023). SGLT2 inhibitors and finerenone in non-diabetic CKD: A step into the (near) future? *Clinical Kidney Journal*, *17*(1). https://doi.org/10.1093/ckj/sfad272

Verma, A., Vaidya, A., Subudhi, S., & Waikar, S. S. (2022). Aldosterone in chronic kidney disease and renal outcomes. *European Heart Journal*, *43*(38), 3781-3791. https://doi.org/10.1093/eurheartj/ehac352

Yang, J., & Young, M. J. (2016). Mineralocorticoid receptor antagonists — pharmacodynamics and pharmacokinetic differences. *Current Opinion in Pharmacology*, *27*, 78-85. https://doi.org/10.1016/j.coph.2016.02.005

Yang, P., Huang, T., & Xu, G. (2016). The novel mineralocorticoid receptor antagonist finerenone in diabetic kidney disease: Progress and challenges. *Metabolism*, *65*(9), 1342-1349. https://doi.org/10.1016/j.metabol.2016.06.001

Zhang, M.-Z., Bao, W., Zheng, Q.-Y., Wang, Y.-H., & Sun, L.-Y. (2022). Efficacy and safety of finerenone in chronic kidney disease: A systematic review and meta-analysis of randomized clinical trials. *Frontiers in Pharmacology*, *13*. https://doi.org/10.3389/fphar.2022.819327

What the COVID-19 Pandemic Taught Us About Economic Status, Human Development, and Emergency Response Plans By Yakin Kim

Abstract

For many years it was believed that a country's wealth, level of human development, and healthcare systems were indicators of how well the country could handle an emergency like a pandemic. However, the COVID-19 pandemic revealed that those parameters alone were not enough to estimate preparedness. For example, before COVID-19, the United States was one of the wealthiest countries in the world and held high positions on the Global Health Security (GHS) Index and the Human Development Index (HDI). Yet, the country's COVID-19 response was slow and ineffective. South Korea, another high-income country, was ranked lower on the same indices, but implemented a more effective response plan. Even more shocking was the case of Vietnam; despite being a low-middle income country ranked far below both countries on the same indices, Vietnam implemented a COVID-19 response plan that was just as effective as South Korea's and more effective than the United States. The purpose of this paper is to understand why the outcomes for these three countries were so different and make suggestions to improve response plans around the globe. Research methods include an analysis of existing literature and statistics from the GHS Index and HDI. Results show that income and advanced medical technology alone are not the most important when implementing a pandemic response plan. Leadership, governance, integration of the healthcare system, and communication are additional factors that can significantly impact the effectiveness of the response plan.

Introduction

The sudden emergence of COVID-19 tested every country's emergency response plan. Some countries like South Korea (henceforth Korea) and Vietnam, which were among the first countries with COVID-19 cases after China, were lauded for their initial government response and safety measures. Other countries, like the United States, suffered from a disorganized response that allowed the virus to rapidly circulate through the country. The difference between Korea's and the US responses are shocking because both are high-income countries with an abundance of resources. Why were the outcomes so different? Additionally, how did Vietnam, a low-middle income country, manage such a strong initial response where wealthier countries failed? Using South Korea, the US, and Vietnam as case studies, this paper seeks to address these questions by analyzing factors that modulated each country's COVID-19 response. While culture also played an important role in the effectiveness of the government's COVID-19 response, cultural differences fall outside of the scope of the current study. The first section includes background information about the major factors that play a role in emergency responses according to recent scholarship. The second section compares the cases of the two high-income countries, Korea and the US, and the third section examines Vietnam's startling success despite its economic standing. Finally, the fourth section contains policy suggestions to improve outcomes for lower income countries throughout Southeast Asia similar to Vietnam.

Factors that Influence Emergency Response

One may wonder why a country's social and economic standing plays a role in its emergency response and prevention. First, disasters, whether natural, medical, or manmade, can cause a great deal of damage which require financial resources to repair. The better off a country is financially, the more quickly and comprehensively they can recover from an unexpected disaster. According to the World Bank, investing in measures to manage risks of disasters before they happen "substantially outweigh the costs and are generally two to ten times higher," ("Economics for Disaster Prevention and Preparedness in Europe"). Thus, it is financially savvy for countries to invest in disaster prevention and preparedness rather than paying for the aftermath. It goes without saying, then, that countries with more money to invest in disaster prevention and preparedness should fare better than countries with fewer financial resources. They can spend more on prevention and can afford to spend more on the aftermath if necessary. Secondly, a country's social development, which includes factors such as "literacy or education level, infant mortality rate, agricultural productivity, life expectancy, population growth rate, health, poverty level, employment level, gender equality and labour force participation rate" can also play a role in emergency responses because these factors dictate the number and quality of workers who can be mobilized to deal with the emergency, and how well the general public can comply with emergency measures (Yadav & Badar). This is why socially developed, high-income countries like the United States and Korea were expected to handle emergencies like COVID-19 well relative to a country like Vietnam.

As much as resources matter when it comes to handling emergencies, however, scholars suggest that other factors matter as well. Amul et al. writes that leadership and governance are responsible for creating and executing emergency response policy as well as allocating resources. Even if a country has the funds to help those affected by an emergency, lack of policy, corrupt, ineffective leaders, or overly complicated processes can prevent funds from being used. Another important factor for emergency response is public communication because it "reinforces compliance with transmission‑reducing measures." In other words, clear communication between the government and other stakeholders is necessary to make sure that the population is following safety precautions and procedures. Finally, the state of the healthcare system and technological integration are both tantamount for guaranteeing the victims and patients to high-quality medical care that will prevent contraction of disease and/or loss of life (Amul et al.) During COVID-19, this factor played a large role in each country's ability to respond to the emergency because the disease placed unprecedented strain on the medical system. The most flexible and accessible systems, and of course, appropriately staffed and funded, were the ones that pivoted to the needs of the population best. With these factors in mind, let us examine how two countries, despite being high-income, suffered different effects of COVID-19.

Differences between the US and Korean Responses

Despite differences in GDP, both Korea and the United States are categorized as "high income OECD countries" by the World Bank (Bajpai and Kvilhaug). Because of this categorization, it is reasonable to expect each country has resources to invest in preventative and countermeasures of emergencies. As mentioned in the previous section, several other statistics indicate factors that may impact the expected effectiveness of the COVID-19 emergency response. For example, the Global Health Security (GHS) index evaluates how well 195 countries are prepared for epidemics and pandemics using 6 indicators: health system, compliance with international norms, prevention, detection and reporting, rapid response, and risk environment (Isaac et al.) According to the latest evaluation, the US is #1 with a score of 75.9/100, and Korea ranks #9 with a score of 65.4 (GHS Index 2021). The second important statistic is provided by the Human Development Index, a composite index based on measures like average life expectancy, education levels, standard of living, etc. Recall that these factors, which comprise a country's social development, can be important in emergency response plans because they indicate the size and quality of the dispatchable workforce. In 2019, the US was #15 on the HDI and Korea sat at #22 (HDI 2019). Thus, according to both the GHS and HDI, the US was expected to handle the COVID-19 pandemic more effectively than Korea.

However, this is not what happened. Korea's response was far more effective than that of the US even though indices indicated they were "less prepared." Scholarship attributes Korea's success to their coordinated all-in government response and their effective communication with the public. First, when COVID-19 was initially reported in Wuhan, China, the Korean Center for Disease Control and Prevention coordinated with all ministries, regional governments, and city governments to create a task force that was sent to Wuhan to study the outbreak. Second, the government also started a screening program and began extensive testing before the first confirmed case that extended to even those who were asymptomatic; this "enabled them to implement early isolation and quarantine and minimized the chance for community transmission" (Isaac et al.) Third, the integration of technology in the healthcare and government systems enabled them to fill in the gaps of patient-reported contact tracing. Using CCTVs, phone call logs, and credit card transaction logs, authorities were able to identify and contact citizens who may have come in contact with a positive COVID-19 case even if the patient themselves did not report everything in full. Finally, the government's transparent and wide-spread communication with the public about new policies and guidelines minimized confusion and panic among the populace and promoted compliance with safety measures. Thus, even though Korea had fewer resources and is considered less developed than the US, the country limited its caseload to 10,752 as of April 28, 2020. Compare this to the US which had a caseload of 1,010,507 by the same date despite having a lower population density (Isaac et al.)

On the US side, the reaction to COVID-19 was slow and ineffective. COVID-19 was not declared a public health emergency until the end of January 2020, a full two months after the first instance of the disease in Wuhan, China. Foreign nationals from China, Iran, and some European countries were banned from entering the country, and only citizens returning home from those regions were required to undergo testing. Testing centers and channels for mask distribution were only established by March 2020. Second, despite its top placement on the GHS index, the decentralized US healthcare system was a major impediment to emergency response, containment, and treatment of the disease. Each state had different containment and testing policies and enforcement standards also differed greatly from state to state. The third issue was leadership and governance. Because of the federal government's limited power and internal disagreement about the best way to move forward, it also took an exorbitant amount of time for the federal government to give guidelines and aid to the state governments. When guidelines and suggestions did arrive, some were contradictory or false, so there was confusion on an administrative level that amplified the panic of the general public and caused more harm than good. For example, though health officials in the country generally supported the use of masks to reduce transmission, President Trump openly questioned the masks' effectiveness of public television. The president's statement led to state government officials also questioning or discouraging mask usage in areas under their jurisdiction, such as the Lt. Governor of North Carolina (Song & Choi). In conclusion, even though the US had more resources and the ability to handle a pandemic well according to several indices, the country failed to implement a successful emergency response plan due to poor leadership, governance, and communication as well as a decentralized, inflexible healthcare system.

The Unexpected Success of Vietnam

On all counts, Vietnam should have been the least effective at handling the COVID-19 pandemic. The World Bank classifies it as a low-middle income country, thus the country had fewer resources than the US and Korea to allocate towards their pandemic emergency response plan ("The World Bank in Vietnam"). It is no surprise, then, that it has a GHS index score of 42.9/100, putting it at #65 out of a total of 195 countries. Note that this position in the ranking is far below the placement of the US and Korea (GHS Index 2021). When compared to other Southeast Asian countries, Vietnam (2.60) has just as many hospital beds per 1,000 people as Singapore (2.40), a high-income country, but fewer physicians (0.83 compared to Singapore's 2.29) (Amul et al.) Finally, according to the HDI in 2018 Vietnam ranked #118 out of 189 countries and territories, placing it solidly in the second-highest development category and mere hundredths of a point shy of the highest development category (UNDP). Despite these social and economic factors, Vietnam managed to respond more effectively to the crisis than the US and just as well as Korea.

Due to its geographical proximity and strong connection with China, Vietnam was one of the first countries to be affected by the spread of COVID-19. Like Korea, the Vietnamese government acted quickly. As soon as the first case was reported in Wuhan, the government placed restrictions on travelers from the area, and by January 2020, Vietnam had established an COVID‐19 Response Plan and Technical Treatment and Care Guidelines as well as assembled a taskforce group to coordinate actions on different levels of governance (Ha et al.) After the first confirmed cases, no visas were issued to Chinese citizens and flights to China were suspended, and a mandatory 14-day quarantine was implemented for arrivals to the country. The government began its emergency measures to ensure later access to testing and vaccinations; they deployed military and local government officials to provide free meals, testing, and other essentials to those in need (Turner et al.) Lockdowns and face masks became also mandatory to prevent the spread of the virus. The quick and decisive actions resulted in low rates of COVID-19 during the first wave of the pandemic–with a caseload of only 1,850 by the end of February 2021 (Turner et al.)

Despite not having as many resources as the US and Korea, Vietnam has several other factors that contributed to the success of its emergency response. First, it has strong leadership and governance; similar to Korea, Vietnam employed an all-in government approach. Because Vietnam is a one-party country and a strong central government, it was easier for all government actors on multiple levels to work together towards a common goal. In fact, the Prime Minister Nguyen Xuan Phuc declared war on COVID when he said that his administration was "fighting the epidemic is like fighting against the enemy" (Amul et al.) Second, Vietnam's healthcare system was prepared for a pandemic due to their previous experience with SARS. After SARS ravaged the country, Vietnam has invested in "a public health emergency operations center, a public health surveillance system and a rapidly implemented epidemic management plan" (Amul et al.) Thus, all hospitals across the country had uniform protocols, safety measures, and protective equipment, and treatment for patients was covered by the national healthcare for Vietnamese citizens. Because of this kind of preparation, Vietnam's healthcare system was able to more reliably handle the public health crisis of COVID-19. The third factor that ensured their success was effective communication. There was clear and consistent communication about COVID-19 on all available media outlets including Facebook, Zalo (a local app) and state-controlled media. The comprehensive communication plan raised awareness about the virus across the country and informed the public about symptoms, safety measures, and what to do in case of infection. This boosted the public's compliance and ultimately helped ensure their wellbeing (Turner et al.)

Suggestions and Conclusion

The three case studies have shown that it is not about the number of resources but how the resources are used. Despite being high-income countries, the US and Korea experienced different COVID-19 trajectories during the pandemic due to different styles of leadership, governance, and healthcare, which ultimately led to different emergency response plans and implementation. Certain metrics predicted that Korea was less prepared to handle a pandemic than the US, yet the reality showed that the opposite was true. The same could be said of Vietnam; as a low-middle income country, it should have struggled to handle the pandemic more than Korea or the US. Yet, the government contained the spread of the virus just as well as Korea and better than the US. This extraordinary turn of events begs the question: what can similar countries learn from Vietnam's success?

The first thing to note is that Vietnam's strategies cannot simply be copied by other countries because emergency response plans and policies are created within the political and economic context of the country. Thus, Vietnam's strategies, which rely on a strong one-party central government would not be applicable to a country with a divided multiple-party system. However, there are still general lessons that other low and middle income countries can learn from Vietnam. First, just as the World Bank suggests, it pays to invest in disaster prevention and preparedness. Because Vietnam invested in their healthcare system after SARS, the country had comprehensive plans and plenty of supplies to help them survive the first wave of COVID. Now that the pandemic has abated, countries should use this time to prepare for another pandemic accordingly. Second, countries should take quick and decisive action at the first sign of danger. Both Korea and Vietnam mobilized task forces to study the virus as soon as it was becoming known in Wuhan and simultaneously implemented travel restrictions and safety measures within their own borders. Given our increasingly connected world, it is important for countries to take outbreaks around the world seriously because rates of transmission are faster than ever. Third, clear and consistent communication is of tantamount importance. The government and medical professionals disseminated the same information to the public about the virus, symptoms, and protocol for infected persons. While countries can differ on the channels and methods of communication, they should strive to present a cohesive front to the public to prevent panic, instill trust, and improve compliance. Regardless of a country's economic situation, these three steps are the ones that can make all the difference in dire emergency situations.

Works Cited

Amul, Gianna Gayle, et al. "Responses to COVID‐19 in Southeast Asia: diverse paths and ongoing challenges." *Asian Economic Policy Review* 17.1 (2022): 90-110.

Bajpai, Prableen, and Suzanne Kvilhaug. "Emerging Markets: Analyzing South Korea's GDP." *Investopedia*, 8 April 2024, https://www.investopedia.com/articles/investing/091115/emerging-markets-analyzing-south-koreas-gdp.asp.

"Economics for Disaster Prevention and Preparedness in Europe." *World Bank*, 4 June 2021, https://www.worldbank.org/en/news/feature/2021/06/04/economics-for-disaster-prevention-and-preparedness-in-europe.

Global Health Security Index. *GHS Index: Homepage*, https://ghsindex.org.

Ha, Bui Thi Thu, et al. "Combating the COVID-19 epidemic: experiences from Vietnam." *International journal of environmental research and public health* 17.9 (2020): 3125.

Issac, Alwin, et al. "The pandemic league of COVID-19: Korea versus the United States, with lessons for the entire world." *Journal of Preventive Medicine and Public Health* 53.4 (2020): 228

Song, Sooho, and Yunhee Choi. "Differences in the COVID-19 pandemic response between South Korea and the United States: A comparative analysis of culture and policies." *Journal of Asian and African Studies* 58.2 (2023): 196-213.

Turner, Mark, Seung-Ho Kwon, and Michael O'donnell. "State effectiveness and crises in East and Southeast Asia: the case of COVID-19." *Sustainability* 14.12 (2022): 7216.

UNDP. "Human Development Report 2019 - Viet Nam briefing note." *United Nations in Viet Nam*, 9 December 2019, https://vietnam.un.org/en/27780-human-development-report-2019-viet-nam-briefing-note.

United Nations Development Programme Human development reports: 2019 Human Development Index ranking. Available from: http://hdr.undp.org/en/content/2019-human-development-index-ranking.

"World Bank in Vietnam" *World Bank*, 19 April 2024, https://www.worldbank.org/en/country/vietnam/overview.

Yadav, Arti, and Badar Alam Iqbal. "Socio-economic scenario of South Asia: An overview of impacts of COVID-19." *South Asian Survey* 28.1 (2021): 20-37.

Imperialism, Capital, and Sovereignty-Analyzation of the Japanese Economic Control of Hanyehping Steelworks in the Perspective of Chinese Marxists, By Hanchang Wu

Abstract

This paper analyzes Japanese economic attempts to take control of the Hanyehping Steelworks located in Central China from 1899 to 1931, in the context of the indirect interventions on China by Japan in the period as well as the rising nationalistic and anti-imperialist sentiments among Chinese intelligentsia. The discussion within the paper will mostly focus on the writings of prominent members of the Chinese Communist Party (CCP) during the 1920s, as most of its members were still cooperating with the Guangzhou Kuomintang (KMT) Government to prepare for the Northern Expedition. Their observations covered the Japanese economic policy on the company, the spontaneous actions of Japanese civilian merchants and Zaibatsu, and the decision-making of the Company's Board. The period of this study, from 1899 to 1931, witnessed China's transition from an aging monarchy to various warlord cliques, as well as before Japan's direct military actions on China proper. This paper aims to unveil the complex economic and geopolitical circumstances that characterized this tumultuous period of the Far East.

Core Argument

Hanyehping as a "national" enterprise, was both situated geographically in central China and metaphorically on the center stage of the national spotlight amidst the era of intense foreign presence. Hanyehping is the abbreviated name for three separate industrial entities in central China: Hanyang Steelworks, Tayeh Iron Mine (both in Hubei province), and Pingxiang Coal Mine (in Jiangxi province), It was established in the last decade of the nineteenth century as a part of the Qing self-strengthening movement aiming at the buildup of elementary industrial facilities. Its establishment and growth were deeply connected to the early phase of Japanese interventions in China, and it serves as an excellent example of how Japan exerted economic imperialism on other parts of Asia it did not directly control. As such, its story also fits the narrative early members of the Chinese Communist Party (CCP) tried to convey to the Chinese proletariat and intelligentsia alike, which was an "imminent foreign plot to subjugate the nation". Cases of foreign economic interventions in China were the core reasons for the early CCP's agenda of supporting the anti-imperialist National Revolution. This essay will analyze this series of historical events mainly from the perspective of several Chinese Marxist activists, and in the process examine their arguments regarding imperialism and warlordism that plagued China.

The Chinese Communist Party, founded in 1921 by a group of left-wing intellectuals that mostly received some form of Western education, was at the forefront of giving impassioned speeches and opinions regarding the entrenched foreign influence and warlord collaborators and posed some of the harshest commentaries on the matters of Hanyehping. Inspired by the successful Bolshevik takeover in Russia, CCP members like Li Dazhao, Liu Shaoqi, and Qu Qiubai devoted their efforts to spreading the novel Marxist theories (which were unfamiliar to most Chinese in the 1920s), organizing labor unions among workers in the mostly British or Japanese-owned factories, protesting the ruling Beiyang Government and various factions of warlords based in the rich coastal areas of China (whom they believed to have treacherous back-alley connections with the imperialists and betray the interest of ordinary Chinese populace). Hanyehping Ironworks, being the vanguard of Chinese-owned strategic industry, was at the center stage of attention for these activists; its geopolitical importance, weakness in operations, and apparent Japanese interest in furthering its control over the decade all sparked fury and calls for changes among progressive public opinions in China.

Liu Shaoqi, the future President of the PRC and one of the leading figures of the CCP, deeply observed the circumstances of Hanyehping in the 1920s. He was "tasked by the Hunan District Executive Committee of the Party… to unionize workers in the Anyuan Coal Mine[1], and successfully led two strikes in September 1922 and January 1925… He was elected chairman of the Hanyehping Federation of Trade Unions …in Anyuan in September 1924, and enjoyed a high reputation in both labor and capital circles." From his experience and observations among the Hanyehping workers, Liu wrote two journals in 1924 and published them respectively in Changsha and Shanghai; One is a retrospection of the reasons for the company's management failures as well as a rally call towards the larger populace, the other is a collection of suggestions towards the Steelworks' future operations. In his *To Save the Hanyehping Company*, Liu stressed the importance of maintaining the company's independence and financial liabilities: "The policy of economic invasion implemented by the various imperialist powers to hinder Chinese industry and turn the nation forever into a market for their trade goods are… very much apparent… To prevent the great powers from controlling the world's steel and obstructing the development of China's industry, Hanyehping's existence is crucial…The stakes are so high, that every citizen of this nation should contribute: First, to its long-term survival; Second, to secure its independence from foreign firms." Liu is addressing the Japanese policy of issuing loans to Hanyehping and controlling its operations in the process. Facing the opportunities of Hanyehping desperately searching for funds in the late 1890s, "the government (of Japan) secretly ordered the Industrial Bank of Japan to extend huge loans", a decision which killed two birds with one stone. It not only enabled Japanese access to "relatively untapped Asian ores and… upon cementing vertical bonds", but also "kept the ore beyond the reach of Western firms that were potential competitors". However, it is interesting to note that commentators like Liu focused more on the aspect that such economic actions by the Japanese hindered China's own industrial development, instead of strengthening the Japanese industry. Similar sentiments appeared in his essay multiple times, such as "Hanyehping did not sell iron to Japan, but being forced into a supply obligation through a debtor relationship; Suffering economic losses and being controlled all the time, and since then Japan could also influence the Chinese national destiny."

When Liu wrote this piece of work in 1924, Hanyehping was suffering from the problem of drastically shrinking steel demand on the international market due to the end of the First World War, and as a result, its business stagnated. Considering this, Liu analyzed potential markets

Hanyehping could utilize. China (due to warlordism), Europe (lack of demand for steel due to war devastation), and the US (which has a strong steel industry itself) have been ruled out; Therefore, he put the bet on Japan and the rising Soviet Union. It is interesting how Liu unwillingly admitted the inevitability of Japan's involvement: "Japan (being the only major market) still needs iron from Hanyehping at this time…due to the recent earthquake and (the need of) future wars…However, Japan's relationship with Hanyehping was too close, bound by various treaties and the enjoyment of priority rights of credits…In this way, Japan would be a not-so-optimal choice for Hanyehping business." One of the similarities among the various incidents of Japanese encroachments in China was that the Chinese recipients of Japanese aid /military intervention /loans usually willingly accepted the offers while considering it a "necessary evil", and the case is no different here at Hanyehping. In Hanyehping's case, the company's general manager, Sheng Xuanhuai, faced a similar scenario around 1900 when he negotiated the loan agreement with Yawata Steel. Sheng "released the prospectus for soliciting commercial shares to the public twice… However, until the end of the year, no single penny of shares had been raised." Sheng certainly foresaw the grave consequences of acquiring foreign capital, "On August 9, 1900, he sent a secret letter to Zhang Zhidong: 'Ever since the acquisition of Kaiping by the Westerners, they have plotted to covet every power plant and mine... The lack of prevention, maintenance, or remedy (on our side) has truly revealed the inadequacy of our response.'" Similarly, three decades later, Liu and other activists called passionately for patriotic donations by wealthy Chinese to pull the company out of the economic quagmire and to remove its dependency on Japan. He wrote, "Furthermore, citizens can follow the method of 'accumulating gold to redeem the railroads', by gathering shares among the people to pay off the Japanese debts and have the management and supervision entirely handled by the Chinese. This would then rely on the patriotism and enthusiasm of the Chinese people, as well as the efforts of propaganda from all sectors." yet their efforts were futile due to the indifference of the Chinese elites and the widespread poverty of the masses. As a last resort, Liu also pointed out that America's intention to contain Japanese expansionism may spur them into an investment, and that "the various conflicts of interest between the two nations after the Great War makes an eventual military conflict inevitable" (a brilliant observation for someone in the 1920s), Hanyehping should "utilize diplomatic flexibility to the utmost to deal with the Japanese-American tensions and free itself from Japanese yoke". In truth, the Japanese grip on Hanyehping was simply too tight, and the company finally broke loose after Japan's defeat in the Pacific War.

It should be noted that while activists like Liu Shaoqi emphasized the significance of Hanyehping Steelworks, most of their essays and proclamations regarding the Company should be put in a much bigger context: anti-imperialism, anti-warlordism, and supporting the National Revolution led by the Kuomintang, based in Guangdong. In their perspective, warlordism and imperialism is two sides of the same coin. Liu Shaoqi spoke at the Third Chinese Workers' Convention in 1926: "Over the past year… especially during the May 30th Incident, the Zhi and Feng warlords suppressed the working class with unprecedented cruelty. From these facts, we can understand that domestic warlords are the tools and lackeys of imperialism; imperialism and domestic warlords are always the enemies of the masses, that is, the targets of the National Revolution. At the same time, it can be proved that domestic warlords must be defeated before the anti-imperialist movement can prevail." Apart from collaborating with the foreign powers that Liu mentioned, the various Cliques that divided central and northern China severely hindered the transportation of trans-provincial enterprises like Hanyehping, making the buildup of national industry even more arduous. And changes were coming to China that year: The KMT government initiated a series of daring military operations that were known as the "Northern Expedition" which targeted the aforementioned Zhili and Fengtian warlords, and in the period of two years, de jure united China under one banner.

Qu Qiubai, early theorist of the CCP leadership, observed the international imperialist system in China from the Marxists perspective in his essay *Various Methods of Imperialist Encroachment on China*: "The steps of imperialism are as follows: 1) Forcefully opening markets; 2) Monopolizing raw materials; 3) Transferring capital; 4) Cultural invasion… If we discuss China's international position, this phenomenon is extremely evident. Especially because China's geographical location on the globe and the era in which it encountered Europe and America, which was neither too early nor too late, made it effectively an 'international colony.' Consequently, various imperialist powers have employed all sorts of methods to invade China, competing with each other, but ironically this competition maintained a balance of power, allowing China to barely survive… In reality, China's survival is solely due to the balance of power among the imperialist nations, with none daring to strike first, and their mutual restraint." And Japan, being the relative newcomer to this imperialist race, had to put in extra effort to hold the advantage. "Japan was forced to adopt imperialist tendencies since its domestic bourgeoise was created…its operations in China were mostly mercantile…its population flocked Chinese lands of southern Manchuria and Shandong; its goods are cheap and inferior in quality, well suited for the purchasing power of the Chinese; every characteristic mentioned conflicts with the newborn Chinese national industry." Yet Qu also emphasized the limitations of Japanese diplomatic maneuvers on China, that is, the presence of overwhelming (and established) British and American influence and strong international pressures whenever Japan tried to further its influence on China, as seen in Japan's "desperate yet unsuccessful attempt to grab Shandong" following the end of the Great War. "Japan was not rich enough to invest in China (from a capitalist perspective), yet they financially engaged in China in such a gorging manner just because it wanted to compete with Western powers. Its financial capabilities were only enough to deal with their claims on the so-called "Manchurian-Mongolian Special Interest" … all of the above-mentioned (aggressions on China) were ripe sources of propaganda for other imperialist nations to sow anti-Japanese sentiments among the Chinese populace…therefore Japan was far from capable of initiating a culture invasion." While it is clear why from his perspective Japan was forcing its pace on asserting its influence on China, Qu Qiubai did not mention a fact about Japanese economic imperialism, that is, the investments and loans on companies like Hanyehping never merely came from Japan's treasury, but from the eager collaborating

Zaibatsus as well. Civilian capital served as Japan's "white gloves" while dealing with Hanyehping: "Though functioning in a 'dummy role,' the trading firms acquired the concessions granted by the Chinese, such as Okura's marketing rights to Pingxiang coal." In general, Qu criticized Japanese imperialism (along with the ambitions of Britain and the United States) heavily for its profit-driven nature and malice towards the Chinese nation, and called for the establishment of an "independent republic for the people and by the people" while economically cooperating with the Soviet Union.

Senior Marxist scholar and organizer of the founding of CCP, Li Dazhao, had expressed his views on Hanyehping in his essay "*Memorial of National Humiliations*" even earlier: "As of now, the viable way of strengthening the nation could only be coal and iron. Hanyehping's natural resources could be of great use in the military industry. Japan wished to monopolize the enterprise, sanctioning our weaponry, and therefore nullifying our hopes of national rejuvenation. Due to (Hanyehping's) poor management at the time, hastily borrowed foreign capital, thus creating today's disastrous fate. Reflecting on the past, how can one not be heartbroken? Alas! Foreign debt is truly a medium for the nation's ruin." From those words filled with patriotism and xenophobia, the Chinese Marxists presented a distinct view on Japanese control of Hanyehping: it was not just an effort to acquire cheap resources for the empire, the whole affair was a malicious imperialist plot by taking advantage of the weak and pathetic Company management and nullify the Chinese industry in its infancy, ultimately disarming the Chinese nation from an all-out invasion in the future.

In some way, it was the fierce opposition of public opinion, strikes, and boycotts led by CCP members and other Chinese activists that made the Japanese control of Hanyehping rough and ineffective, and the unruly attitude of China signaled the failure of the indirect policies of encroachment on China performed by the civilian Japanese government in the 1920s. Such conclusions made by Japan paved the way for their more aggressive and opportunistic military actions in Northern China that eventually led to an open war with Nanjing and the eventual complete occupation of Hanyehping. In this sense, the development around Hanyehping directed the fate of Sino-Japanese relations in a much larger context than economics and industry.

Works Cited

Gao Zhonghua, Liu Shaoqi’s thoughts on solving the dilemma of Hanyehping Company in 1924, CCP Website of Party-building, accessed April 30, 2024

http://www.dangjian.com/shouye/zhuanti/zhuantiku/dangshixuexijiaoyu/202208/t20220818_6454274.shtml

Liu Shaoqi, To Save the Hanyehping Company, Online Archive of Chinese Marxism, Section 2, 3, 4, 6, accessed April 30, 2024.

https://www.marxists.org/chinese/liushaoqi/1967/004.htm

Bernard Elbaum, How Godzilla Ate Pittsburgh: The Long Rise of the Japanese Iron and Steel Industry, 1900-1973, Social Science Japan Journal, Vol. 10, No. 2, Oct. 2007, 248, accessed April 30, 2024.

https://www.jstor.org/stable/30209572

Yi Huili, Sheng Xuanhuai, Hanyang Steelworks and Japanese Loans, China Merchants History Museum Website, Shanghai Academy of Social Sciences Press, 2002, Section 1, 3, accessed April 30, 2024.

https://1872.cmhk.com/shuyuan/308.html

Liu Shaoqi, Development of the Chinese Labor Movement in the Year 1926, Online Archive of Chinese Marxism, “Conclusion and Tasks”, accessed April 30, 2024.

https://www.marxists.org/chinese/liushaoqi/mia-chinese-lsq-192605.htm

Qu Qiubai, Various Methods of Imperialist Encroachment on China, Online Archive of Chinese Marxism, Section 1, 2, accessed April 30, 2024.

https://www.marxists.org/chinese/quqiubai/mia-chinese-qqb-19230526.htm

William D. Wray. Japan's Big-Three Service Enterprises in China, 1896-1936. pp. 50, University of Rochester Blackboard Archive, accessed April 30, 2024.

Li Dazhao. Memorial of National Humiliations. Wiki source, 1915.6, accessed April 30, 2024.

https://zh.wikisource.org/zhhans/%E5%9C%8B%E6%B0%91%E4%B9%8B%E8%96%AA%E8%86%BD

Impact of Sociocultural Factors on Tobacco Consumption in India By Leela Sharma

Abstract

Cancer is an epidemic that impacts the physical, emotional, financial, and social well-being of not just the patient but also their entire family. Cancer incidence rates are growing worldwide, with many thousands of patients dying each year from cancer. According to the WHO, in 2022, there were an estimated 20 million new cancer cases and 9.7 million deaths [1]. By 2050, the estimated number of new cancer cases is expected to be over 35 million, or an increase of 77% from 2022. Governments are spending millions on research to find both medical and procedural solutions to prevent and mitigate the impact of cancer. However, there are a myriad of cultural factors that play a crucial part in the growth of cancer. By understanding these cultural paradigms, we can create awareness and aim to educate and teach, help change attitudes or beliefs, and set up individuals to form healthy lifestyle habits. This paper summarizes the critical overall global trends in cancer with a particular focus on lung cancer in India and attempts to make a connection between the socio-cultural factors that influence tobacco consumption, which in turn directly impacts lung cancer rates.

Introduction

Cancer is an epidemic that impacts the physical, emotional, financial, and social well-being of not just the patient but also their entire family. Cancer incidence rates are growing worldwide, with many patients dying each year from cancer. The World Health Organization (WHO), through its GLOBOCAN database, keeps meticulous records of cancer incidence and mortality rates. This database is the source of truth for researchers worldwide. According to the WHO, in 2022, there were an estimated 20 million new cancer cases and 9.7 million deaths. The estimated number of people who were alive within five years following a cancer diagnosis was 53.5 million. About 1 in 5 people developed cancer in their lifetime; approximately 1 in 9 men and 1 in 12 women died from the disease. By 2050, the estimated number of new cancer cases is expected to be over 35 million, or an increase of 77% from 2022 [1].

Governments are spending millions on research to find both medical and procedural solutions to prevent and mitigate the impact of cancer. Many organizations focus on the clinical side of cancer prevention, early detection, and treatment strategies and methods. However, there are a myriad of cultural factors that play a crucial part in the growth of cancer. Many of these factors fall on the prevention or awareness building side of cancer and focus on lifestyle, food habits, and physical activity, many of which are influenced by cultural practices. By understanding these cultural paradigms, we can create awareness and aim to educate and teach, help change attitudes or beliefs, and set up individuals to form healthy lifestyle habits. Changing current behaviors, which can only be accomplished by creating sufficient public awareness, can be a precursor for prevention.

This paper summarizes the key overall global trends in cancer with a particular focus on lung cancer in India and attempts to make a connection between the socio-cultural factors that influence tobacco consumption, which in turn directly impacts lung cancer rates. The impact of culture on overall cancer rates in India presents unique challenges given the diversity of cultures and subcultures in the country, as well as the vital role that cultural practices play in Indian families. Further, cancer is typically diagnosed during later stages in India, and consequently, the mortality rates are higher, and recovery rates are lower. As a result, understanding the cultural factors in India and the role of education can play a pivotal role in prevention and early diagnosis, eventually ensuring better treatment outcomes and cures.

Causes of Cancer

Cancer, a disease characterized by the uncontrolled growth of abnormal cells due to a mutation in the DNA of a single cell, poses a significant threat to health. The spread of these abnormal cells to other parts of the body and different organs, a process known as metastasis, is the primary cause of cancer-related deaths. This underscores the importance of early detection and intervention in combating the disease.

Reasons for mutation in genes are a combination of an individual's genetic and external factors. According to WHO, external factors can be grouped into three main categories: (1) Physical Carcinogens (radiation, ultraviolet rays, etc.); (2) Chemical Carcinogens (asbestos, tobacco, smoke, alcohol, arsenic, etc.), and (3) Biological Carcinogens (certain bacteria, viruses or parasites). Many of these carcinogens can cause damage to the DNA, which can, in turn, cause cancer. It's equally important to understand the amount of duration of exposure to these carcinogens as that can determine the likelihood of developing cancer.

Cancer Trends

Cancer rates in the world are increasing at an alarming rate. Figure 1 shows the cancer death rate in comparison to other diseases. Cancer is the second leading cause of death globally, next only to cardiovascular diseases, globally for the past thirty four years.

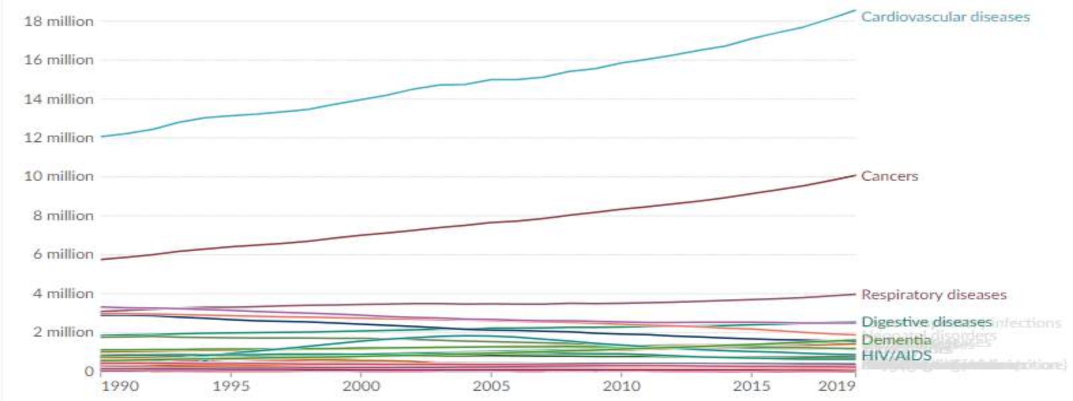


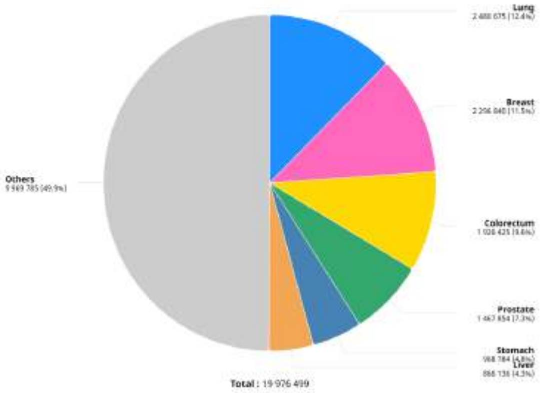
Figure 1: Causes of Death, World 1990 - 2019. [2]

Cancer Data Collection

Cancer data is collected via cancer registries. A cancer registry is an information system designed for collecting, storing, and managing data on persons with cancer[3]. Registries are data warehouses used extensively in cancer surveillance and are the bedrock on which all cancer research is based. Registries are also used to create cancer prevention and intervention programs. The WHO has mandated the International Agency for Research (IARC) as the world's premier keeper of cancer data [4]. The IARC, through its Cancer Surveillance Branch (CSU), systematically collects, analyses, interprets, and disseminates cancer data and statistics worldwide. CSU works with countries globally to collect data on cancer registries. The detailed research that the CSU develops can be accessed through the Global Cancer Observatory (GCO), a global online database that houses global cancer statistics, which helps in cancer research and control. The GCO maintains a global online database called GLOBOCAN, primarily used by researchers worldwide as the source of truth for cancer rate data.

Incidence of Cancer

In 2022, there were an estimated 20 million new cancer cases and 9.7 million deaths. The most commonly diagnosed cancers worldwide were female breast cancer (2.27 million cases), lung (2.5 million), colorectal cancer (1.9 million), and prostate cancers (1.5 million); the most common causes of cancer death were lung (1.81 million deaths), colorectal cancer (0.9 million), liver (0.8 million) and breast cancers (0.7 million) [5].



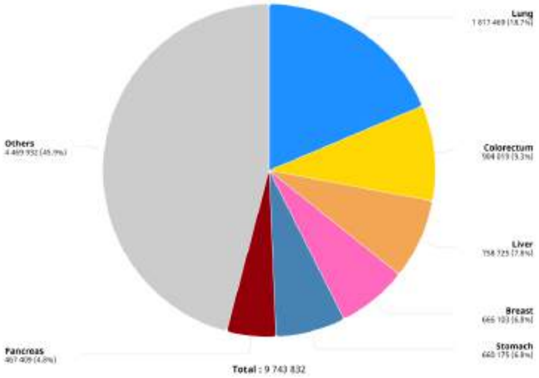


Figure 2: Cancer Incidence by cancer type, 2022 [6]

Figure 3: Cancer Mortality by cancer type, 2022 [6]

Cancer is the leading cause of death in high or very high Human Development Index (HDI) countries like the USA, Canada, France, Germany, Argentina, Chile, Australia, New Zealand etc. The Human Development Index (HDI) is a composite index of three basic dimensions of human development: a long and healthy life (based on life expectancy at birth), education (based on average and expected years of schooling), and a decent standard of living (based on gross national income per capita). The development levels of countries can be considered according to four tiers of HDI: low, medium, high, and very high HDI [7]. However, the cancer mortality rates in these countries are declining, primarily due to better prevention strategies, early detection, and more effective treatment. Cancer is the second leading cause of death after cardiovascular diseases in medium to high HDI countries like Brazil, China, and India, and many eastern European countries. In these countries, mortality rates are still increasing. The country with the top number of cancer cases is China, with 4.8 million cases, followed by the US, with 2.4 million cases, and India, with 1.4 million cases.

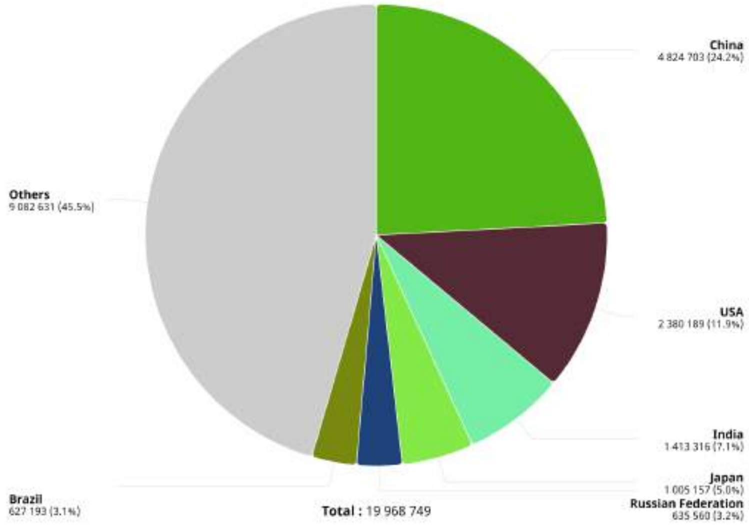


Figure 4: Cancer Incidence by Country, 2022 [6]

Cancer in India

In 2022, the estimated number of cancer cases in India was 1,461,427. Given the population of India, this equates to a Crude Incidence Rate (CIR) of 100.4 per 100,000. CIR is calculated by dividing the total number of cases by the total population at risk and is a useful metric to compare cancer incidence rates. The incidence of cancer was larger among women, estimated at 749,251 or CIR of 105.4 per 100,000, as compared to men, estimated at 712,176 or CIR of 95.6 per 100,000. The CIR of India overall was 107.0 per 100,000 [8].

The Cumulative Risk (CR) of a person developing cancer in their lifetime between 0 and 74 yr was 1 in every 9 persons for all types of cancer for both men and women. The CR is a standard metric used to predict the incidence of cancer. Absent any other competing cause of death, CR risk refers to the likelihood that a given individual will be diagnosed with cancer during their lifetime between the ages of 0 and 74 years. The CR is a good indicator of the general incidents of cancer in a geographic area or population but is not appropriate for comparison of different populations or areas with large differences in age distributions.

The Age-Standardized Rate (ASR) for all of India is 107. ASR, a hypothetical rate, is used to compare incidence/mortality over time across different countries, assuming a population with a standard age structure, and all other factors remaining unchanged. ASRs account for differences in age structure, which can have a strong influence on the risk of cancer or other factors, making it a key metric for international comparisons.

Types of Cancer

Upon further breaking down the data between the different types of cancer, we can see that the leading causes of cancer overall in people of both genders were cancers of the digestive system (288,054), followed by breast cancer (221,757), genital system (218,319), oral cavity and pharynx (198,438) and respiratory system (143,062) [8].

Breakdown by Gender

In women, the overwhelming leading cause of cancer is breast cancer at 216,108 cases or 28.8%, followed by cancers of the genital system at 163,694 cases or 21.9% and cancers of the digestive system at 116,029 cases or 15.5%. The cumulative risk of a woman to develop breast cancer in her lifetime from birth to 74 years of age was 1 in every 29 women.

In men, the overwhelming leading cause of cancer is lung cancer at 75,474 cases or 10.6%, followed by cancers of the mouth at 60,164 cases or 8.5% and prostate cancer at 43,691 cases or 6.13%. The cumulative risk of a man to develop lung cancer during his lifetime between 0 and 74 years was one in every 67 for lung cancers in men [8].

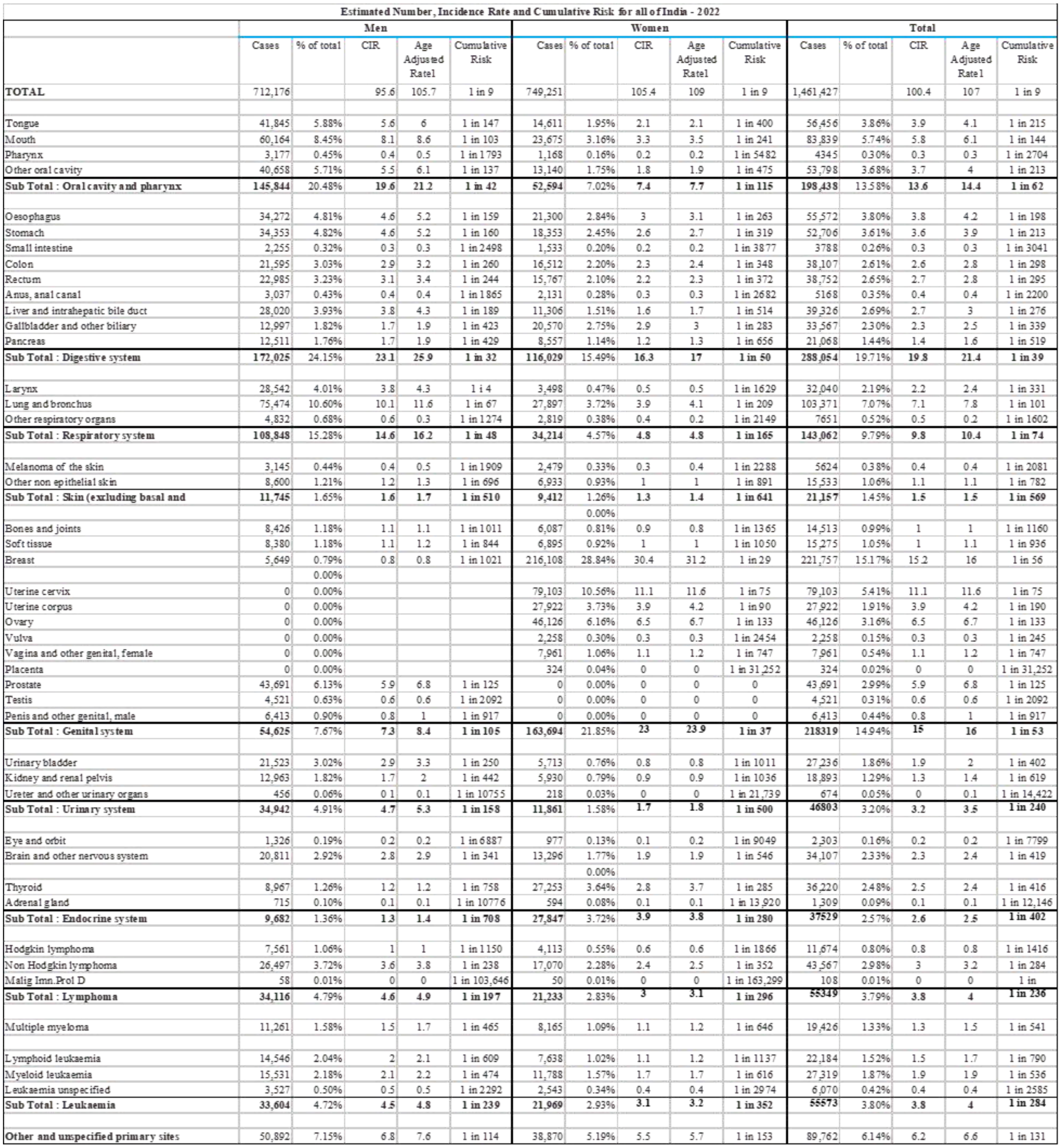
Breakdown by Age

For children ages 1-14 years old, the leading type of cancer was Lymphoid Leukemia for both boys and girls at 29.3% and 24.3% respectively. For girls and women, the leading type of cancer from age 15 onwards is breast cancer at 27.3% among patients ranging from 15 to 39 years of age, 33% for those ranging from 40 to 64 years of age and 23% for those aged 65 years and older. For boys and men, the leading cause of cancer was oral cancer at 12% for 15-39 years of age, and lung cancer for men 40+ years. This breakdown shows the apparent direct correlation between the rates of cancer diagnoses and age for both sexes. The 15 to 39 years old age group has reported the least amount of diagnoses while the eldest age group for both sexes show the most frequent cancer rates [8].

Breakdown by Year

Cancer cases in India are increasing over the years. On comparing cases in 2015, 2020 and estimates for 2025, overall, cases increased at a compounded annual growth rate (CAGR) of approximately 3% annually. The compound annual growth rate measures the annual growth rate when only the value at the end of the period and the value at the beginning of the period are provided. This is one of the most accurate ways of measuring the growth rate.

The estimated rate of increase in cancer cases in women from 2022 to 2025 is slightly higher at 3.13% versus men which is at 2.96%. However the rate of increase of cases is decreasing. Between 2015 and 2020, the overall cancer cases in India grew at an annual rate of 3.25% while from 2020 to 2025 it is expected to grow at an annual rate of 3.05%. While the rate of increase might be lower, the number of cancer cases is still expected to increase from 1.46 million in 2022 to 1.57 million cases in 2025. The cancer estimates are prepared by Globoscan [8].

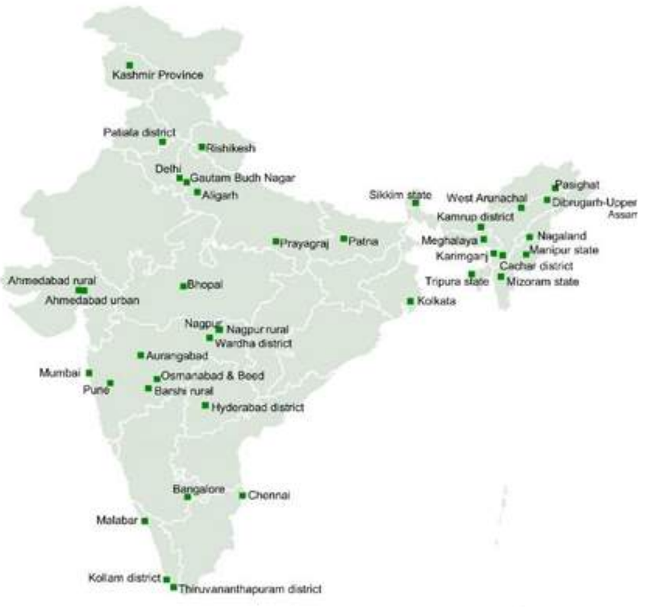


Estimated Number, Incidence Rate and Cumulative Risk for all of India - 2022

	Men					Women					Total				
	Cases	% of total	CIR	Age Adjusted Rate1	Cumulative Risk	Cases	% of total	CIR	Age Adjusted Rate1	Cumulative Risk	Cases	% of total	CIR	Age Adjusted Rate1	Cumulative Risk
TOTAL	712,176		95.6	105.7	1 in 9	749,251		105.4	109	1 in 9	1,461,427		100.4	107	1 in 9
Tongue	41,845	5.88%	5.6	6	1 in 147	14,611	1.95%	2.1	2.1	1 in 400	56,456	3.86%	3.9	4.1	1 in 215
Mouth	60,164	8.45%	8.1	8.6	1 in 103	23,675	3.16%	3.3	3.5	1 in 241	83,839	5.74%	5.8	6.1	1 in 144
Pharynx	3,177	0.45%	0.4	0.5	1 in 1793	1,168	0.16%	0.2	0.2	1 in 5482	4345	0.30%	0.3	0.3	1 in 2704
Other oral cavity	40,658	5.71%	5.5	6.1	1 in 137	13,140	1.75%	1.8	1.9	1 in 475	53,798	3.68%	3.7	4	1 in 213
Sub Total : Oral cavity and pharynx	**145,844**	20.48%	**19.6**	**21.2**	**1 in 42**	**52,594**	7.02%	**7.4**	**7.7**	**1 in 115**	**198,438**	13.58%	**13.6**	**14.4**	**1 in 62**
Oesophagus	34,272	4.81%	4.6	5.2	1 in 159	21,300	2.84%	3	3.1	1 in 263	55,572	3.80%	3.8	4.2	1 in 198
Stomach	34,353	4.82%	4.6	5.2	1 in 160	18,353	2.45%	2.6	2.7	1 in 319	52,706	3.61%	3.6	3.9	1 in 213
Small intestine	2,255	0.32%	0.3	0.3	1 in 2498	1,533	0.20%	0.2	0.2	1 in 3877	3788	0.26%	0.3	0.3	1 in 3041
Colon	21,595	3.03%	2.9	3.2	1 in 260	16,512	2.20%	2.3	2.4	1 in 348	38,107	2.61%	2.6	2.8	1 in 298
Rectum	22,985	3.23%	3.1	3.4	1 in 244	15,767	2.10%	2.2	2.3	1 in 372	38,752	2.65%	2.7	2.8	1 in 295
Anus, anal canal	3,037	0.43%	0.4	0.4	1 in 1865	2,131	0.28%	0.3	0.3	1 in 2682	5168	0.35%	0.4	0.4	1 in 2200
Liver and intrahepatic bile duct	28,020	3.93%	3.8	4.3	1 in 189	11,306	1.51%	1.6	1.7	1 in 514	39,326	2.69%	2.7	3	1 in 276
Gallbladder and other biliary	12,997	1.82%	1.7	1.9	1 in 423	20,570	2.75%	2.9	3	1 in 283	33,567	2.30%	2.3	2.5	1 in 339
Pancreas	12,511	1.76%	1.7	1.9	1 in 429	8,557	1.14%	1.2	1.3	1 in 656	21,068	1.44%	1.4	1.6	1 in 519
Sub Total : Digestive system	**172,025**	24.15%	**23.1**	**25.9**	**1 in 32**	**116,029**	15.49%	**16.3**	**17**	**1 in 50**	**288,054**	19.71%	**19.8**	**21.4**	**1 in 39**
Larynx	28,542	4.01%	3.8	4.3	1 i 4	3,498	0.47%	0.5	0.5	1 in 1629	32,040	2.19%	2.2	2.4	1 in 331
Lung and bronchus	75,474	10.60%	10.1	11.6	1 in 67	27,897	3.72%	3.9	4.1	1 in 209	103,371	7.07%	7.1	7.8	1 in 101
Other respiratory organs	4,832	0.68%	0.6	0.3	1 in 1274	2,819	0.38%	0.4	0.2	1 in 2149	7651	0.52%	0.5	0.2	1 in 1602
Sub Total : Respiratory system	**108,848**	15.28%	**14.6**	**16.2**	**1 in 48**	**34,214**	4.57%	**4.8**	**4.8**	**1 in 165**	**143,062**	9.79%	**9.8**	**10.4**	**1 in 74**
Melanoma of the skin	3,145	0.44%	0.4	0.5	1 in 1909	2,479	0.33%	0.3	0.4	1 in 2288	5624	0.38%	0.4	0.4	1 in 2081
Other non epithelial skin	8,600	1.21%	1.2	1.3	1 in 696	6,933	0.93%	1	1	1 in 891	15,533	1.06%	1.1	1.1	1 in 782
Sub Total : Skin (exluding basal and	**11,745**	1.65%	**1.6**	**1.7**	**1 in 510**	**9,412**	1.26%	**1.3**	**1.4**	**1 in 641**	**21,157**	1.45%	**1.5**	**1.5**	**1 in 569**
							0.00%								
Bones and joints	8,426	1.18%	1.1	1.1	1 in 1011	6,087	0.81%	0.9	0.8	1 in 1365	14,513	0.99%	1	1	1 in 1160
Soft tissue	8,380	1.18%	1.1	1.2	1 in 844	6,895	0.92%	1	1	1 in 1050	15,275	1.05%	1	1.1	1 in 936
Breast	5,649	0.79%	0.8	0.8	1 in 1021	216,108	28.84%	30.4	31.2	1 in 29	221,757	15.17%	15.2	16	1 in 56
		0.00%													
Uterine cervix	0	0.00%				79,103	10.56%	11.1	11.6	1 in 75	79,103	5.41%	11.1	11.6	1 in 75
Uterine corpus	0	0.00%				27,922	3.73%	3.9	4.2	1 in 90	27,922	1.91%	3.9	4.2	1 in 190
Ovary	0	0.00%				46,126	6.16%	6.5	6.7	1 in 133	46,126	3.16%	6.5	6.7	1 in 133
Vulva	0	0.00%				2,258	0.30%	0.3	0.3	1 in 2454	2,258	0.15%	0.3	0.3	1 in 245
Vagina and other genital, female	0	0.00%				7,961	1.06%	1.1	1.2	1 in 747	7,961	0.54%	1.1	1.2	1 in 747
Placenta	0	0.00%				324	0.04%	0	0	1 in 31,252	324	0.02%	0	0	1 in 31,252
Prostate	43,691	6.13%	5.9	6.8	1 in 125	0	0.00%	0	0	0	43,691	2.99%	5.9	6.8	1 in 125
Testis	4,521	0.63%	0.6	0.6	1 in 2092	0	0.00%	0	0	0	4,521	0.31%	0.6	0.6	1 in 2092
Penis and other genital, male	6,413	0.90%	0.8	1	1 in 917	0	0.00%	0	0	0	6,413	0.44%	0.8	1	1 in 917
Sub Total : Genital system	**54,625**	7.67%	**7.3**	**8.4**	**1 in 105**	**163,694**	21.85%	**23**	**23.9**	**1 in 37**	**218319**	14.94%	**15**	**16**	**1 in 53**
Urinary bladder	21,523	3.02%	2.9	3.3	1 in 250	5,713	0.76%	0.8	0.8	1 in 1011	27,236	1.86%	1.9	2	1 in 402
Kidney and renal pelvis	12,963	1.82%	1.7	2	1 in 442	5,930	0.79%	0.9	0.9	1 in 1036	18,893	1.29%	1.3	1.4	1 in 619
Ureter and other urinary organs	456	0.06%	0.1	0.1	1 in 10755	218	0.03%	0	0	1 in 21,739	674	0.05%	0	0.1	1 in 14,422
Sub Total : Urinary system	**34,942**	4.91%	**4.7**	**5.3**	**1 in 158**	**11,861**	1.58%	**1.7**	**1.8**	**1 in 500**	**46803**	3.20%	**3.2**	**3.5**	**1 in 240**
Eye and orbit	1,326	0.19%	0.2	0.2	1 in 6887	977	0.13%	0.1	0.2	1 in 9049	2,303	0.16%	0.2	0.2	1 in 7799
Brain and other nervous system	20,811	2.92%	2.8	2.9	1 in 341	13,296	1.77%	1.9	1.9	1 in 546	34,107	2.33%	2.3	2.4	1 in 419
							0.00%								
Thyroid	8,967	1.26%	1.2	1.2	1 in 758	27,253	3.64%	2.8	3.7	1 in 285	36,220	2.48%	2.5	2.4	1 in 416
Adrenal gland	715	0.10%	0.1	0.1	1 in 10776	594	0.08%	0.1	0.1	1 in 13,920	1,309	0.09%	0.1	0.1	1 in 12,146
Sub Total : Endocrine system	**9,682**	1.36%	**1.3**	**1.4**	**1 in 708**	**27,847**	3.72%	**3.9**	**3.8**	**1 in 280**	**37529**	2.57%	**2.6**	**2.5**	**1 in 402**
Hodgkin lymphoma	7,561	1.06%	1	1	1 in 1150	4,113	0.55%	0.6	0.6	1 in 1866	11,674	0.80%	0.8	0.8	1 in 1416
Non Hodgkin lymphoma	26,497	3.72%	3.6	3.8	1 in 238	17,070	2.28%	2.4	2.5	1 in 352	43,567	2.98%	3	3.2	1 in 284
Malig Imn.Prol D	58	0.01%	0	0	1 in 103,646	50	0.01%	0	0	1 in 163,299	108	0.01%	0	0	1 in
Sub Total : Lymphoma	**34,116**	4.79%	**4.6**	**4.9**	**1 in 197**	**21,233**	2.83%	**3**	**3.1**	**1 in 296**	**55349**	3.79%	**3.8**	**4**	**1 in 236**
Multiple myeloma	11,261	1.58%	1.5	1.7	1 in 465	8,165	1.09%	1.1	1.2	1 in 646	19,426	1.33%	1.3	1.5	1 in 541
Lymphoid leukaemia	14,546	2.04%	2	2.1	1 in 609	7,638	1.02%	1.1	1.2	1 in 1137	22,184	1.52%	1.5	1.7	1 in 790
Myeloid leukaemia	15,531	2.18%	2.1	2.2	1 in 474	11,788	1.57%	1.7	1.7	1 in 616	27,319	1.87%	1.9	1.9	1 in 536
Leukaemia unspecified	3,527	0.50%	0.5	0.5	1 in 2292	2,543	0.34%	0.4	0.4	1 in 2974	6,070	0.42%	0.4	0.4	1 in 2585
Sub Total : Leukaemia	**33,604**	4.72%	**4.5**	**4.8**	**1 in 239**	**21,969**	2.93%	**3.1**	**3.2**	**1 in 352**	**55573**	3.80%	**3.8**	**4**	**1 in 284**
Other and unspecified primary sites	50,892	7.15%	6.8	7.6	1 in 114	38,870	5.19%	5.5	5.7	1 in 153	89,762	6.14%	6.2	6.6	1 in 131

Figure 5: Estimated Number, Incidence Rate and Cumulative Risk for all of India - 2022 [8]

Similar to global cancer registries, cancer data in India is also collected using Population Based Cancer Registries (PBCR). PBCRs collect data on cancer incidence and mortality in their respective geographic areas on a continuous basis. They provide information on burden and trends of cancer in the population over time and projected future estimates of incident number of cases. There are 38 PBCRs functioning in India [9]. In addition there are 213 Hospital Based Cancer Registries (HBCR), that provide data on cancer patterns in hospital, clinical, pathological and treatment related details of cancer patients. HBCRs also contribute to the PBCRs and the combined data becomes part of GLOBOCAN data from India. PBCRs is a very labor intensive process and is backward looking. There is no real time data collection but rather trained registry staff collect data from hospitals, laboratories, etc. Because of this manual collection of data, usually there is a lag of 2-4 years between actual PBCR data and when it gets reflected in GLOBOSCAN. One of the biggest issues with this methodology is the lack of consistent PBCR coverage within all states in India.



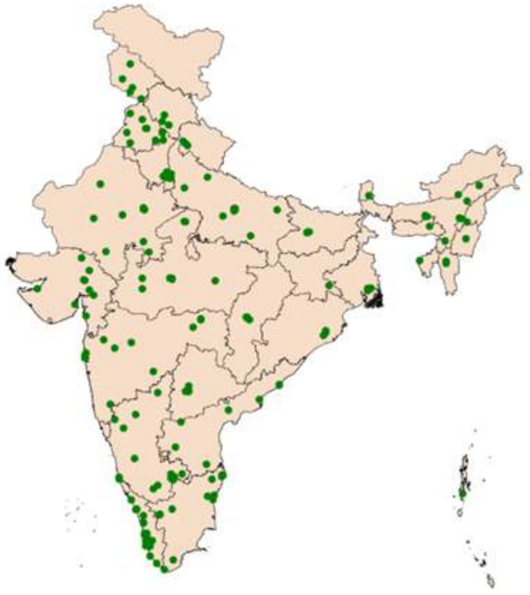


Figure 6: PBCR Network in India [9]

Figure 7: HBCR Network in India [9]

Lung Cancer

Among all types of cancers, Lung Cancer is responsible for the most deaths worldwide with a mortality rate of 1.8 million deaths in 2022. In 1950, a groundbreaking study conducted by Richard Doll and Austin Bradford Hill (English epidemiologist and statisticians, 1897–1991), suggested that tobacco smoking and exposure to outdoor pollutants were the two main causes of lung cancer. Since then prevalence of both these key risk factors has increased over time and has contributed to the increase in lung cancer.

The primary cause of lung cancer is tobacco smoking, which is responsible for 63% of all deaths and more than 90% of lung cancer deaths in countries where smoking is prevalent [10]. Thus, by studying smoking patterns across cultures, we can make a determination of lung cancer trends. For example, in high to very high HDI countries, where smoking first began and then subsequently decreased, lung cancer incidence and mortality trends have followed similar patterns for countries like the USA, Canada, Australia, New Zealand, etc. See figures 8 and 9 below for Age-standardized (a) incidence rates and (b) mortality rates per 100,000 person-years by calendar year in selected countries for lung cancer in men, from 1975–2012 [10].

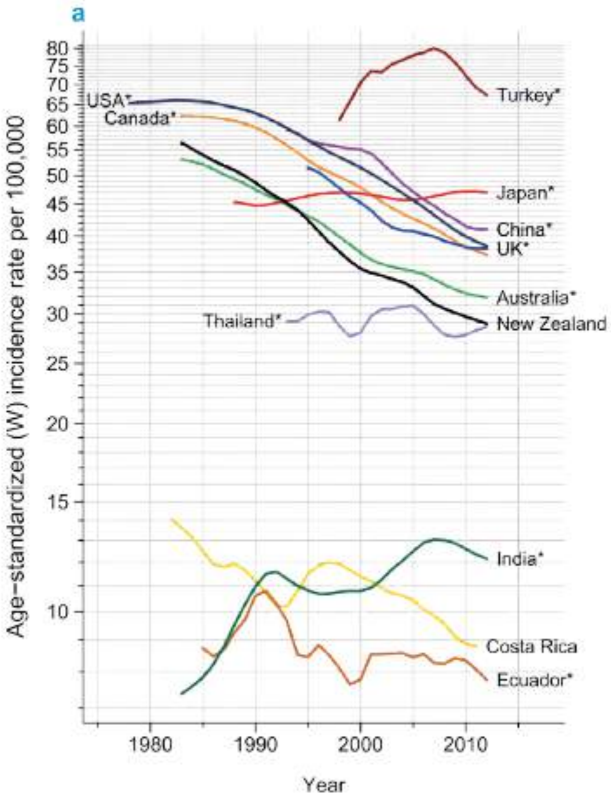


Figure 8: Age Standardized Cancer Incidence Rates [10]

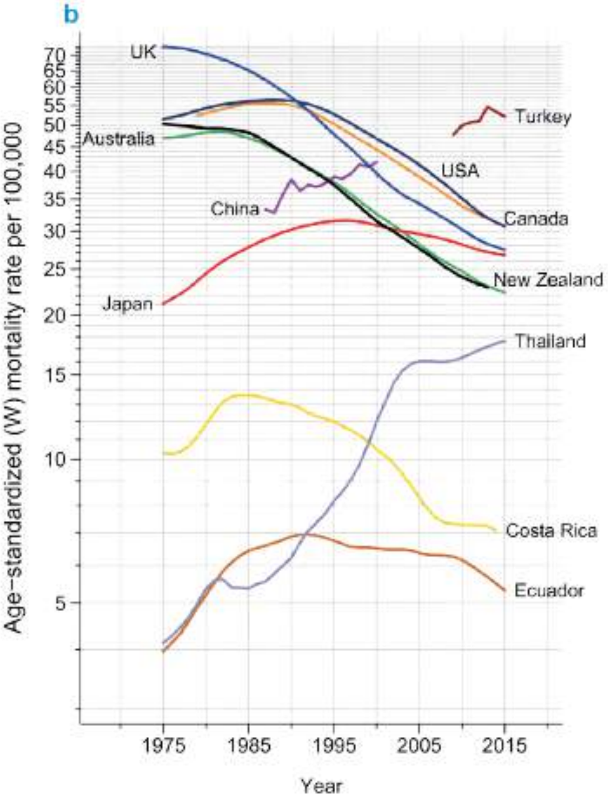


Figure 9: Age Standardized Cancer Mortality Rates [10]

In countries with mid or lower HDI index, like India, Ecuador, etc. the incidence rates have remained somewhat stable, after a large spike in the 1980s, reflecting that smoking trends in those countries have stabilized in recent years.

Tobacco Usage and Lung Cancer

Since smoking is one of the primary contributors to lung cancer, it is essential to understand the prevalence of tobacco consumption to have meaningful discussions on disease prevention and control. In the world, there are 1.1 billion adult smokers and at least 303 million users of tobacco.

What is it about tobacco that makes it attractive to people? Tobacco contains Nicotine, a naturally occurring ingredient that is found in tobacco. Nicotine is so highly addictive that just after three or four cigarettes, adolescents can begin to become addicted. Tobacco usage is harmful in all forms, whether it is smoked, chewed, applied on gums and teeth, or mixed with other ingredients. While cigarettes are the largest manufactured tobacco products in the world (96% of the total value of sales), there are other traditional methods of smoking and chewing tobacco that are prevalent in the world, especially in India and other parts of Southeast Asia. Excessive use of tobacco, either via smoking or other ways, can lead to cancer. As a result, in addition to lung cancer, tobacco is also associated with a large number of other cancers, such as lip, tongue, mouth, oropharynx, larynx, esophagus, and urinary bladder.

The recent trend in tobacco/nicotine is from Electronic Nicotine Delivery Systems (ENDS) or e-cigarettes as they are more popularly known. These are primary devices that heat a liquid to create an aerosol that is inhaled by the user. The liquid contains Nicotine (but not tobacco) and other chemicals that may be toxic to people's health. While these products are being marketed as 'cleaner,' there is not sufficient evidence to support the claim that these products are risk-free. The long-term health impact of these remains unknown.

Tobacco Usage in India

Tobacco usage is a major public health challenge in India. Approximately 28.6% of adults use different types of tobacco, which is higher than the global rate of 23.4%. According to WHO's Global Adult Tobacco Fact Sheet from 2016-2017, which is a global standard for monitoring adult tobacco use and tracking key tobacco control indicators, men share a higher proportion of tobacco use, with 42.4% of men as compared to only 14.2% of women make up the total 28.6% of adults that use all types of tobacco. In addition to tobacco use, secondhand smoke is also a deadly carcinogen, as 38.7% are exposed to secondhand smoke at home, while 30.2% of adults who work indoors are exposed to secondhand smoke at their workplace [11].

Socio-Cultural Practices influencing Tobacco usage in India

With the mixture of different cultures and societal practices, India's tobacco usage is very diverse and complex, with a variety of smoking forms and Smokeless Tobacco products (SLT). The Government of India has set up comprehensive tobacco use cessation measures under the National Tobacco Control Programme and implemented the Cigarettes and Other Tobacco Products Act, but tobacco use continues to be widespread.

Tobacco usage in India has distinct patterns in urban and rural areas. In urban areas, tobacco is smoked in the form of cigarettes, while in rural areas, the consumption of bidi and hookah is more prevalent. Bidis (pronounced bee-dees) are conical shaped, small, hand-rolled cigarettes that contain a small amount of flaked tobacco (0.2 grams) and are hand wrapped in tendu or temburni leaf (Diospyros melanoxylon, a plant that is native to Asia) and tied with a string. Bidis are manufactured in India and other Southeast Asian countries. Bidis are less expensive and more heavily consumed than traditional commercial cigarettes. Every year, around 750 billion to 1.2 trillion bidis are produced and consumed in India, accounting for nearly half (~48%) of tobacco consumption and making them much more popular than conventional cigarettes (which account for ~14% of tobacco consumption) [12]. Even though bidis have a small amount of tobacco, they have more tar and hence can deliver more carbon monoxide than regular cigarettes. They, therefore, carry a greater risk of causing cancer.

Due to the significant price difference between cigarettes and bidis, people who are wealthier, more educated, and have decent jobs are more likely to smoke cigarettes. On the other hand, people who are less educated with poor socioeconomic status are more likely to have a habit of bidi smoking. In India, which is a developing country with wide ranges in income levels, while 2.2% of the overall adult population engages in daily cigarette smoking, the percentage of the population that smokes bidis every day is much higher at 6.4%. In India, smoking tobacco is more prevalent in urban areas than it is in rural areas [13].

The prevalence of Smokeless Tobacco (SLT) in India is large and growing. About 26% of all adults in India use SLT by chewing, applying it to the teeth and gums, or by sniffing. The use of SLT among males (33%) is higher than that of females (18%). In rural areas, 29% of adults use SLT, whereas 18% use it in urban areas [14]. The usage of SLT is driven by misconceptions that it is good for dental health, helps in weight reduction, and suppresses hunger. These beliefs are more prevalent in lower socioeconomic strata and in rural areas due to a lack of education [13].

There are many socioeconomic factors that influence tobacco consumption in India. Tobacco consumption is more prevalent among the population with lower socioeconomic status [15]. Socioeconomic positions are measured by various indicators such as education, occupation, and income level. Education is the most widely used and obvious indicator of socioeconomic position. It is easy to measure universally and is applicable to every individual regardless of age, gender, cultural background, etc. It also has a strong correlation to tobacco use, both in India and in other parts of the world [16]. Typically, the indicators are linked and used interchangeably to measure socioeconomic position. For example, an individual who is educated will have a decent job that pays him enough income to support a higher lifestyle than a person, who is not educated enough and thus has a lower paying job.

Individuals with lower socioeconomic status have more inclination towards consumption of more tobacco and also less likely to be successful in quit attempts [13]. These individuals are less educated on the dangers of tobacco, are less exposed to resources that speak to the harmful effects of tobacco, and, as a result, will likely have a very strong addiction and less motivation to quit. It is likely that these individuals, given their lack of education, do not hold a good paying job and as a result might also have other psychological and behavioral issues such as lack of self-efficacy, confidence to not get swayed by slick marketing efforts coming from the very powerful tobacco industry. It is also likely that such individuals do not have adequate community support.

There is a complex interplay of family and society that also influences the adoption of tobacco in India. Inequalities arising from disparate social status can lead to differences in tobacco use. In certain parts of India, especially North India, it was socially acceptable for the dominant elder male members of the family to smoke the ‘hookah’ (a water pipe used for heating or vaporizing and then smoking tobacco). It was not acceptable for younger members of the family to consume tobacco in the presence of elder family members, since it is seen as a sign of disrespect. Due to both the historical context and positioning by cigarette manufacturers, smoking is considered more masculine. As a consequence, smoking incidence is higher among men than women across geographies, urban/ rural and socio-economic status. Smoking by women in India has traditionally and till today continues to be socially frowned upon.

Historically, different generations of a family would never sit together at home and smoke together. From the 20th century, due to economic pressures and families moving to different parts of the country in search of jobs, the structure of the joint families started to split into nuclear families, which has reduced the social taboo of smoking in front of elders and likely increased the consumption of tobacco and smoking.

Prevalence of smoking in India is the highest in North Eastern States, with Mizoram, Tripura, Manipur, Meghalaya, all at above 50% of population above 15 years of age using tobacco. Mizoram, Meghalaya, Manipur, and Nagaland also have more than a third of youths (15-24 years) using tobacco, and the majority of usage is in SLT. The biggest driver of tobacco usage is the effect of parental use of tobacco. Youths who live with mothers who have been using any tobacco are at 3.4 times more risk of using tobacco themselves. Youths who live with fathers who use tobacco are 1.14 times more likely to use any tobacco. Cultural factors play a large part in driving high tobacco usage in this region. Not only is tobacco usage socially acceptable, it is also an integral part of the culture of the region. Contributory socioeconomic and cultural factors also include lower economic profile of the region as well as higher incidence of alcohol usage (which correlates highly to tobacco use) as well as lower taxes on SLT products and beedis, which make tobacco usage more affordable [17].

Conclusion

Cancer is a deadly disease that has remained elusive from a cure. The community of doctors, scientists and researchers are doing some groundbreaking and innovative research to change the current state of the epidemic. However, prevention remains a key strategy in battling this disease, especially lung cancer. Tobacco consumption remains the primary cause of lung cancer. In India, socio-cultural factors play a key role in influencing tobacco consumption. It is imperative that we understand the socio-cultural factors while studying lung cancer trends in India. Any change in behavior has to include an understanding of the cultural factors that impact tobacco consumption in India.

****

Works Cited

[1] "Global Cancer Burden Growing, amidst Mounting Need for Services." *World Health Organization*, World Health Organization, www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services. Accessed 7 June 2023.

[2] ourworldindata.org/cancer. Accessed 7 July 2023.

[3] "What Is a Cancer Registry?" SEER, seer.cancer.gov/registries/cancer_registry/. Accessed 23 June 2023.

[4] "Cancer Surveillance Branch (CSU)." World Health Organization, World Health Organization, www.iarc.who.int/branches-csu/. Accessed 23 June 2023.

[5] "Cancer Today." International Agency for Research on Cancer, gco.iarc.fr/today/. Accessed 24 June 2023.

[6] "Cancer Today." *World Health Organization*, World Health Organization, gco.iarc.who.int/today. Accessed 7 July 2023.

[7] WORLD HEALTH ORGANIZATION: REGIONAL OFFICE FOR EUROPE. *World Cancer Report: Cancer Research for Cancer Development*. IARC, 2020.

[8] Mathur, Prashant, et al. "Cancer incidence estimates for 2022 & projection for 2025: Result from National Cancer Registry Programme, India." *Indian Journal of Medical Research*, vol. 156, no. 4, 20 Aug. 2022, pp. 598–607, https://doi.org/10.4103/ijmr.ijmr_1821_22.

[9] *Ncdirindia*, www.ncdirindia.org/All_Reports/AR/AH_2022_2023.pdf. Accessed 30 June 2023.

[10] World Health Organization, International Agency for Research on Cancer, Lyon, France, 2020, *World Cancer Report*.

[11] WHO. Tobacco Free Initiative. Global Adult Tobacco Survey (GATS) India report 2016–2017. Geneva, Switzerland: World Health Organization.

[12] Pednekar MS, Gupta PC, Yeole BB, Hébert JR. Association of tobacco habits, including bidi smoking, with overall and site-specific cancer incidence: results from the Mumbai cohort study. Cancer Causes Control. 2011 Jun;22(6):859-68. doi: 10.1007/s10552-011-9756-1. Epub 2011 Mar 24. PMID: 21431915; PMCID: PMC3756904.

[13] Shah S, Dave B, Shah R, Mehta TR, Dave R. Socioeconomic and cultural impact of tobacco in India. J Family Med Prim Care. 2018 Nov-Dec;7(6):1173-1176. doi: 10.4103/jfmpc.jfmpc_36_18. PMID: 30613493; PMCID: PMC6293949.

[14] WHO. Tobacco Free Initiative. Global Adult Tobacco Survey (GATS) India report 2009–2010. Geneva, Switzerland: World Health Organization; 2011.

[15] Hiscock R, Bauld L, Amos A, Fidler JA, Munafò M. Socioeconomic status and smoking: a review. Ann N Y Acad Sci. 2012 Feb;1248:107-23. doi: 10.1111/j.1749-6632.2011.06202.x. Epub 2011 Nov 17. PMID: 22092035.

[16] Sorensen G, Gupta PC, Pednekar MS. Social disparities in tobacco use in Mumbai, India: the roles of occupation, education, and gender. Am J Public Health. 2005 Jun;95(6):1003-8. doi: 10.2105/AJPH.2004.045039. PMID: 15914825; PMCID: PMC1449300.

Diagnosing Respiratory Diseases using Machine Learning Algorithms and Signal Processing on Lung Sounds By Anusha Sundar

Abstract

Respiratory illnesses are a major contributor to global mortality and morbidity. Accurate and timely diagnosis is crucial for improving the outcomes of lung diseases. Due to the paucity of healthcare resources in developing countries, innovative approaches such as analysis of respiratory sounds by Machine learning algorithms should be employed to diagnose lung conditions. This study aims to explore Machine Learning models and modeling techniques to predict lung diseases using audio recordings of respiratory sounds. The audio recordings were converted to spectrograms as inputs to Convolutional Neural Network (CNN) models for prediction. A variety of experiments were conducted to determine the optimal modeling procedure. The experiments considered various factors, including spectrogram types, sampling methods, CNN model architectures, and the incorporation of demographic information. It was concluded that STFT spectrograms worked better than MFCC spectrograms and Weighted Random Sampling worked better than Random Down Sampling. The Simple CNN Model performed better than both the hyperparameter-tuned Model and the pre-trained RESNET models. The model performance was also enhanced with the inclusion of the demographic information. This study successfully applies Machine Learning to predict lung diseases using respiratory audio recordings and documents the optimal modeling approach.

Introduction

Respiratory diseases, or diseases of the lungs, are common causes of morbidity and mortality worldwide. According to the WHO, respiratory diseases accounted for over 8 million deaths in 2019 (*The Top 10 Causes of Death*). The spectrum of respiratory diseases includes chronic conditions such as COPD and bronchiectasis as well as acute conditions such as pneumonia, bronchiolitis, and upper respiratory tract infections.

Chronic respiratory diseases were the third leading cause of death responsible for 4.0 million deaths globally in 2019 (Momtazmanesh et al.). Within this class, Chronic Obstructive Pulmonary Disease (COPD) is the most prevalent and accounted for 3.3 million deaths (Momtazmanesh et al.). COPD is a heterogenous group of obstructive lung conditions which present as dyspnea and cough (*COPD - Symptoms | NHLBI, NIH*) Common risk factors for COPD include history of smoking, exposure to biomass, and air pollution (*GOLD-2021-Chapter-1.Pdf*). It results in frequent physician office visits and multiple hospitalizations due to acute exacerbations (Albert Richard K. et al.).

Acute respiratory conditions, notably pneumonia, are lung infections caused by bacteria, viruses, and fungi and manifest with symptoms such as cough, fever, and dyspnea (Association). The annual incidence of pneumonia is estimated at 151 million new cases per year, of which 11–20 million (7–13%) cases are severe enough to necessitate hospitalization (Bhutta). Pneumonia stands as the leading cause of child mortality, responsible for nearly 15% of all deaths among children under the age of 5 worldwide (Liu et al.). In 2015, developing countries such as India, Nigeria, Indonesia, Pakistan, China and Ethiopia contributed to over 54% of pneumonia cases and 49% of deaths from pneumonia globally (McAllister et al.). Key factors exacerbating mortality rates in these regions include overcrowding, inadequate sanitation, poor nutrition (McAllister et al.), and most importantly, delayed access to healthcare (Graham et al.).

Despite the distinct pathophysiology of these lung diseases, they often present with similar symptoms such as cough and dyspnea, making accurate diagnosis challenging (Celli et al.). This is especially true in low-income countries, where healthcare services commonly adopt a three-tiered organization system consisting of community health centers, district hospitals, and regional central hospitals (Aston). In community health centers, where the majority of patients are treated, initial evaluation is conducted by first-level healthcare workers. Because of their limited medical training, these healthcare workers are often not trained in simple medical skills such as lung auscultation (Aston). Due to the lack of specificity of clinical symptoms of respiratory diseases, community health workers are unable to accurately diagnose them (Bhutta). Consequently, respiratory diseases are often misdiagnosed and under-reported (Mulupi et al.).

Background

One possible way to mitigate the issue of erroneous diagnosis in underserved areas is to equip community health centers with digital stethoscopes for lung auscultation. Digital Signal Processing (DSP) and Machine Learning algorithms can then be used to analyze and classify the recorded lung sounds. DSP can decompose audio signals into its constituent frequency components (*FFT*), at a level of detail that is impossible for humans to distinguish. Machine Learning algorithms, in turn, can be trained to learn these subtle differences in signal characteristics and accurately predict the underlying disease.

A drawback of this approach is that lung auscultation alone has a low sensitivity for diagnosing pneumonia. A metaregression analysis of 34 studies found that the sensitivity of lung auscultation in diagnosing pneumonia is 37% and specificity 89% (Arts et al.). In other words, 63% of patients with pneumonia would go undetected if we were to solely rely on lung auscultation. Furthermore, physicians often combine lung auscultation with patient demographics and clinical symptoms in the diagnostic process (Htun et al.). It would therefore be prudent to incorporate this information into the predictors for Machine Learning algorithms. Earlier research studies have relied either exclusively on lung auscultation (Alqudah et al.) (Arts et al.) (Huang et al.) or on a combination of patient symptoms and lung auscultation (Stokes et al.) to train Machine Learning algorithms. However, none of the prior studies have combined demographics and lung sounds to train Machine Learning algorithms for diagnosing respiratory conditions.

The objective of this study is 1) to develop Machine Learning algorithms to classify respiratory diseases based on lung sounds as well as patient demographic information, and 2) to explore various modeling techniques to optimize performance.

Data Source

The algorithms in this study were trained using the International Conference on Biomedical and Health Informatics (ICBHI) Scientific Challenge database (He et al.). This respiratory sound database is one of the largest publicly available databases on lung auscultation sounds, and was created by two research teams in Portugal and Greece.

Dataset Characteristics

The dataset comprises 920 annotated recordings (.wav audio files sampled at 22050 Hz) obtained from 126 patients, with durations ranging from 10 to 90 seconds. These recordings add up to a total of 5.5 hours. The dataset encompasses a mix of clean respiratory sounds and noisy recordings designed to replicate real-world conditions. Patients across all age groups—children, adults, and the elderly—are represented in the dataset. For each patient, the dataset includes their diagnosis, demographic details and description of their respiratory cycle.

Every patient has been assigned a diagnosis by pulmonologists and cardiologists who reviewed the audio recordings. Table 1 shows the number of patients for each diagnosis category.

Diagnosis	**Number of patients**
COPD	64
Healthy	26
URTI (Upper respiratory tract infection)	14
Bronchiectasis	7
Bronchiolitis	6
Pneumonia	6
LRTI (Lower respiratory tract infection)	2
Asthma	1
Total	**126**

Table 1: Distribution of diagnosis among the 126 patients in the database

The dataset also includes demographic information for each patient, such as age, sex, adult body mass index (BMI) (in kg/m^2), child weight (in kg) and child height (in cm).

Furthermore, each audio file contains multiple respiratory cycles and every respiratory cycle has the following annotations: time of beginning of respiratory cycle(s), time of end of respiratory cycle(s), presence/absence of crackles in the respiratory cycle (Presence=1, Absence=0) and lastly presence/absence of wheezes in the respiratory cycle (Presence=1, Absence=0). There are a total of 6898 respiratory cycles in total. Among these cycles, 1864 include crackles, 886 include wheezes, and 506 include both crackles and wheezes. The mean duration of a respiratory cycle is 2.70 seconds, and the standard deviation is 1.17 seconds.

Methodology

The overall flow of data is shown in Figure 1. The audio segments are extracted from the raw audio data and processed into spectrograms in the form of RGB images, which are then sent as inputs to a Convolutional Neural Network (CNN) model for prediction.

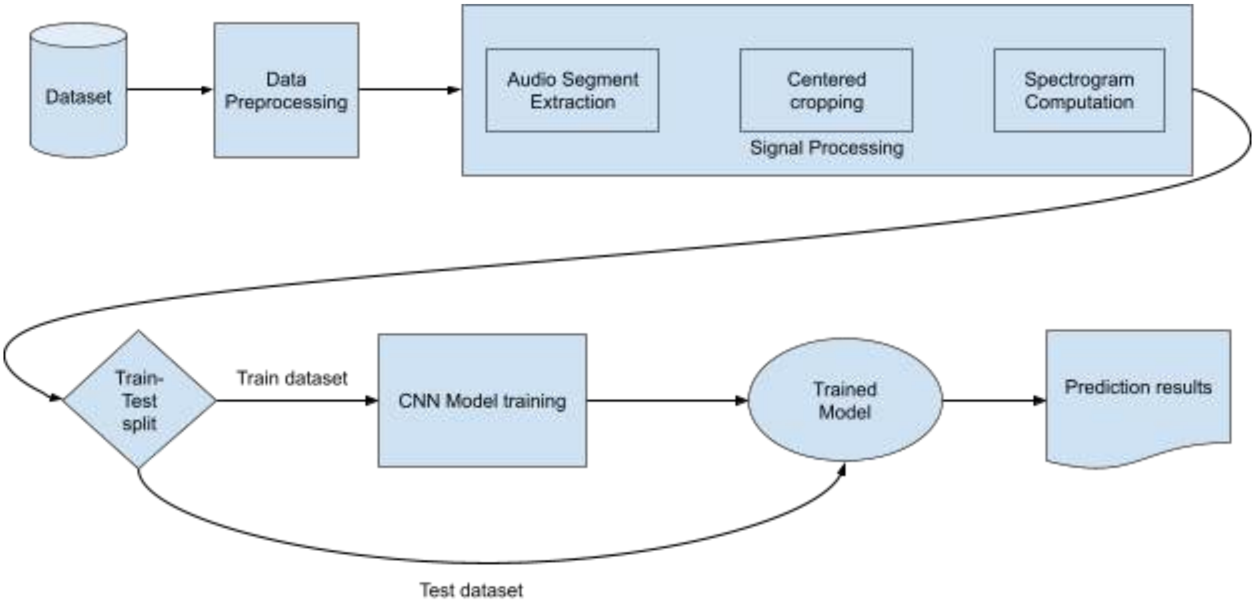


Figure 1: Flowchart of the Methodology

Data Preprocessing

Very few patients were diagnosed with LRTI (2 patients) and Asthma (1 patient). Hence, the audio files corresponding to these patients were excluded from the dataset.

In the demographics data, BMI values were not present for children. For such patients, if weight and height data was available, BMI was calculated as follows, as detailed in (Peterson et al.).

$$BMI = \frac{Weight\ (kg)}{Height^2\ (m^2)}$$

Some features had missing values. For features containing categorical values (“Sex” feature), the missing values were replaced with the mode of the feature values over all patients. For features containing numerical values (“Age” and “BMI” features), the missing values were replaced with the mean of the feature values over all patients. The “Age” and “BMI” feature values were standardized by removing the mean and scaling to unit variance.

Feature Engineering - Signal Processing

To increase the size of the dataset, a sample was defined as the audio signal contained in a single respiratory cycle, from which spectrograms were extracted and sent to the Machine Learning model for disease prediction.

Each audio file contains multiple respiratory cycles, and each respiratory cycle was extracted using its beginning and end times listed in the annotation data. The respiratory cycles varied in duration, with a mean of 2.70 ± 1.17 seconds. These audio segments were cropped to have the duration of the shortest respiratory cycle found across all recordings, so as to generate spectrograms of equal duration. This was accomplished by first locating the center of the segment and then cropping on both ends.

STFT Spectrogram

For each respiratory cycle, the Short-time Fourier Transform (STFT) spectrogram was computed using the following parameters:

- FFT size = 2048
- Number of audio samples between adjacent frames (hop length) = 512
- Window length = 2048
- Hann window

This resulted in a spectrogram with a frequency and time resolution of approximately 10.77Hz and 93ms, respectively. The resulting STFT matrix was uniformly converted to a 500× 200 pixel spectrogram image (Figure 2).

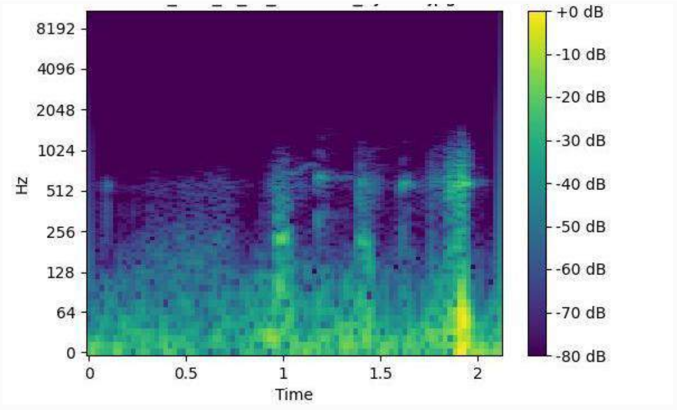


Figure 2: Sample STFT spectrogram image

MFCC Spectrogram

For each respiratory cycle, the Mel-frequency Cepstral Coefficients (MFCC) spectrogram (Abdul and Al-Talabani) was computed by using the following parameters:

- Number of MFCCs = 40
- FFT size = 2048
- Number of audio samples between adjacent frames (hop length) = 512

- Window length = 2048
- Hann window

The MFCC matrix was uniformly converted to a 500×200 pixel spectrogram image (Figure 3).

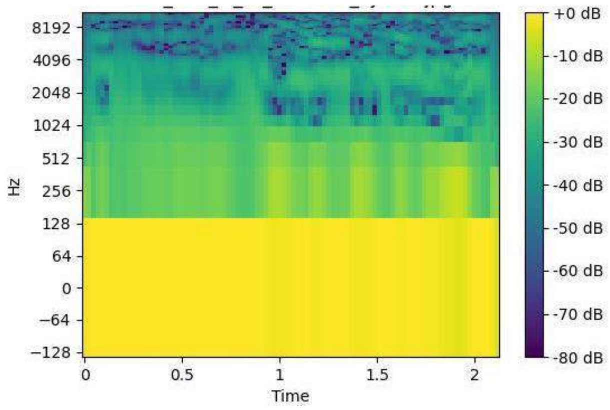


Figure 3: Sample MFCC spectrogram image

Train-Test split

A total of 6799 spectrogram images were generated after splitting the audio files into individual respiratory cycles (Table 2).

Diagnosis	**Number of spectrogram images**	
COPD	5685	83.6%
Healthy	322	4.7%
Pneumonia	285	4.2%
URTI	243	3.6%
Bronchiolitis	160	2.4%
Bronchiectasis	104	1.5%
Total	**6799**	**100%**

Table 2: Number of Spectrogram images per diagnosis

As shown in Table 2, there is a large imbalance in the distribution of the target classes. Therefore, Stratified Sampling was used when splitting the spectrogram images into train, validation, and test datasets to preserve the relative class frequencies.

The input dataset (spectrogram images) was first split into Train+Validation (80%) and Test (20%) datasets. The Train+Validation dataset was further split into Train (80%) and Validation (20%) datasets (Table 3).

Diagnosis	**Number of spectrogram images**			
	Train dataset	**Validation dataset**	**Test dataset**	**Percentages**
COPD	3638	910	1137	~83.6%
Healthy	206	52	64	~4.7%
Pneumonia	182	46	57	~4.2%
URTI	155	39	49	~3.6%
Bronchiolitis	103	25	32	~2.4%
Bronchiectasis	67	16	21	~1.5%
Total	**4351**	**1088**	**1360**	**100%**

Table 3: Spectrogram images in the various datasets

Models

A Convolutional Neural Network (CNN) (Krizhevsky et al.) is a class of neural networks that specializes in finding patterns in images and can be effective for classifying audio, time-series, and signal data.

Simple CNN Model

A Simple CNN model with the architecture shown in Figure 4 was defined.

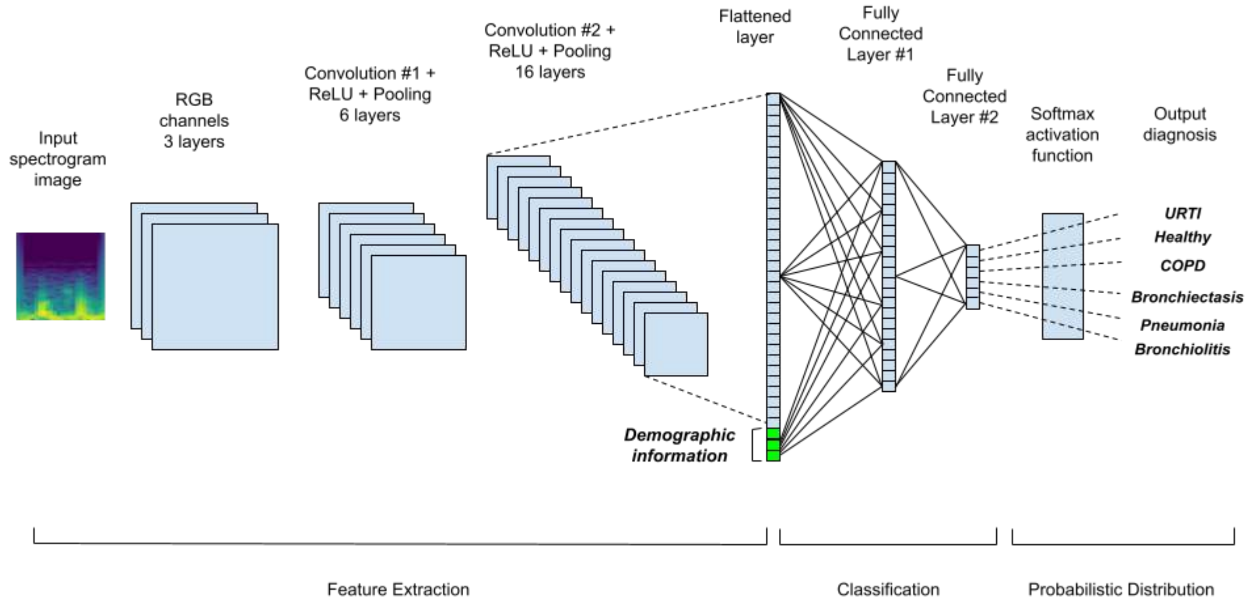


Figure 4: Simple CNN model architecture

For both Convolutional layers, a kernel size of 5 x 5 with stride of 1 and padding of 0 was used. The Rectified Linear Unit (ReLU) non-linearity layers were placed directly after the Convolutional layers to introduce non-linearity to the activation map. For both Pooling layers, max pool operation with a kernel size of 2, stride of 2 and zero padding was used. The first Fully Connected Layer was configured to produce 120 output features, the second Fully Connected Layer to produce 84 output features, and the third Fully Connected Layer to produce 6 features corresponding to the 6 diagnosis classes. The model was trained using Cross Entropy (Mao et al.) as the loss function, and Stochastic Gradient Descent (SGD) (Ruder) as the optimizer with a learning rate of 0.001 and momentum of 0.9.

RESNET Model

The ResNet-50 pre-trained model from “Deep Residual Learning for Image Recognition” paper was built using the ResNet50_Weights.IMAGENET1K_V2 weights and fine-tuned using the spectrogram data. The last Fully Connected layer in the model was replaced with a Fully Connected layer that produced 6 features corresponding to the 6 diagnosis classes. This last Fully Connected layer was made trainable, while all other model parameters were frozen. The model was trained using Cross Entropy (Mao et al.) as the loss function, and Adam (Kingma and Ba) as the optimizer with a learning rate of 0.01.

Addition of Demographic Information

For each patient, a 3x1 demographic vector was built containing the demographic information (Age, Sex and BMI). The output of the first Fully Connected Layer that produced

120 features was extended with this demographic vector resulting in a total of 123 output features. These 123 features were input to the second Fully Connected Layer to produce 84 output features. The third Fully Connected Layer produced 6 features corresponding to the 6 diagnosis classes.

Modeling Experiments

A variety of experiments were conducted to determine the optimal modeling procedure. Several modeling factors were considered:

1. Spectrogram type
    a. STFT
    b. MFCC
2. Sampling method (to address data class imbalance)
    a. Random downsampling
    b. Weighted random sampling
3. Model architecture
    a. Simple model
    The Simple model was built as described in the Simple CNN Model section.
    b. Model with hyperparameter tuning
    Using Ray Tune, a tool to perform hyperparameter tuning at scale, the hyperparameters that resulted in the best CNN model performance were determined and the optimized model was used for prediction.
    c. RESNET model
    The RESNET model used was as described in the RESNET Model section.
4. Demographic information
Experiments were performed to check how the inclusion of patient demographic information into the models affected the performance and accuracy of diagnosis.
    a. Without demographic information
    b. With demographic information

The experiments were organized in a tree-like fashion (Fig. 5), with each factor listed above acting as a point of bifurcation. An experiment could be discontinued at any point in the modeling stage if the results were deemed to be too poor.

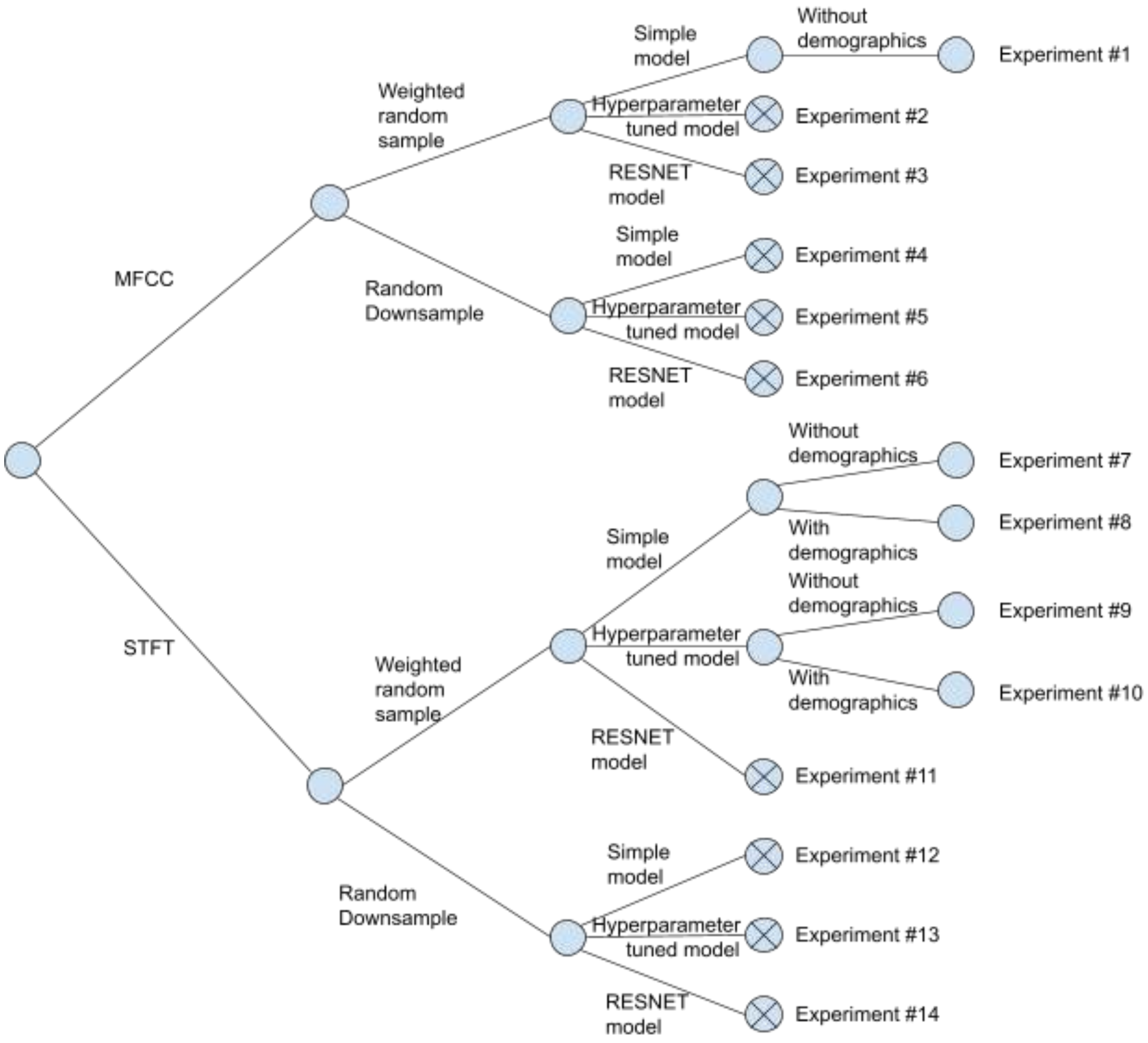


Figure 5: Flow chart of modeling experiments

Nodes = data/results. Branches = choice of methodology. Cross marks = termination of experiments.

Hyperparameter Tuning

Hyperparameter tuning was performed with the following search space for the various CNN Model parameters (Table 4).

Parameter	**Values**
Number of Fully Connected Layers #1	64, 128, 256, 512

Number of Fully Connected Layers #2	64, 128, 256, 512
Learning Rate	Log-uniform distribution between 0.001 and 0.1
Image batch size	2, 4, 8, 16

Table 4: Parameters used for hyperparameter tuning

Model Training and Evaluation

The spectrogram images were split into Train+Validation (80%) and Test (20%) datasets. The Train+Validation dataset was further split into Train (80%) and Validation (20%) datasets. The Model was trained for multiple epochs. In each epoch, each of the spectrogram images in the Train dataset was input to the Model to train the Model and a loss value was computed using the Cross Entropy Loss function (Mao et al.) for that image. “Train loss” was computed as the average of the loss values of all the images in the Train dataset. Then each of the images in the Validation dataset was input to the Model and a loss value was computed for that image. “Validation loss” was computed as the average of the loss values of all the images in the Validation dataset. It was observed that the Validation loss reached its minimum value within 20 epochs in all experiments, and hence all experiments were run for 20 epochs. The trained Model with the minimum “Validation loss” was saved as the best Model for that experiment.

Each of the spectrogram images in the Test dataset was input to the saved best Model and the diagnosis predicted by the model was compared against the actual diagnosis of the patient. A Classification report with Precision, Recall, F1-score and Accuracy metrics was computed. These four metrics were used to determine the performance of the Model in each experiment.

Results

The parameters used for the various experiments are shown in Table 5.

Experiment	**Spectrogram type**	**Sampling type**	**Model**	**Demographics**	**# of FC Layer 1**	**# of FC Layer 2**	**Learning rate**
1	MFCC	Weighted Random Sampling	Simple CNN Model	Not included	120	84	0.001
2	MFCC	Random DownSampling	Simple CNN Model	Not included	120	84	0.001

3	STFT	Weighted Random Sampling	Simple CNN Model	Not included	120	84	0.001
4	STFT	Weighted Random Sampling	Simple CNN Model	Included	120	84	0.001
5	STFT	Weighted Random Sampling	Hyperparameter tuned Model	Not included	64	128	0.001180
6	STFT	Weighted Random Sampling	Hyperparameter tuned Model	Included	131	512	0.002543
7	STFT	Weighted Random Sampling	RESNET Model	Not included	-	-	0.01
8	STFT	Random DownSampling	Simple CNN Model	Not included	120	84	0.001
9	STFT	Random DownSampling	RESNET Model	Not included	-	-	0.01

Table 5: Parameters used for the experiments

The results of all the experiments are shown below. Each experiment contains the performance metrics, curve of Training and Validation losses and the confusion matrix.

Experiment #1

	Precision	Recall	F1-score	Support
URTI	0.35	0.57	0.43	49
Healthy	0.58	0.33	0.42	64
COPD	0.95	1.00	0.97	1137
Bronchiectasis	0.83	0.24	0.37	21
Pneumonia	0.50	0.11	0.17	57
Bronchiolitis	0.41	0.41	0.41	32
Accuracy			0.89	1360
Macro average	0.60	0.44	0.46	1360
Weighted average	0.88	0.89	0.87	1360

Table 6: Performance metrics for Experiment #1

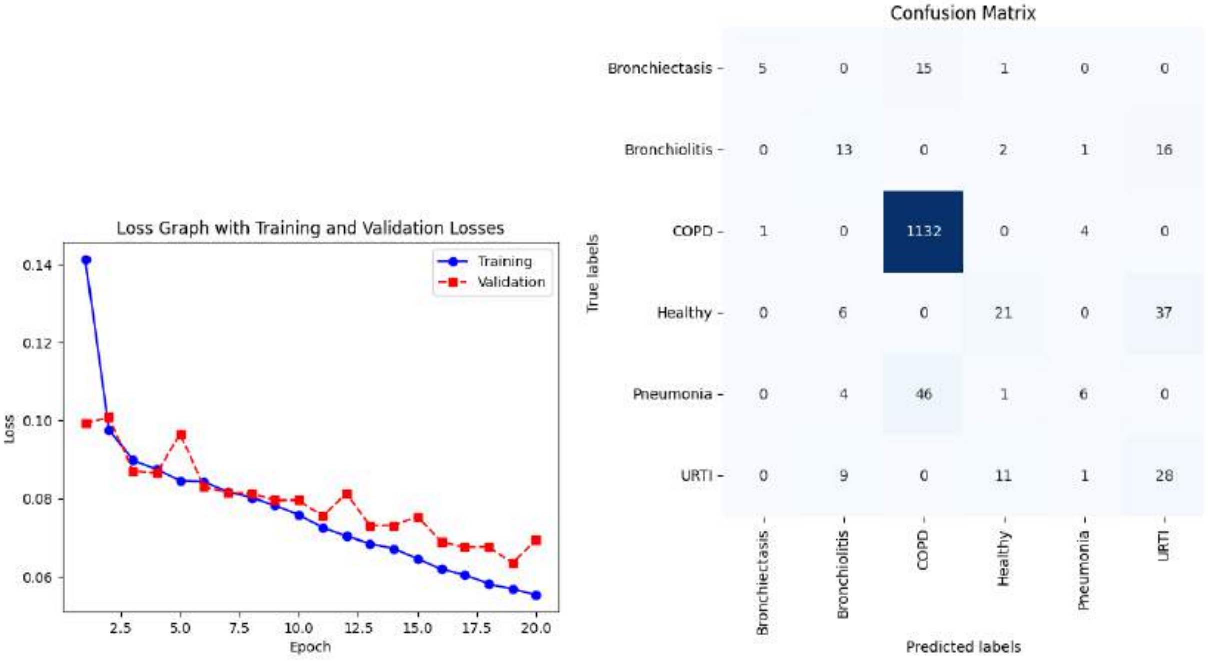


Figure 6: Training curve and Confusion matrix for Experiment #1

Experiments #2 to #6
Results were deemed too poor for further analysis.
Experiment #7

	Precision	Recall	F1-score	Support
URTI	0.52	0.47	0.49	49
Healthy	0.50	0.58	0.54	64
COPD	0.97	0.99	0.98	1137
Bronchiectasis	0.94	0.76	0.84	21
Pneumonia	0.78	0.56	0.65	57
Bronchiolitis	0.72	0.56	0.63	32
Accuracy			0.92	1360
Macro average	0.74	0.65	0.69	1360
Weighted average	0.92	0.92	0.91	1360

Table 7: Performance metrics for Experiment #7

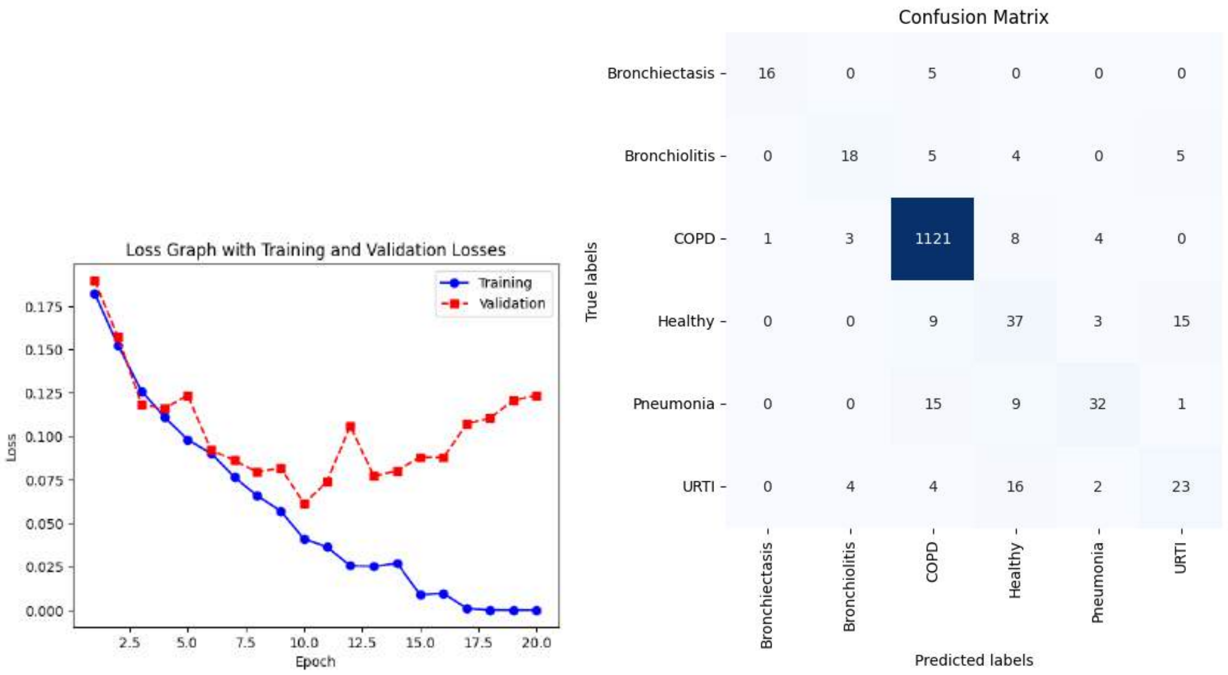


Figure 7: Training curve and Confusion matrix for Experiment #7

Experiment #8

	Precision	Recall	F1-score	Support
URTI	0.60	0.55	0.57	49
Healthy	0.65	0.83	0.73	64
COPD	0.98	0.99	0.99	1137
Bronchiectasis	1.00	0.76	0.86	21
Pneumonia	0.82	0.63	0.71	57
Bronchiolitis	0.86	0.59	0.70	32
Accuracy			0.94	1360
Macro average	0.82	0.73	0.76	1360
Weighted average	0.94	0.94	0.94	1360

Table 8: Performance metrics for Experiment #8

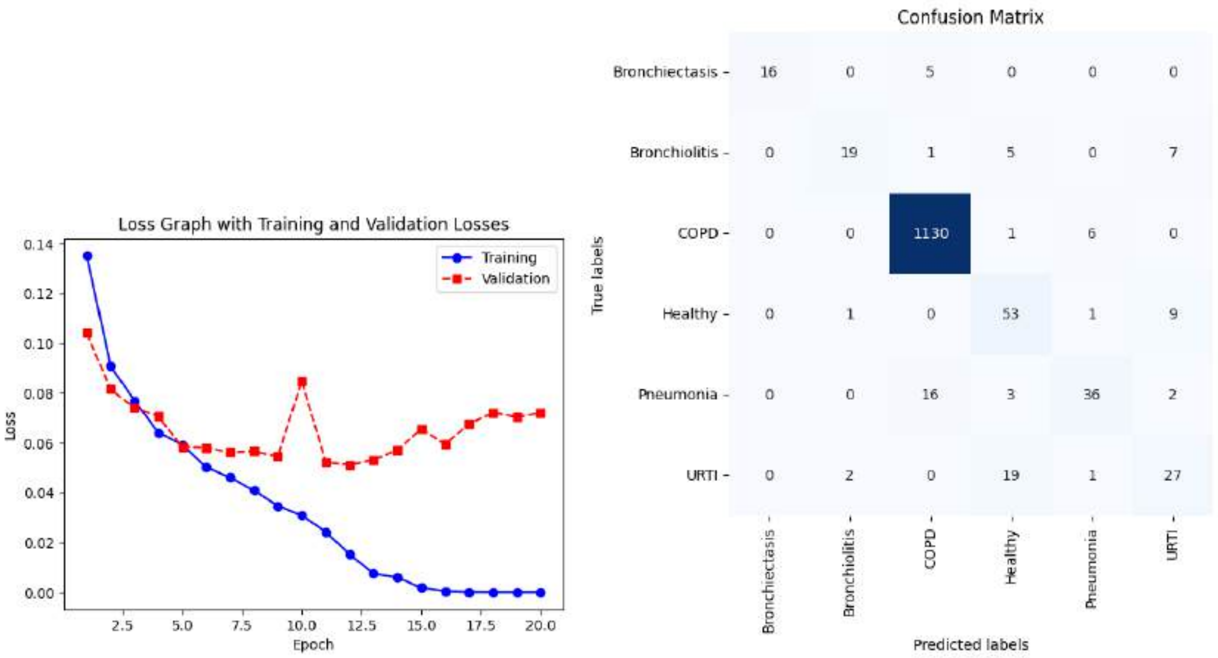


Figure 8: Training curve and Confusion matrix for Experiment #8

Experiment #9

	Precision	Recall	F1-score	Support
URTI	0.45	0.78	0.57	49
Healthy	0.67	0.38	0.48	64
COPD	0.96	0.98	0.97	1137
Bronchiectasis	0.88	0.67	0.76	21
Pneumonia	0.62	0.54	0.58	57
Bronchiolitis	0.91	0.31	0.47	32
Accuracy			0.91	1360
Macro average	0.75	0.61	0.64	1360
Weighted average	0.91	0.91	0.90	1360

Table 9: Performance metrics for Experiment #9

Experiment #10

	Precision	Recall	F1-score	Support
URTI	0.54	0.63	0.58	49
Healthy	0.67	0.70	0.69	64
COPD	0.96	1.00	0.98	1137
Bronchiectasis	0.88	0.71	0.79	21
Pneumonia	0.90	0.16	0.27	57
Bronchiolitis	0.77	0.62	0.69	32
Accuracy			0.92	1360
Macro average	0.79	0.64	0.67	1360
Weighted average	0.92	0.92	0.91	1360

Table 10: Performance metrics for Experiment #10

Experiments #11 to #14

Results were deemed too poor for further analysis.

Discussion

The Macro average F1-score was used as the metric to compare the performance of various experiments. A F1-score of 0 was used for experiments whose results were deemed too poor. The F1-score of a node in the experiment tree was calculated as the average of the F1-scores of all the experiments stemming from that node. The resulting F1-scores for various experiment nodes are shown in Figure 11.

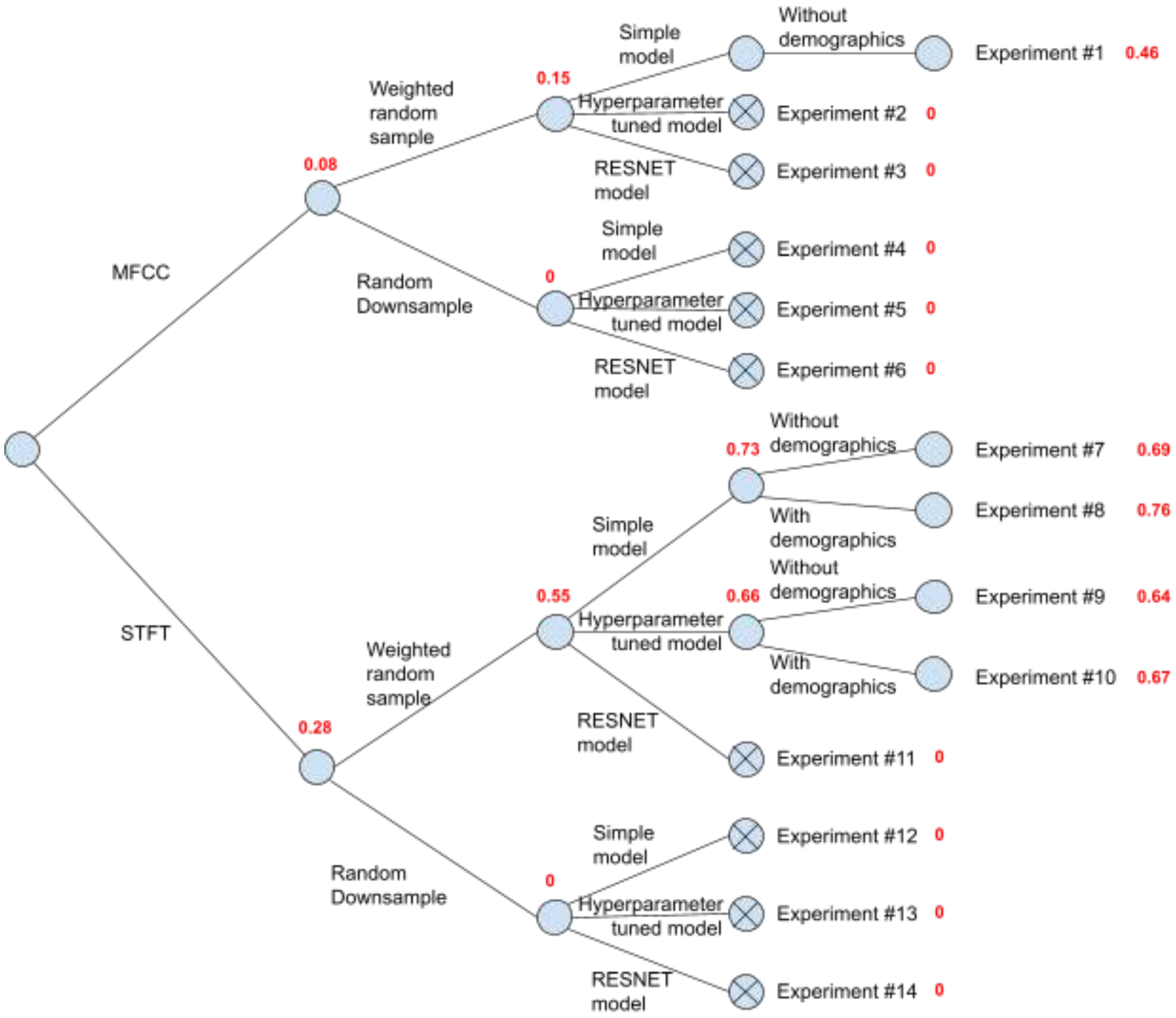


Figure 11: F1-scores for different experiment nodes

It is clear from Figure 11 that the best results are found in experiments 7 - 10. These results point to the following four conclusions: 1) STFT spectrograms work better than MFCC spectrograms, 2) Weighted Random Sampling works better than Random Down Sampling, 3) The Simple CNN Model performed better than both the hyperparameter-tuned Model and the pre-trained RESNET models, and 4) Model performance is enhanced with the inclusion of the demographic information.

Some potential interpretations and explanations of these results are offered here: 1) The MFCC is known for characterizing human speech vocalizations (Abdul and Al-Talabani) but may not generalize as well to other sounds as the STFT, hence the significant difference in performance. 2) Weighted random sampling is superior to Random downsampling by ensuring equal representations of samples from different classes in training batches, resulting in much better performance. 3) The Simple CNN out-performed the pretrained and hyperparameter optimized models, likely due to the small size of the dataset and the fact that spectrograms are not normally found in data used for pretraining foundational CNN models. 4) The inclusion of demographic information resulted in a lift in model performance for both Experiment #10 (vs. Experiment #9) and Experiment #8 (vs. Experiment #7). This clearly indicates that this additional information is useful in this disease classification context.

These conclusions provide a set of comprehensive modeling guidelines for achieving superior model performance with this dataset.

The best performing experiment was Experiment #8 with the Model parameters shown in Table 11.

Parameter	**Value**
Spectrogram type	STFT
Sampling type	Weighted Random Sampling
Model	Simple CNN model
Demographics	Included
Fully Connected Layer #1	120 layers
Fully Connected Layer #2	84 layers
Learning Rate	0.001

Table 11: Experiment #8 Model parameters

The average F1-score for various diagnosis types across all experiments are shown in Table 12.

Diagnosis	**F1-score**	**# of samples**
URTI	0.53	49
Healthy	0.57	64
COPD	0.98	1137
Bronchiectasis	0.72	21

Pneumonia	0.48	57
Bronchiolitis	0.58	32

Table 12: Average F1-score for various diagnosis types across all experiments

As shown in Table 12, the performance was the best for the COPD class, which contains the highest number of samples. The most difficult disease to classify is Pneumonia, followed by Bronchiolitis, and Bronchiectasis. The lower performance of the other classes is likely due to poor representation resulting from their small sample sizes. This is a fundamental limitation of the current study. However, the findings of this study in terms of modeling strategies still hold important value and will be applicable to larger, more representative datasets.

Conclusion

Respiratory illnesses are a major source of mortality and morbidity worldwide and it is imperative to diagnose them accurately and in a timely fashion. Machine Learning has the potential to make automated and accurate diagnostic predictions using respiratory sounds. This study explored various deep learning models and modeling techniques to classify respiratory diseases using audio recordings from the lungs. Annotated audio recordings of respiratory sounds from multiple patients were subject to different signal processing techniques to convert them to spectrogram images, which were used to train CNN models. It was found that STFT spectrograms performed better than MFCC spectrograms. The Simple CNN model with parameters as shown in Table 11 performed the best in identifying the diagnosis. Furthermore, inclusion of patient demographic information into the CNN model enhanced the model performance. In the future, model performance may be improved by training on larger datasets that are balanced across classes. With access to more GPUs, experiments with different epochs and batch sizes can be performed to identify optimal model parameters. In summary, this study successfully applied machine learning to predict respiratory diseases using audio recordings and also documented the optimal modeling approach.

Acknowledgements

This endeavor would not have been possible without the excellent guidance and mentorship of Joe Xiao throughout the research process. I am deeply indebted to him for his invaluable insights and continuous encouragement.

Works Cited

Abdul, Zrar Kh., and Abdulbasit K. Al-Talabani. "Mel Frequency Cepstral Coefficient and Its Applications: A Review." *IEEE Access*, vol. 10, 2022, pp. 122136–58. *DOI.org (Crossref)*, https://doi.org/10.1109/ACCESS.2022.3223444.

Albert Richard K., et al. "Azithromycin for Prevention of Exacerbations of COPD." *New England Journal of Medicine*, vol. 365, no. 8, 2011, pp. 689–98. *Taylor and Francis+NEJM*, https://doi.org/10.1056/NEJMoa1104623.

Alqudah, Ali Mohammad, et al. "Deep Learning Models for Detecting Respiratory Pathologies from Raw Lung Auscultation Sounds." *Soft Computing*, vol. 26, no. 24, Dec. 2022, pp. 13405–29. *Springer Link*, https://doi.org/10.1007/s00500-022-07499-6.

Arts, Luca, et al. "The Diagnostic Accuracy of Lung Auscultation in Adult Patients with Acute Pulmonary Pathologies: A Meta-Analysis." *Scientific Reports*, vol. 10, no. 1, Apr. 2020, p. 7347. *www.nature.com*, https://doi.org/10.1038/s41598-020-64405-6.

Association, American Lung. *Learn About Pneumonia*. https://www.lung.org/lung-health-diseases/lung-disease-lookup/pneumonia/learn-about-pneumonia. Accessed 25 May 2024.

Aston, Stephen J. "Pneumonia in the Developing World: Characteristic Features and Approach to Management." *Respirology*, vol. 22, no. 7, 2017, pp. 1276–87. *Wiley Online Library*, https://doi.org/10.1111/resp.13112.

Bhutta, Zulfiqar A. "Dealing with Childhood Pneumonia in Developing Countries: How Can We Make a Difference?" *Archives of Disease in Childhood*, vol. 92, no. 4, Apr. 2007, pp. 286–88. *PubMed Central*, https://doi.org/10.1136/adc.2006.111849.

Celli, Bartolome R., et al. "Differential Diagnosis of Suspected Chronic Obstructive Pulmonary Disease Exacerbations in the Acute Care Setting: Best Practice." *American Journal of Respiratory and Critical Care Medicine*, vol. 207, no. 9, pp. 1134–44. *PubMed Central*, https://doi.org/10.1164/rccm.202209-1795CI.

COPD - Symptoms | NHLBI, NIH. 25 Oct. 2023, https://www.nhlbi.nih.gov/health/copd/symptoms.

FFT. https://www.nti-audio.com/en/support/know-how/fast-fourier-transform-fft. Accessed 26 May 2024.

GOLD-2021-Chapter-1.Pdf. https://goldcopd.org/wp-content/uploads/2021/05/GOLD-2021-Chapter-1.pdf. Accessed 25 May 2024.

Graham, Stephen M., et al. "Challenges to Improving Case Management of Childhood Pneumonia at Health Facilities in Resource-Limited Settings." *Bulletin of the World Health Organization*, vol. 86, no. 5, May 2008, pp. 349–55. *PubMed Central*, https://doi.org/10.2471/BLT.07.048512.

He, Kaiming, et al. *Deep Residual Learning for Image Recognition*. arXiv:1512.03385, arXiv, 10 Dec. 2015. *arXiv.org*, https://doi.org/10.48550/arXiv.1512.03385.

Htun, Tha Pyai, et al. "Clinical Features for Diagnosis of Pneumonia among Adults in Primary

Care Setting: A Systematic and Meta-Review." *Scientific Reports*, vol. 9, May 2019, p. 7600. *PubMed Central*, https://doi.org/10.1038/s41598-019-44145-y.

Huang, Dong-Min, et al. "Deep Learning-Based Lung Sound Analysis for Intelligent Stethoscope." *Military Medical Research*, vol. 10, no. 1, Sept. 2023, p. 44. *BioMed Central*, https://doi.org/10.1186/s40779-023-00479-3.

Kingma, Diederik P., and Jimmy Ba. "Adam: A Method for Stochastic Optimization." *arXiv.Org*, 22 Dec. 2014, https://arxiv.org/abs/1412.6980v9.

Krizhevsky, Alex, et al. "ImageNet Classification with Deep Convolutional Neural Networks." *Advances in Neural Information Processing Systems*, vol. 25, Curran Associates, Inc., 2012. *Neural Information Processing Systems*, https://proceedings.neurips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html.

Liu, Li, et al. "Global, Regional, and National Causes of under-5 Mortality in 2000–15: An Updated Systematic Analysis with Implications for the Sustainable Development Goals." *Lancet (London, England)*, vol. 388, no. 10063, Dec. 2016, pp. 3027–35. *PubMed Central*, https://doi.org/10.1016/S0140-6736(16)31593-8.

Mao, Anqi, et al. *Cross-Entropy Loss Functions: Theoretical Analysis and Applications*. arXiv:2304.07288, arXiv, 19 June 2023. *arXiv.org*, https://doi.org/10.48550/arXiv.2304.07288.

McAllister, David A., et al. "Global, Regional, and National Estimates of Pneumonia Morbidity and Mortality in Children Younger than 5 Years between 2000 and 2015: A Systematic Analysis." *The Lancet Global Health*, vol. 7, no. 1, Jan. 2019, pp. e47–57. *www.thelancet.com*, https://doi.org/10.1016/S2214-109X(18)30408-X.

Momtazmanesh, Sara, et al. "Global Burden of Chronic Respiratory Diseases and Risk Factors, 1990–2019: An Update from the Global Burden of Disease Study 2019." *eClinicalMedicine*, vol. 59, May 2023. *www.thelancet.com*, https://doi.org/10.1016/j.eclinm.2023.101936.

Mulupi, Stephen, et al. "What Are the Barriers to the Diagnosis and Management of Chronic Respiratory Disease in Sub-Saharan Africa? A Qualitative Study with Healthcare Workers, National and Regional Policy Stakeholders in Five Countries." *BMJ Open*, vol. 12, no. 7, July 2022, p. e052105. *PubMed Central*, https://doi.org/10.1136/bmjopen-2021-052105.

Peterson, Courtney M., et al. "Universal Equation for Estimating Ideal Body Weight and Body Weight at Any BMI1." *The American Journal of Clinical Nutrition*, vol. 103, no. 5, May 2016, pp. 1197–203. *PubMed Central*, https://doi.org/10.3945/ajcn.115.121178.

Ruder, Sebastian. "An Overview of Gradient Descent Optimization Algorithms." *arXiv.Org*, 15 Sept. 2016, https://arxiv.org/abs/1609.04747v2.

Stokes, Katy, et al. "A Machine Learning Model for Supporting Symptom-Based Referral and Diagnosis of Bronchitis and Pneumonia in Limited Resource Settings." *Biocybernetics and Biomedical Engineering*, vol. 41, no. 4, Oct. 2021, pp. 1288–302. *ScienceDirect*, https://doi.org/10.1016/j.bbe.2021.09.002.

The Top 10 Causes of Death. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. Accessed 20 May 2024.

Israeli Judicial Reform and Soft Power Diplomacy By Daniella Biblin

Introduction

For the greater part of 2023, Israel has been characterized by turbulent debate surrounding the proposed and partially executed judicial reforms. The Israeli system of government is structured around three branches independent from each other by law: the legislative, executive, and judicial. The Supreme Court is based on 15 judges appointed by the Judicial Selection Committee, and has had the ability to act independently while enforcing legal checks on other branches ("Israel Government & Politics"). Beginning in January of 2023, however, the Israeli Knesset introduced plans to reform the political system through a modified power balance between the judiciary and government. The first of five proposals, the "reasonableness" bill, was passed in July to officially remove the power of the judiciary to review and cancel decisions made by the executive branch that they deem "unreasonable" ("Israel's 'Reasonableness' Legislation"). The remaining proposals would allow Knesset to overrule judicial decisions with votes from 61 out of 120 members (Fuchs, Amir). Thousands of Israelis have gathered to protest reforms since January, with up to 160,000 gathered at several demonstrations, and the debates have polarized Israel more than ever before (Mellen, Ruby, et al).

While the October 7th attacks and subsequent war between Israel and Hamas have evidently taken over recent headlines, the conflict is critical in understanding both sides of the reforms. Given the horrific events of the conflicts, the connection between reforms and Israel's ability to achieve its national objectives through soft power has become increasingly pressing. The political changes and controversy have put established political concepts and theories to the test, including the impacts of a major domestic overhaul by a sovereign state on its international relationships. As exemplified through the ongoing war and Israel's lasting global stance, Israel is highly dependent on leveraging soft power to achieve its military, economic, and political objectives. Because international image is a driving force behind state-to-state relations, government decisions have impacts far beyond Israeli borders. While reforms might be advantageous for circumstances in Israel, limited power for each branch is a democratic principle itself. If the changes undermine the democracy that foreign states evidently value, to what extent will the 2023 Israeli judicial reform undermine Israel's ability to leverage soft power diplomacy?

I became interested in this topic because much of my family lives in Israel, and recent conversations with relatives overseas have been related to the widely discussed and debated reforms. With relatives in support and against reforms, I have heard public responses to the changes, as well as different perspectives on its broader implications during a conflicted time. For such a domestically and internally oriented reform, it has been interesting to see core arguments from an international perspective, such as whether reforms threaten Israeli relations with allies and its position on the global stage. My engagement activities were focused on exploring the intersection between soft power and changing political systems, specifically through contrasting views on the issue. For the first part of my engagement, I listened to a live recording of a Knesset hearing, which was a discussion between several members of Knesset. These discussions provided insight into a perspective against reforms, as members expressed concerns and felt as though the reforms threaten basic democratic principles. Next, I interviewed a practicing attorney and legal advisor for the Knesset, who has first-hand witnessed disagreements surrounding reforms, and favors increased power within the parliament. To better understand crucial differences between these two parts, I will use political ideas including legitimacy and soft power in evaluating the first perspective, while using sovereignty and structural realism for the second.

Soft power and legitimacy

Over the course of debates surrounding judicial reforms, a widespread concern has been stability and legitimacy. Many political leaders, and much of the Israeli population, argue that domestic turbulence weakens the perception that allies have of a sovereign state. While listening to Knesset dialogue for the first part of my engagement, reforms were discussed as a sign of disorganization, with members referring to the changes as a “radical upheaval” and “disruption.” When considering global implications, specifically through a soft power approach, the relationship between the U.S. and Israel is a central point. Knesset members explained that Israel is the only truly democratic and highly stable country in the Middle East, which has allowed Israel to develop diplomatic partnerships unlike any other country in the region. They argue that the government is uprooting a system that is well-established and accepted as legitimate by its strongest allies. In March of 2023, President Biden expressed his remarks on the reforms, claiming that he was “very concerned” about the health of the Israeli democracy and that “they cannot continue down this road” (“Remarks by President Biden”).

Several members believe that reforms undermine soft power diplomacy by challenging political ideologies and an institution that has been seen as legitimate for years. Members discussed the possibility that global powers including the U.S. may be reluctant to provide the same financial support and military aid to a nation that is unstable and undermines democratic principles, such as balanced executive power. One member claimed that, if Israel continues, they are undermining decades of U.S. support in building “the only democracy in the Middle East.” This perspective emphasizes that reforms threaten the foreign relationship Israel values the most; members mentioned that “Israel is almost entirely dependent on the U.S. and its allies” with the U.S. providing Israel with over $130 billion in bilateral aid since its founding (“U.S. Security Cooperation with Israel”). They further communicated that reforms will inevitably “weaken leverage that Israeli political leaders have at the negotiation table, making the job of leaders much harder.”

In an increasingly globalized world, soft power has become undoubtedly connected to state legitimacy and international perceptions. In contrast to hard power, soft power depends on diplomatic relations with strong allies. The primary arguments used against reforms have relied on soft power theories; a nation will only have the ability to leverage soft power if its ideology and institutions are accepted by other global powers. One way to better understand this perspective is through political scientist Joseph Nye, who defines soft power as the ability to achieve global power objectives through non-coercive means. Nye argued that truly effective ways to create change, either national or global, are through attraction and legitimacy. He refers to intangible power resources, "cultural attraction, ideology, and international institutions," as the most powerful policies (Nye, Joseph S). Members from the hearing reflected a soft power approach by prioritizing legitimacy through stability and democratic principles. Similar to ideas from Nye, members communicated that nations only achieve objectives if their power is perceived as legitimate on an international scale.

Sovereignty and structural realism

In contrast to Knesset discussions, the interview I conducted with a Knesset attorney indicated that reforms would only strengthen Israeli democracy in the long-term, without threatening soft power diplomacy. When asked about the impacts of reforms on democratic principles, she explained that, in reality, reforms move Israel closer towards a true democracy. She believes Israel has diverged away from its democratic roots, largely because of the structure and governance of the Supreme Court. According to the interview, a democracy relies on the will of the people, but the current system allows "15 highly political judges who are chosen by a committee of 9 highly political figures" to override the will of the people. The Knesset, which is elected by the people and is responsible for representing the people, should have the final say on the law of the land, rather than an exclusive group of unelected political figures; "democracy strengthens a country, rule by a small group weakens a country."

Interestingly, an idea prevalent throughout the interview was sovereignty. While sovereignty is traditionally defined as the ability for a state to govern its territory and people, government sovereignty is important to consider. The attorney implied that, while Israel has attained sovereignty over its territory and people, strengthening executive government sovereignty will allow leadership to execute decisions and represent the majority of the population. As a result, government sovereignty is integral in returning to a democratic system in Israel. When considering the U.S. branches, which are structured around legal limitations and power separation, it is important to note that Supreme Court justices are chosen directly by the President and approved by the Senate, both of which are elected by the people ("The Judicial Branch"). Furthermore, justices are mandated to make decisions based on a Constitution with over two centuries of precedent, and Congress has the power to remove justices and override judicial decisions through amendments ("The Judicial Branch"). As a result, government sovereignty is just as foundational for the U.S. democracy.

After asking whether "domestic turbulence" and "radical reforms" would undermine Israel's diplomatic relations overseas, the attorney stated that "the risk is outweighed by a strengthened democracy and legitimate decision-making." She explained that opposing the reforms is equivalent to granting the judiciary, which is self-selected and isolated from the people, to unchecked power. The reforms prevent countless more arbitrary judicial decisions through a power balance that only allows the judiciary to vote against truly unreasonable legislation. In the long-term, a redesigned democracy will be admired and respected internationally, which actually aligns with Nye's soft power theory. When asked about concerns from Knesset discussions, specifically foreign reactions to reforms based on international views of an effective democracy, the interviewer pointed out that Israeli relations during the October 7th attacks and ongoing war have been unaffected by reforms. In terms of aid and partnerships, reforms have not emerged in any negotiations with international allies. Instead, it has been evident that geopolitical interests have far outweighed perceptions on domestic reforms.

An idea the interviewer also emphasized was national progress. The attorney referred to reforms as "necessary" for Israel to progress towards a stronger nation. Considering the idea of progress through political theories, the interview reflected aspects of structural realism rather than liberalism. While Hans Morgenthau is a political scientist known for his theory on classical realism, his ideals are foundational to the realist approach. In his book, *Politics Among Nations*, Morgenthau wrote that "international politics, like all politics, is a struggle for power" (Morgenthau, Hans J). As the interview suggested, reforms are essential on a global stage based on "power politics." Instead of prioritizing international perceptions, this perspective views reforms as a way to restore an effective political system and execute decisions that can set the nation ahead. Without reforms, in the long-term Israel is much more likely to become threatened by a weaker country than any implications on soft power. Further focusing on political theories, neorealism or structural realism has been indirectly used to support reforms. Neorealism, as defined by theorist Kenneth Waltz, is the idea that states have no choice but to pursue greater power because of the international system (Waltz, Kenneth N).

Conclusion

Judicial reforms undeniably sparked political controversy within Israel, but also caught the attention of practically every global power around the world, especially Israeli allies. Soft power relies on ideological appeals, institutional legitimacy, international image, agreement on democratic values, and many more state-to-state factors; however, proponents of the reforms tend to prioritize a realist approach, through which the progressing strength of Israel surpasses short term implications on soft power. From this perspective, powerful nations are based on sovereign governance with the ability to make ultimate decisions for the state, which will naturally lead to more and stronger relations.

Israel's ability to leverage soft power diplomacy has likely been affected in negotiations and minor discussions with allies. The true effect on soft power is a complex balance between a multitude of factors, including the political climate, conflict at hand, and scale of proposed reforms. After conducting this investigation in the context of 2023, it seems that the scale of the conflict and adversaries outweighs potential implications of reforms on soft power. In reality, leaders must consider the relationship between diplomatic impacts of policies on global perception and long-term domestic benefits of the policies.

Works Cited

Biblin, Daniella, and Daniel Wolfson. “Attorney Interview.” 15 Dec. 2023.

Fuchs, Amir. “The Override Clause Explainer.” *The Israel Democracy Institute*, 11 Nov. 2022, en.idi.org.il/articles/46387.

“Israel Government & Politics: How Does the Israeli Government Work?” *Jewish Virtual Library*, www.jewishvirtuallibrary.org/how-does-the-israeli-government-work.

“Knesset Hearing Recording.” December 7, 2023. Translated from Hebrew to English by Daniella Biblin.

Mellen, Ruby, et al. “Protests Rocked Israel for 29 Consecutive Weeks. There’s More to Come.” *The Washington Post*, 24 July 2023, www.washingtonpost.com/world/2023/07/24/israel-protests-judicial-overhaul/.

Morgenthau, Hans J. “Political Power.” *Politics Among Nations The Struggle For Power And Peace*, Alfred A. Knopf, New York City, New York, 1948, pp. 13–18.

Nye, Joseph S. “Soft Power.” *Foreign Policy*, no. 80, autumn 1990, pp. 153–171, https://doi.org/10.2307/1148580.

“Remarks by President Biden Before Air Force One Departure.” *The White House*, The United States Government, 28 Mar. 2023, www.whitehouse.gov/briefing-room/speeches-remarks/2023/03/28/remarks-by-president-biden-before-air-force-one-departure-26/.

“The Judicial Branch.” *The White House*, The United States Government, www.whitehouse.gov/about-the-white-house/our-government/the-judicial-branch/.

“U.S. Security Cooperation with Israel.” *Bureau of Political-Military Affairs*, U.S. Department of State, 19 Oct. 2023, www.state.gov/u-s-security-cooperation-with-israel/.

Waltz, Kenneth N. “Structural Realism after the Cold War.” *International Security*, vol. 25, no. 1, summer 2000, pp. 5–41.

“What Is Israel’s ‘Reasonableness’ Legislation and Why Is It so Contentious?” *American Jewish Committee Global Voice*, 23 July 2023, www.ajc.org/news/what-is-israels-reasonableness-legislation-and-why-is-it-so-contentious.

21: Find common ground By Claire Suh

Tyranny has a way of pitting individuals against each other, and makes it hard to listen. In a system designed to quiet all opinions, it is easy to feel the urgent need to make yourself heard under such circumstances. Do not fall for it. Among the cacophony of voices willing themselves to be acknowledged, it yields a fragmented community unwilling to listen and unable to move forward and reject tyranny. Rather, listen to others, gather opinions, work together, and find common ground.

Dr. Sun-Yat Sen was a Chinese revolutionary, statesman, political philosopher, and founder of the Kuomintang party. Although he was not an active participant in the Xinhai Revolution, which ended the last imperial dynasty of China, he greatly stressed its necessity and fueled the spirit of its occurrence. He is associated as being the father of modern China, as he served as the first provisional president following the fall of the Qing dynasty. Despite the chaos following the lack of clear leadership, and the resulting period of fragmentation known as the Warlord Era, Dr. Sun-Yat Sen worked in collaboration with the Chinese Communist Party, the Soviet Union, and the nationalist party of China upon the common desire of Chinese unity and industrialization.

The necessity of finding common ground amongst different communities is as difficult as it is essential. Oftentimes, tyranny causes individuals to be discredited and infused with blatant propaganda in order to lull the sea of objections that may rise to the questionable actions of rule. Lost amongst such actions is the ability to think, as individuals are not taught how to process information, but rather are given large amounts of deemed facts to know as truth. In this way, any questions or differing approaches appear as foreign knowledge, dangerous, and threatening, and forms distinct segregation within communities. Although individuals under tyranny may maintain the same agenda, contrasting perspectives become a barrier as the tyrants work to achieve just that. Debilitation of thought, dissociation, and consequently an inability to resist.

The separation of communities is as chaotic as it is tactical. A move made by tyrants to ensure total control over whatever group or nation it is they chose. Take the Chinese Civil war and its primary leaders Chiang Kai-Shek and Mao Zedong, who encouraged their followers of the essential evil of the other side and the need for victory. Through propaganda enforced by both party leaders, individuals are driven to participate in the fight against one another, as they remain distinctly separate in animosity. The odd irony of such a combative approach is that both sides, aside from their political morals, had similar desires for the future of China, and were in collaboration up until the death of Dr. Sun-Yat Sen. What is observed here is the isolation of two groups with a common goal, restricted by the total control held by two different party leaders who are compelled by a competitive desire to outdo one another. Both without tolerance, they facilitate the spread of similar ideology to their followers to achieve this agenda.

Not only this, but the fragmentation of communities as a result of tyranny serves to prevent collective action towards its oppression. Such is the context of The Great Leap Forward, which caused widespread starvation throughout China and resulted in the breakdown of the Chinese economy. From this came forward division amongst the communist party and within the community, as some began to question the efficiency of a communist society. As later seen in the Cultural Revolution, Mao Zedong publically removed and punished those who rendered any doubt in their minds, attesting to the futile nature of independent thought. Individuals under tyranny cannot simply work by themselves to overthrow the presence of such power, but must come together to learn how to tear down its walls as one.

This being said, tyranny is not an immovable force. By the collective contributions to resistance for a common cause, individuals hold the power to reverse the actions of their tyrants and break through the walls of isolation from each other and from themselves. For example, Chiang Kai-Shek was forced to retreat into Taiwan with the Kuomintang party in 1949, where they have resided ever since. The communist party, who had been on the brink of defeat between 1934 and 1935, achieved a miracle turnaround, which is mostly hypothesized as possible due to their ability to better empathize and rally the support of the rural and otherwise general population. Thus, by the success of the Communist Party, the myth regarding the indestructible nature of tyranny is shattered upon the collaboration of many.

Perhaps the most valuable takeaway from communication, is the value of empathy and the strength entailed by human emotions. Know the priority. Just as Dr. Sun-Yat Sen did not allow his disagreements to translate into his actions for the sake of the bigger picture, do not get swept up by differences. Learn with it, grow with it, and let it work as a molding tool to build a bigger and stronger community than before. Resist tyranny by resisting the urge to reject thought, and take all perspectives into consideration. What is different is not necessarily dangerous, and it should be held in close account. Do not fall prey into the trap of ignorance, and allow yourself the opportunity to learn from others. Although resistance may seem a vague topic to act on, the first step is to form a community, hear others, form solutions, and make compromises. Build together, and do it together upon the common ground.

Works Cited

Britannica, The Editors of Encyclopaedia. "Great Leap Forward." Encyclopedia Britannica, 26 Feb. 2024, https://www.britannica.com/event/Great-Leap-Forward. Accessed 30 May 2024.

Lamb, Stefanie. "Introduction to the Cultural Revolution." *Stanford Program on International and Cross-Cultural Education*, Dec. 2005, spice.fsi.stanford.edu/docs/introduction_to_the_cultural_revolution. Accessed 30 May 2024.

Overview of Chinese History 1911 - 1949 | the 20th Century | World History | Khan Academy. *Khan Academy*, www.youtube.com/watch?v=a9QtIfPIQl4. Accessed 30 May 2024.

Wang, Yi Chu. "Sun Yat-sen." Encyclopedia Britannica, 10 Apr. 2024, https://www.britannica.com/biography/Sun-Yat-sen. Accessed 30 May 2024.

Entrepreneurship As Empowerment: How Women's Entrepreneurship Develops From Necessity to Better Their Societal Position By Marina Guzzi

According to the U.S. Small Business Administration, women business owners own over 12 million businesses in 2023, a number that would not be possible without the hard work of past female entrepreneurs in the late 19th and early 20th century. Entrepreneurship can be defined in many ways, but the most accurate is the relentless pursuit of new ideas to fill a need in the market. After the Civil War, American entrepreneurship began to flourish and continued expanding due to growth in communication and transportation. Westward expansion further amplified entrepreneurial opportunities through railroads, banking, and land acquisition, creating a new age of capitalists and innovators. This narrative, however, has largely excluded women, who played a crucial role in the development of American entrepreneurship, especially in the beauty and service sectors. By looking at the two primary motives behind starting a business- necessity or opportunity- it becomes clear which one women have experienced. Often subject to unequal wages or job offerings in white-collar fields, American women have found the most success in business ownership, though they start from necessity. Two people in particular, Madam C.J. Walker and Lydia Pinkham, embodied these ideals for themselves and others, using their entrepreneurship to improve women's quality of life. Walker, the first female self-made millionaire in American history, used her wealth to help others begin their own businesses. Pinkham, who sold a homemade remedy, bettered women's health and increased knowledge surrounding the female body. Therefore, although women usually start businesses out of necessity, female entrepreneurship in America has an impact beyond the money, as it strengthens communities, empowers women, and establishes a precedent for more to start their own businesses. Walker and Pinkham are pioneers of this philanthropic work, boosting women's confidence and financial independence.

Women have not always had this freedom in the workforce. From the colonial period to the American Revolution, women worked within the homes, caring for children, cooking, and making clothes for her family to wear. During the Industrial Revolution in the 1840s and 1850s, factory labor began to consume America, leading many women to join the workforce. They were a coveted option for employers, as factory owners could pay them much less than their male counterparts and had experience with some tasks like sewing. Still, only 15% of American women held a job outside their house by 1850, though that number was 10% in 1840. Despite their small numbers, though, women have always been trailblazers in the labor force. The Lowell Mills Girls were the first women to unionize and strike, doing so in the 1830s- decades before the mass movement for labor rights. They fought for higher wages, regulations on working hours, and safer working conditions. In the short term, they didn't win much (both of their two strikes were crushed, only won a ten-hour work day in New Hampshire, where it couldn't be easily enforced) their long-term legacy prevails. As one employee said, "They have at last learnt the lesson which a bitter experience teaches, not to those who style themselves their 'natural protectors' are they to look for the needful help, but to the strong and resolute of their own sex"

(AFL-CIO). The mill girls established that women did not need to tolerate injustice in the workplace. As pioneers in the women's labor movement, they set a standard for future advocacy, which Walker and Pinkham followed.

First, Madam C. J. Walker began her entrepreneurial endeavors due to necessity, as she had no formal education, did not make enough money to survive, and wanted to help other black women with hair loss. Orphaned at age seven, Walker soon became determined to improve her situation. She lived with her older sister, Louvenia, in Vicksburg, Mississippi, and married at fourteen, in part to escape Louvenia's abusive husband. Walker's husband died five years later, leaving her to raise their daughter alone. She worked as a laundress for eighteen years in St. Louis before moving to Denver with $1.50 in savings to start her own business. Walker was determined to earn enough to send her daughter to school, a privilege she did not receive. With only three months of formal education, options were limited for Walker, and her job as a laundress earned just over a dollar a day. The idea for a hair care business stemmed from her own baldness problem, an issue common in African American women due to poor diet, stress, illness. The current products often caused damage or scalp disease, induced by the harsh clothing cleaners from her laundry work. Most Black people in the late 1800s also lacked indoor plumbing, heating, and electricity, so bathing could be infrequent. While Walker saw a clear opportunity for a thriving business and took it, this was not a choice. She said in a speech to the National Negro Business League, "I am a woman who came from the cotton fields of the South. From there, I was promoted to the washtub. From there, I was promoted to the cook kitchen. And from there, I promoted myself into the business of manufacturing hair goods and preparations. I have built my own factory on my own ground."[1] She worked her way from nothing to an empire out of necessity. In order to provide her daughter with education, she had to remain determined and focused. This motivation is a reality for many other female entrepreneurs in America.

Like Walker, Lydia Pinkham was forced to start her business because of the little information available on women's health, the validity of current treatments, and her personal financial situation. The idea for a medicine arose from necessity as women had no place to turn to for information on their bodies, specifically pregnancy. Women's health was a taboo topic, leaving women uninformed on their pregnancy or uncomfortable asking additional questions. Nutritional knowledge was slim, frequently leading to chronic indigestion, fatigue, anemia, and food poisoning. Furthermore, consistent good health was rare for women in the 19th century, as germ theory had not yet been understood. Menstrual cramps, a common ailment, were solved through ovary removals, a procedure with a 40% mortality rate. Women noticed the scope of patients killed by doctors and widespread public dissatisfaction with physicians swept the nation. Gynecology, a new field in the 1870s, was acutely limited in how it could help: basic gynecological exams were subject to strict rules to protect women's virtue, meaning their clothes remained entirely on for the duration of the appointment. Pinkham thought that male doctors were careless and insensitive towards women's reproductive health, an issue that only a woman

[1] Records of the National Negro Business League, Part I, Annual Conference Proceedings and Organizational Records, National Negro Business League 1900-1919, Manuscript Division, Library of Congress, Washington, DC.

could understand.[2] Finally, an obsessive concern with women's weakness created an image of the surviving middle-aged women as frail and emotional. Many medical books at the time continued to invalidate their concerns, writing it off as "hysteria," a condition diagnosed from tantrums and erratic behavior. The common physician viewed women as incapable of giving birth.[3] Therefore, many women turned towards handcrafted remedies, consisting of roots, herbs, and alcohol (for preservation), which Pinkham was especially successful in producing them as family, friends, and neighbors turned to her to cure ailments- menstrual cramps, menopause, or pregnancy problems- free of charge. Pinkham was convinced that she was the reliable, comforting, and trustworthy source of medicinal help and advice that women needed. Pinkham's husband lost his job from the Panic of 1873, the first "Great Depression" which was caused by the rapid expansion of railroads. Currency was valued from metal and banks issued paper money backed by the supply of gold and silver. To finance the Civil War, President Abraham Lincoln began to print paper money called "Greenbacks," allowing rapid railroad growth and speculation from bonds sold by banks. Soon, construction costs heavily increased, faster than the financing could keep up. Banks that had sold bonds failed, igniting withdrawals and other firms and industries to collapse while the unemployment rate rose to 14% in 1876.[4] Her husband, Isaac, was sued and almost arrested for his inability to repay his mortgage, leaving him unable to work. Pinkham's family needed to financially recover, so when a friend came knocking on the door offering money for her Vegetable Compound, her business was created. Not only out of financial necessity, but out of necessity for the female gender's well being.

Though the Madam C. J. Walker Manufacturing Company only sold hair products, it impacted the community in ways beyond that, such as providing business opportunities for other black women, creating an inclusive community center, and empowering African Americans. Walker saw a severe need for black women to enhance their self-esteem, and her hair products provided a solution. Their hair, after using her product, became much healthier, therefore boosting appearance-based self-esteem. In advertisements, Walker rejected the stereotypical depiction of African Americans as docile and cheerful servants, instead opting for more realistic imagery. She used realistic representations of African American women, which acknowledged the expanding importance of their buying power and dignified, rather than degraded them[5]. By placing herself on her packaging, she conveyed a clear message that there is something beautiful about herself, and something beautiful about you as well. Through her business, Walker also worked diligently to improve black women's economic place in society. A self-proclaimed progressive employer, she claimed: "I am endeavoring to provide employment for hundreds of women of my race."[6] The company used a franchise system where thousands of Walker Agents

[2] Rainey Horwitz, "Lydia Pinkham's Vegetable Compound (1873-1906)," Embryo Project Encyclopedia, last modified May 20, 2017, https://embryo.asu.edu/pages/lydia-pinkhams-vegetable-compound-1873-1906.

[3] Cynthia J. Davis, "Health and Medicine," in *American History through Literature 1820-1870*, ed. Janet Gabler-Hover and Robert Sattelmeyer (Detroit, MI: Charles Scribner's Sons, 2006),, Gale in Context: U.S. History.

[4] Library of Congress Research Guides, "The Panic of 1873," Library of Congress, https://guides.loc.gov/this-month-in-business-history/september/panic-of-1873.

[5] Cynthia J. Davis, "Health and Medicine," in *American History through Literature 1820-1870*, ed. Janet Gabler-Hover and Robert Sattelmeyer (Detroit, MI: Charles Scribner's Sons, 2006), Gale in Context: U.S. History.

[6] Mark David Higbee, "W. E. B. Du Bois, F. B. Ransom, the Madam Walker Company, and Black Business Leadership in the 1930s," *Indiana Magazine of History* 89, no. 2 (1993) JSTOR.

ran their own shops selling her products door-to-door, earning an exceptionally comfortable wage for living- more than most other African American women. She enabled over 5,000 black women throughout the entire country to become financially and personally independent through her resourceful business. Walker was also extremely impactful beyond hair care as she blurred the lines between entrepreneurship, politics, and social justice. Her salons served as crucial information-sharing and organization hubs for black people, and her first factory in Indianapolis created such a prosperous black business community that it could support three black newspapers. She donated $1,000 to fund the construction of a black YMCA, a contribution equivalent to roughly $25,000 today.[7] Walker was also unafraid to vocalize her political opinions, writing a telegram to President Woodrow Wilson protesting the race riots in East St. Louis. The next year, she sent another one arguing for better treatment of black soldiers in WWI. At the annual national conferences held for all Walker agents, she openly condemned lynching and racism. Her persistence and confidence in her efforts reveals her encompassing impact. While her original goal was to provide black women with effective products, it soon shifted to encompass the betterment of the entire black community through a higher quality of life. As an ambitious woman, Walker had many dreams she successfully accomplished. However, one of her largest goals- the construction of a cultural and social center, also the business's headquarters- was left unachieved. After her death in 1919, her company (now under control of her daughter, A'Lelia Bundles) constructed a massive, four-story and 48,000 square foot building. The Walker building provided an unsegregated restaurant, movie theater, and shops, all of which were an uncommon sight in a highly segregated downtown Indianapolis.[8] During its construction, most materials were purchased from black-owned businesses, workers were black, and the single building greatly boosted the local economy. It served as the heart of the black community where they could shop, grab food, and congregate in a safe environment, all under the Walker name, demonstrating how her legacy went far beyond hair care. Now named the Madam Walker Legacy Building, it remains an integral part of the community with a large portion of its original 1927 architecture. The building imparts cultural education, advances social justice, fosters entrepreneurship, and empowers youth to become civic leaders and entrepreneurs. Through African American art, such as concerts, showcases, and other events, the theater celebrates cultural diversity and rich heritage and traditions, continuing to honor Madam CJ Walker and the ideals she embodied.[9]

Lydia Pinkham expanded her impact beyond her medicine, illustrated by her honesty towards women's health, extension of education, and trustworthy advice. Her vegetable compound provided women with medicine for their menstrual problems during a time of common distrust towards doctors, who often failed to understand their issues. In conjunction with faulty treatments from professionals, Pinkham understood the need to educate others on female health. She did so by including written pamphlets on diet, health, and exercise with every purchase of the compound. Not only did women using the medicine receive the advice, but the

[7] A'lelia Bundles, "The Life and Times of Madam C. J. Walker," *History News* 58, no. 1 (2003):, JSTOR.

[8] Rita G. Koman and TwHP Staff, "Two American Entrepreneurs: Madam C. J. Walker and J. C. Penney," *OAH Magazine of History* 20, no. 1 (2006): JSTOR.

[9] "MWLC Today," Madam Walker Legacy Center, https://madamwalkerlegacycenter.com/about-us/.

men who often purchased it for them did too, helping to educate husbands and fathers. The clear labeling and language on her packaging informed her audience on the female body and reproductive processes. In addition, she wrote a book titled *Yours For Health* with the sole purpose of educating women on sexual processes. It was a rare source of unsentimental explanations of puberty, conception birth, menopause, and other female health issues that used the scientific terms for every part and function, all of which remain accurate today.[10] She also began an initiative in which women could write to Pinkham herself and receive advice in return. Promising confidentiality in these letters, this encouraged women to ask questions otherwise deemed uncomfortable or taboo. At her peak, she got about 150 letters every day and faithfully answered each one. The answers she gave were not complex- most of them contained common sense, yet they provided women with a straight-forward and trusted source of information- something sparse at the time.

Finally, Pinkham's words helped dispel the myth that women were weak and fragile, empowering them. She formed the idea that one could be healthy and female. As her compound spread amongst Lynn, Massachusetts, "women's weakness" declined. Although her compound contained a high percentage of alcohol- over 19%, as analyzed by the British Medical Association,[11] some ingredients were found to work similarly to estrogen therapy fifty years before it became common practice. The black cohosh specifically has been proven to have estrogenic effects in the appropriate dosage in Pinkham's Vegetable Compound.[12] While it is not as effective as modern medicine, her treatment undoubtedly helped women through their menstrual pains, empowering them in the process. Her medicine also sparked the women's wellness craze, the first product in a now $77.8 billion[13] industry. No price can be put on Pinkham's impact, however, because through her writing and knowledge sharing, she broke down the stigma surrounding female health and instilled confidence and trust into women in the nineteenth century.

This female entrepreneurial drive extends beyond Walker and PInkham, though. Muriel Siebert became the first female to hold a seat on the New York Stock Exchange in 1967 after establishing her own finance firm. With her high position of power, she served as an advocate for women and minorities in the industry and developed a female financial literacy program. Mary Ellen Pleasant, who some consider the first female self-made millionaire (instead of Walker) earned money through smart investments with the sole purpose of using her wealth to help as many people as possible. She achieved this goal through supporting the abolitionist movement and opening laundries and boarding houses, staffed mainly by black women. Her financial and political influence contributed to declaring streetcar segregation illegal and the repeal of a law

[10] "Lady with a Compound," *The American Journal of Nursing* 59, no. 6 (1959): 1, https://doi.org/10.2307/3417615.

[11] "Lady with," 2.

[12] Varro E. Tyler, "Was Lydia E. Pinkham's Vegetable Compound an Effective Remedy?," *Pharmacy in History* 37, no. 1 (1995): 4, JSTOR.

[13] Transparency Market Research, "Women's Health Market Size to Reach USD 130.9 Billion 2031, at a 5.5% CAGR - Exclusive Report by Transparency Market Research Inc.," PR Newswire, last modified November 30, 2023, https://www.prnewswire.com/news-releases/womens-health-market-size-to-reach-usd-130-9-billion-2031--at-a-5-5-cagr--exclusive-report-by-transparency-market-research-inc-302001890.html#:~:text=30%2C%202023%20%2FPRNewswire%2F%20%2D%2D,biological%20issues%20than%20do%20males.

banning Black testimony in California courts. In conjunction with Walker and Pinkham, Siebert and Pleasant utilized female entrepreneurship to better women's place in society.

Female entrepreneurship from leaders like Madam C.J. Walker and Lydia Pinkham has been a crucial factor in enhancing women's positions in society looking further than only the economic benefits. Walker made vast progress for black women by employing thousands of them as salespeople, expanding job opportunities. Her brand worked to create a desegregated community and blurred the lines between business, philanthropy, and politics. Lydia Pinkham and her vegetable compound spread knowledge about women's health and empowered women to take more control over their own wellbeing. Though the efficacy of her product is questionable, Pinkham helped eliminate the idea of women as weak and provided female-to-female advice. These two successful women, though, started their businesses out of necessity when there were no other jobs available, an unfortunate reality of the late 19th and early 20th century before the Suffrage Movement. Female entrepreneurs in America today continued to follow the path set by Walker and Pinkham, combining business with changemaking. The 19th Amendment, which granted women the right to vote, was ratified in 1920- one year after the death of Walker. Perhaps the most important feature of her legacy was her determination to extend the influence of women, and there is no greater vehicle for impact than voting. Though she (nor Pinkham) lived to see the ultimate power of voting women would gain, both undeniably paved the way for this right by bettering women's position in society through their entrepreneurship.

Works Cited

Alexander, Amy. *Fifty Black Women Who Changed America*. 1999.

I used this book to gain more in depth knowledge about Walker, as it dedicates an entire chapter to her. I found information here I hadn't discovered anywhere else in the easily digestible format of a book. This book provided me with the most in depth background knowledge on her and triggered additional research on specific aspects of her life, such as her headquarters in Indianapolis.

American Federation of Labor and Congress of Industrial Organizations. aflcio.org/about/history/labor-history-events/lowell-mill-women-form-union. Accessed 3 Mar. 2024.

This article provided me with necessary knowledge about the Lowell Mills Girls for my contextualization. I wanted to discuss women's labor history and their role in the workforce prior to the emergence of Walker or Pinkham, and the Lowell Mills is a very important part of that. These women were among the first leaders in the women's labor history, and this source gave me a powerful quote from an employee herself, explaining the significance of their strike.

Baskett, Sam S. "Eliza Lucas Pinckney: Portrait of an Eighteenth Century American." *The South Carolina Historical Magazine*, vol. 72, no. 4, 1971, pp. 207-19. *JSTOR*, www.jstor.org/stable/27567072.

This source includes a primary source, a letter written from Eliza Lucas Pinckney to a friend. Often regarded as one of the first female entrepreneurs, many doubted her, including her own husband. She writes, "Pray tell him I think these so, and what he may now think, whims and projects may turn out well by and by- out of so many one may hit," regarding her husband, meaning that he didn't expect any business ventures to succeed. She eventually did succeed, as she developed the indigo industry. This primary source reveals the gender climate at the time and the challenges female entrepreneurs faced.

Bundles, A'lelia. "The Life and Times of Madam C. J. Walker." *History News*, vol. 58, no. 1, 2003, pp. 6-9. *JSTOR*, www.jstor.org/stable/42655535.

Written by her great-great-granddaughter, this is the introduction to a book written about Madam C.J. Walker from a familial perspective. This article is not only thorough in its historical information about her, but also reveals the extent of her inspiration towards black women. Using a flattering and honorous tone, the author delves into many anecdotes about her great-great-grandmother, demonstrating her constant work towards betterment of black women. She also explains Walker's legacy by explaining how she remains inspired by her great-great-grandmother nearly a century later.

Bundles, A'Lelia Perry. "Walker, Madam C. J." *Encyclopedia of African-American Culture and History*, edited by Colin A. Palmer, 2nd ed., vol. 5, Detroit, MI, Macmillan Reference USA, 2006, pp. 2259-60. *Gale in Context: U.S. History*, link.gale.com/apps/doc/CX3444701272/UHIC?u=mlin_m_nnorth&sid=bookmark-UHIC&xid=7565e295.

This biography focused on one female entrepreneur, Madame C.J. Walker, and explains her influence beyond only her business. Although she was an extremely successful entrepreneur, becoming the first African American woman millionaire, the extent of her impact is greater. She advocated for women's economic independence by creating business opportunities when most were working as maids or farmers. This inspired me to do my research about not just the entrepreneurial aspect of these women, but the impact that had as well.

Davis, Cynthia J. "Health and Medicine." *American History through Literature 1820-1870*, edited by Janet Gabler-Hover and Robert Sattelmeyer, vol. 2, Detroit, MI, Charles Scribner's Sons, 2006, pp. 493-500. *Gale in Context: U.S. History*, link.gale.com/apps/doc/CX3450700113/UHIC?u=mlin_m_nnorth&sid=bookmark-UHIC&xid=789bc0e9.

This excerpt from a book was abundantly helpful when researching Lydia Pinkham and the landscape of women's health in the 19th century. It discusses health from 1820-1870 and how Victorian ideals of womanhood, a lack of trust in doctors, and the spread of "hysteria" shaped female wellness. Alternative medicine, it claims, was a rebellion against the new yet unsuccessful treatments, led by Pinkham's vegetable compound. This source shaped my research by providing background knowledge about women's health at the time, which helps understand Pinkham's impact against how health had been.

Hanson, Susan. "Changing Places through Women's Entrepreneurship." *Economic Geography*, vol. 85, no. 3, 2009, pp. 245-67. *JSTOR*, www.jstor.org/stable/40377305.

This journal provides really interesting insight into female entrepreneurship in America, describing the relation between place (physically and socially) and gender and how entrepreneurship is a channel for women to change their place. While this doesn't contain information about any of the specific women I studied, it serves as important background knowledge that I can weave in throughout my essay. From this article, I learned about the different motives for starting a business- opportunity or necessity- and this helped me shape my thesis and argument.

Higbee, Mark David. "W. E. B. Du Bois, F. B. Ransom, the Madam Walker Company, and Black Business Leadership in the 1930s." *Indiana Magazine of History*, vol. 89, no. 2, 1993, pp. 101-24. *JSTOR*, www.jstor.org/stable/27791654.

This journal continued to add to my knowledge of Walker by providing background information on her and explaining her method of advertising. She labeled herself as a "progressive employer," meaning she strived to provide employment for hundreds of women. It also details her impact outside of hair, with the construction of the Walker Building and her work with National Negro

Business League. This article gave me a more thorough explanation of the extent of Walker's impact for women and African Americans in Indianapolis.

Horwitz, Rainey. "Lydia Pinkham's Vegetable Compound (1873-1906)." *Embryo Project Encyclopedia*, Arizona State University, 20 May 2017, embryo.asu.edu/pages/lydia-pinkhams-vegetable-compound-1873-1906.

This encyclopedia gave me a detailed history of Lydia Pinkham's Vegetable Compound, helping me understand the necessity of the product. It explains how she used calculated marketing: clear and accurate labeling and information to advise women who wanted trustworthy service, something hard to find in doctors at the time. She also displayed herself as a caring, grandmotherly figure, assuring women that the medicine was created by a woman, for women. These advertisement methods helped shift power to women by taking away the medical authority away from the male-dominated physician field.

Koman, Rita G., and TwHP Staff. "Two American Entrepreneurs: Madam C. J. Walker and J. C. Penney." *OAH Magazine of History*, vol. 20, no. 1, 2006, pp. 26-35. *JSTOR*, www.jstor.org/stable/25162013.

Though this journal compares two entrepreneurs, Walker and Penny, I used its information on Walker to gain a lot of my context and background knowledge, as well as her impact beyond the business. It describes her business model by explaining how she knew she was dependent on black women as her main customers, leading her to desire to help them succeed like she did. The jobs she provided were extremely competitive wage-wise, and the article goes into more depth about her other influences, such as the construction of the headquarters.

"Lady with a Compound." *The American Journal of Nursing*, vol. 59, no. 6, 1959, pp. 854-55. *JSTOR*, https://doi.org/10.2307/3417615.

This journal article was one of the most important sources of my thesis. It outlines Pinkham's life and accomplishments in depth, and provided me with ample information on her impact beyond her medicine. I learned about her book, her willingness to use the proper anatomical terms, confidential responses to women's letters, and urgency to dispel the myth regarding women's weakness. This shaped my research by displaying the extent of her impact and providing a new lens on her- as not only a "Lady with a Compound," but a lady who helped further the entire gender.

Lewis, Jone Johnson. "Women and Work in Early America." *ThoughtCo*, 11 Sept. 2019, www.thoughtco.com/women-at-work-early-america-3530833.

The introduction of this thesis includes contextualization for the situation of women's labor. This article contributed to that by explaining women's work in early America. By including these historical details, a large contrast emerges between this and the high number of jobs Walker produced, as well as the boundaries Pinkham broke by talking about women's health. Women in the 1700 and 1800s rarely worked outside the house, and if they did it was in traditionally feminine spheres.

Library of Congress Research Guides. "The Panic of 1873." *Library of Congress*, guides.loc.gov/this-month-in-business-history/september/panic-of-1873.

I used this source to help provide background information to Lydia Pinkham's story and the necessity of her business. She started selling her compound when her husband lost his job due to the Panic of 1873, which this article outlines. This article is brief, but gave me enough to explain why her husband became unemployed and the severity of the exigence that led to the creation of Pinkham's Vegetable Compound- a product that would soon transform female lives.

Loscocco, Karyn A., and Joyce Robinson. "Barriers to Women's Small-Business Success in the United States." *Gender and Society*, vol. 5, no. 4, 1991, pp. 511-32. *JSTOR*, www.jstor.org/stable/190098.

This academic journal, written in 1991, outlines the barriers to why it is difficult for female entrepreneurs to be successful. It states that women have more key traits needed to be entrepreneurs, yet make up an obscure amount of all enterprises. Also, it explains that women-owned businesses tend to be in the personal service, beauty, or education sectors, which is an interesting comparison to the women I am researching. These factors transcend time, remaining relevant during the times of all of these women.

Madam C. J. Walker. "I Hope You Will Catch the Inspiration." *Speaking While Female Speech Bank*, speakingwhilefemale.co/business-walker/.

This speech provided me with the evidence I needed to prove Walker's influence on women's confidence. Walker spoke at the 13th National Negro Business League Annual Convention, where Booker T. Washington failed to acknowledge her or any other female entrepreneurs. However, she still spoke and gave a powerful talk on her journey as a woman "from the cotton fields of the south" who promoted herself to an extremely successful business owner. She outlines her final ambition not of becoming uber-rich but helping and inspiring others.

"Mary Ellen Pleasant." *National Park Service*, www.nps.gov/people/mary-ellen-pleasant.htm.

I knew I wanted to briefly focus on a few other female entrepreneurs (besides Walker and Pinkham), and I discovered Mary Ellen Pleasant. Some historians actually consider her the first female black millionaire, not Walker, but regardless of who was first, both are very important. As a housekeeper for prominent white families, Pleasant earned her money by listening to conversation from her employers and then making smart investments based on them. She strived to make as much money as possible to put it back into the black community by opening a library and meeting place, advocating against segregation, and employing many African Americans at laundries and boarding houses. Pleasant was extremely beneficial to the black community, just like Walker.

Mcgee, Suzanne. "7 Trailblazing American Women Entrepreneurs." *History.com*, Feb. 2024, www.history.com/news/successful-american-women-entrepreneurs-history.

Beginning this research, I only had a broad topic in mind without any specifics. I knew I wanted to focus on female entrepreneurs throughout history and their impact, but I didn't know where to start or who to start with. This article provided me with some examples of famous ones, and guided my research towards three women in particular: Madame C.J. Walker, Estée Lauder, Eliza Lucas Pinckney, and Muriel Siebert. From this point, I was able to focus on a few entrepreneurs rather than the entire topic, and pushed me into further, more specific learning.

"MWLC Today." *Madam Walker Legacy Center*, madamwalkerlegacycenter.com/about-us/.

This website is entirely dedicated to the Madam Walker Legacy Center, so I used their about page to learn about what the building is used for today. The center is registered as a National Historical Landmark, surviving almost a century, and has served as a vital African American community-building space for all its years. I learned that it has hosted various black performers and showcased art. It is an "enduring symbol" of Walker's legacy in improving the lives of black Americans.

Peiss, Kathy. "'Vital Industry' and Women's Ventures: Conceptualizing Gender in Twentieth Century Business History." *The Business History Review*, vol. 72, no. 2, 1998, pp. 218-41. *JSTOR*, https://doi.org/10.2307/3116276.

This source provided me with very interesting concepts that shaped my paper. It describes Madam C.J. Walker's company as something called "charismatic capitalism," which means the business combined with a social movement-economic nationalism, racial advancement, and female emancipation. Therefore, this journal helped me to understand how her impact spread beyond the business and into increased opportunities and self-respect, also through its explanation of some of Walker's philanthropic acts and the positive power of her products.

Schlesinger Library, Radcliffe Institute, Harvard University, Cambridge, Mass, compiler. *Pinkham Pamphlets*. Lydia E. Pinkham Medicine Company Records, 1873-1968. *Open Collections Program at Harvard University*, curiosity.lib.harvard.edu/women-working-1800-1930/catalog/45-990097708700203941.

This is a primary source displaying some of Pinkham's pamphlets educating on women's health. I initially hoped to find some of the letters she had written to other women, but I actually discovered that she wanted to keep them so confidential from men that there are few surviving today. However, these pamphlets are fascinating to read as they reveal Pinkham's willingness to use real anatomy terms, as well as dozens of testimonies from happy customers. These reviews show how her compound strengthened women, boosted their morales, and allowed countless women to dispel the myth of weakness.

"The Southern Urban Negro as Consumer." *American Decades Primary Sources*, edited by Cynthia Rose, vol. 3, Detroit, MI, Gale, 2004, pp. 135-40. *Gale in Context: U.S. History*, link.gale.com/apps/doc/CX3490200463/UHIC?u=mlin_m_nnorth&sid=bookmark-UHIC&xid=40b5aa0f.

Displaying different advertisements geared towards African Americans in the 1930s, this primary source explains how consumers were attracted at that time. It demonstrates how Madame CJ Walker used uplifting and empowering depictions of African American women for her hair products, rather than the common alienating depictions, remnant of slavery. These images help my research by expanding on Madame CJ Walker's influence besides her wealth, as she was a positive influence on the community and an inspiration to young African American women.

Transparency Market Research. "Women's Health Market Size to Reach USD 130.9 Billion 2031, at a 5.5% CAGR - Exclusive Report by Transparency Market Research Inc." *PR Newswire*, 30 Nov. 2023, www.prnewswire.com/news-releases/womens-health-market-size-to-reach-usd-130-9-billion-2031--at-a-5-5-cagr--exclusive-report-by-transparency-market-research-inc-302001890.html#:~:text=30%2C%202023%20%2FPRNewswire%2F%20%2D%2D,biological%20issues%20than%20do%20males.

Statistic on value of women's health industry

Tyler, Varro E. "Was Lydia E. Pinkham's Vegetable Compound an Effective Remedy?" *Pharmacy in History*, vol. 37, no. 1, 1995, pp. 24-28. *JSTOR*, www.jstor.org/stable/41111661.

This was a very interesting and educational journal that examined the efficacy of each ingredient in Pinkham's compound, debating if it was a real treatment or not. The author ultimately decides that while it definitely is not the best treatment, especially today, it actually had a few ingredients that have been clinically proven to work. For its time, it was an impressive compound. I used this source to explain Pinkham's medical impact, but also allude to how her real value lay elsewhere, as an advocate for women's wellness.

Predicting Young's, Bulk, and Shear Modulus with Machine Learning By Bryan Chen, Mira Krishnaiah, Maximus Ren

ABSTRACT

Effective utilization of additive manufacturing has the potential to revolutionize design. Desirable attributes associated with additive manufacturing, such as the ability to rapidly prototype parts, construct parts with intricate geometries, and design architected materials, make it attractive to various fields, such as aerospace, automotive, and electronics, among others. However, discovering materials suitable for additive manufacturing proves to be very difficult, especially in the discovery of alloys. Alloys with high elastic moduli can be difficult to machine or print in a way that does not break or deform in the process, and discovering new alloys with the desirable elastic modulus using traditional methods can be costly, time-consuming, and ineffective. To aid the furtherment of materials discovery, we develop a set of artificial neural networks (ANNs) trained to accurately predict Young's Modulus, bulk modulus, and shear modulus of alloys found in the Materials Project database with significant accuracy. This method provides a far cheaper and expeditious computational approach to identify and discover alloys that are useful for additive manufacturing. Two models were created: one for predicting the shear modulus and one for predicting the bulk modulus. The bulk modulus model had a final testing RMSE of 0.3565 ± 0.0907, and the final shear modulus model had a testing RMSE of 0.9212 ± 0.5557.

Keywords: Additive manufacturing, machine learning, design, solid mechanics, experimental mechanics

I. INTRODUCTION

Additive manufacturing can be defined as creating objects by adding material[30] - a more familiar subset of which is 3D printing. Today, AM has a variety of applications including aerospace[31], automotive[32], machinery[33], medical implants[34], and custom parts[35], among many others[1]. Over the years, advancements in AM have significantly reduced development costs both in terms of finances and energy consumption[2]. These advancements are particularly true in the field of alloy design[3]. Alloys, as opposed to pure metals, offer the unique ability to possess tailored properties for a product's specific requirements[9]. Because of these properties, alloys are extensively used for a variety of mechanical applications, spanning disciplines such as bioengineering (shape memory alloys for bone screws, surgical tools, and equipment, etc[4]) and aerospace engineering (aluminum and titanium alloys for jet engines, airframes, etc[5]). They have even been applied to produce household appliances such as kitchen equipment[6], coins[7], musical instruments[8], etc. Consequently, a continuous effort is being dedicated to designing new alloys with properties that allow them to be used in AM thus simplifying, and accelerating the prototyping phase of the design process[3].

Amidst this effort, it is evident that not all alloys are suitable for printing. During AM of metal parts, alloys undergo heating, melting, solidification, and cooling of the entire part, and these conditions can modify or even destroy an alloy, deeming it unsuitable for production[11]. For this reason, searching for alloys with desirable properties–strength, temperature resistance, and elasticity–is both costly and inefficient.

To manufacture an alloy, an alloy must have appropriate elastic moduli, as the elastic moduli of an alloy define how alloys react to different forces, and therefore define if an alloy is practical and suitable for production. However, in materials discovery, the utilization of molecular dynamics to simulate the elastic moduli of a single alloy can be a time-consuming task, requiring days or even weeks of computation on a supercomputer[12]. Traditional methods to compute the elastic moduli of alloys (tensile testing, static torsion testing, utilization of a torsional pendulum, transient bulk creep experiments) suffer from inaccuracy, noise, and inefficiency as well. These methods, and others, often entail research timelines exceeding 10 years, with advanced materials taking upwards of 20 years to develop[3]. Thus, finding ways to optimize the discovery process of new alloys is critical.

In recent years, machine learning (ML) has emerged as a powerful tool to expedite the alloy discovery process by having the capability to target specific attributes of an alloy and predict the values associated with these attributes, classify phase structures, microstructures, and materials, and predict how an alloy will fare under the conditions they were designed for[10]. Open access to applications for creating neural networks (e.g. Tensorflow, Pytorch), and public datasets (e.g. Materials Project[28], CatApp[13]) have opened up new possibilities for ML utilization in research. ML can process large amounts of data and generate predictions based on identified trends, significantly accelerating the discovery process and uncovering hidden patterns that might otherwise go unnoticed. The remarkable benefits offered by ML are particularly advantageous in materials discovery, due to its potency in regards to making accurate predictions. With the power of machine learning, certain attributes of different materials and alloys can be predicted with remarkable amounts of accuracy. Therefore, it has been employed in conjunction with traditional methods (e.g. density functional theory[14], molecular dynamics[15], and multivariate statistics[16]), as well as in a standalone prediction approach[17,18,19]. In the field of alloy discovery, ML has demonstrated success in designing and predicting the elastic properties of various high-performing alloys such as high-entropy alloys[20,21,22], low- and medium-entropy alloys[23], multicomponent alloys[24], and compositionally complex alloys[25]. In this article, we provide the process and challenges it took to create more generalized ML models, able to predict the elastic moduli for any given alloy with promising amounts of accuracy, and therefore further optimize the process for discovering new alloys.

II. METHODS / METHODOLOGY

In this section, we outline how we extracted data from the Materials Project, identified relevant attributes and relevant materials within the dataset, and standardized the data points we obtained. The construction of any machine learning model begins with obtaining an unbiased, detailed, and reliable dataset.

A. Data Curation

Data was sourced from the Materials Project, a publicly accessible database of information regarding the attributes of various materials. Notably, data concerning Young's Modulus for each material was not directly provided; however, this value could be calculated using the bulk and shear moduli which were provided in the dataset[26]. The relevant equation for calculating Young's Modulus using the shear and bulk moduli is found below in Equation 1[27],where E denotes Young's Modulus, κ is the bulk modulus, and μ is the shear modulus.

$$E = \frac{9\kappa\mu}{(3\kappa+\mu)} \qquad (1)$$

We imported data from the Materials Project with the following restrictions and parameters on each material: Is a metal/alloy, Bulk/Shear moduli, and Voigt/Reuss. This set of parameters allowed us to filter through our data, and ensure that only alloys with data on the relevant elastic properties were extracted from the Materials Project database. For each material found where there was data on their elastic properties, the following parameters were extracted and used as inputs for our model, shown in Table 1.

Parameters	Units	Description
Density	g·cm^{-3} (grams per cubic centimeter)	Compactness of material
Atomic Density	atoms·cm^{-3} (atoms per cubic centimeter)	Number of atoms per cm^3
Volume	Å^3 (cubic angstrom)	The amount of space an object occupies
Energy Per Atom	N/A	Energy level of the electrons in atoms
Formation Energy Per Atom	eV/atom (energy per atom)	Energy required to generate that molecule configuration
Energy Above Hull	eV/atom (energy per atom)	Energy released by decomposing the compound to the most stable combination of compounds
Total Magnetization	μB/f.u (Bohr magneton per	Magnetic strength and

	formula unit)	orientation
Normalized Total Magnetization Volume	µB/f.u (Bohr magneton per formula unit)	Magnetic strength and orientation per unit volume.
Elements	N/A	element composition
Symmetry	N/A	type of symmetry

Table 1: Parameters extracted from the Materials Project, given with units and a brief description. This process left us with data on 72853 alloys to use to train and test our model.

Due to the significant variance in the magnitude of different parameters, one additional consideration made was regarding the standardization of all training data. This is necessary as it is well known that machine learning estimators tend to perform poorly if the individual features in the dataset are not normally distributed with zero mean. Accordingly, we scaled our data down to values between 0 and 1, using the Scaler tool from the Sci-kit Learn programming software[29]. This provided an effective and easily deployable way to standardize our data, allowing our ML model to run at greater speeds.

Standardization of a dataset is a common requirement for many machine learning estimators; ML algorithms might perform poorly if the individual features look different from a normally distributed dataset. Given that all the parameters had largely different unit sizes, finding a method to standardize our data and allow our ML model to interpret/process our data was crucial. Standardizing the output data to between 0 and 1 worked well.

B. Integrating Machine Learning

Since Young's modulus was not directly given in the Materials Project dataset, as pre-existing prediction methods for Young's modulus are known to be highly inaccurate[36], we decided to take on an approach in which we predicted bulk and shear modulus and later constructed a model to predict Young’s modulus.

We constructed multiple Artificial Neural Networks (ANN). One key advantage ANNs have compared to other models is their learning ability for complex and nonlinear relationships[15] and parallel processing. As more layers are added to the network, or more nodes are added within each layer, the resulting model becomes more adept at capturing the complex nonlinear relations within the data. These unique advantages therefore make an ANN architecture our algorithm of choice. The trained model was evaluated using K-fold cross-validation. This was due to its resilience against overfitting and split sensitivity. To speed up computational time, 5 was chosen as the number of folds. We chose Root Mean Squared Error (RMSE) as our loss function (in gigapascals).

C. Architecture and Hyperparameter Tuning

To set a benchmark for the performance of our machine learning model, we built a linear regression estimator, shown in Figure 1. With only material density and element composition as the inputs for training, the linear regression model had a testing RMSE of 34.677 ± 0.646 gigapascals for the bulk model and 101.255 ± 7.637 gigapascals for the shear model.

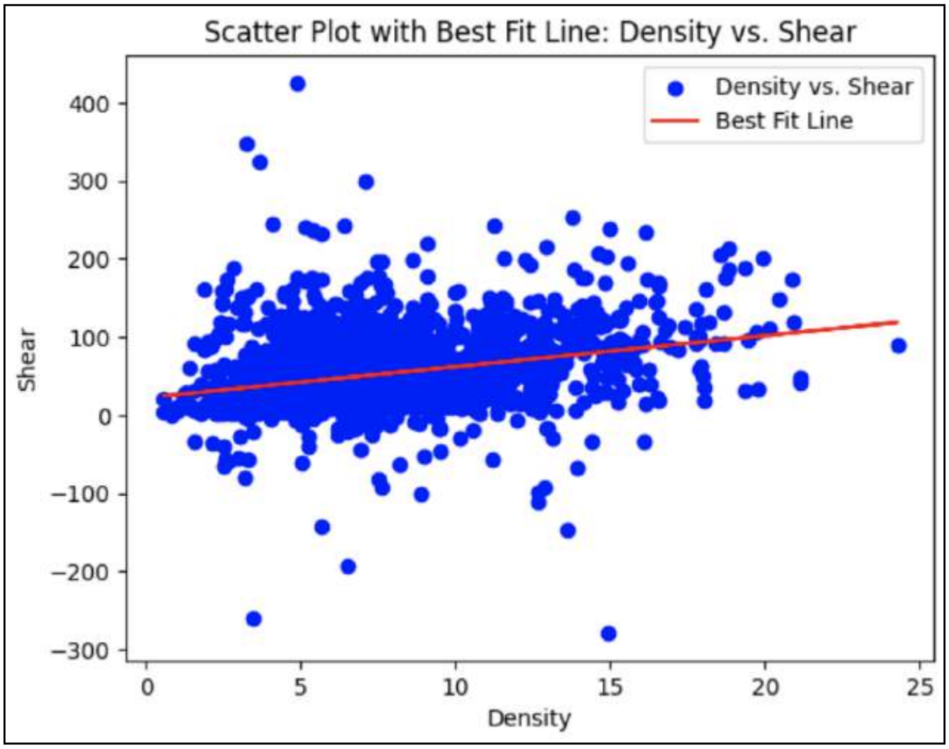


Fig 1: A scatter plot displaying the linear regression model's best-fit line with respect to the shear modulus values over density.

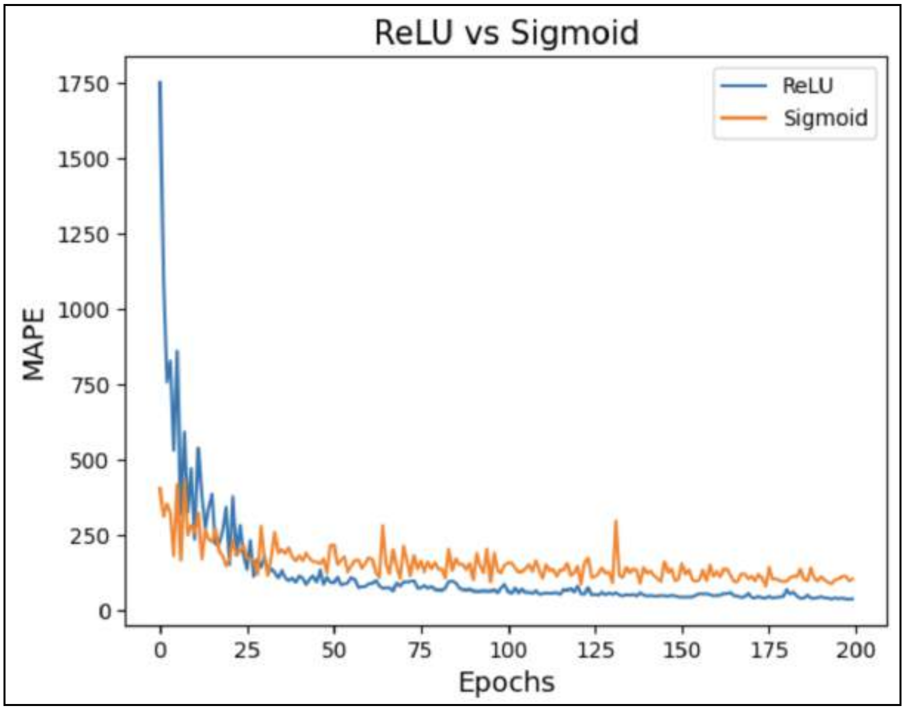


Fig 2: RMSE as a function of epochs, for both the ReLU activation function and the Sigmoid activation function.

We then compared the performance of the model with two activation functions -- ReLU and Sigmoid, as shown in Figure 2. We found that regardless of the number of epochs, ReLU had significantly lower RMSE. Accordingly, ReLU was chosen as our loss function of choice in the final model. Finally, to tune the architecture and hyperparameters, a grid search was performed on both models using the following parameters: Learning rate, layers, layer width, batch size, dropout, and epochs. In total, 128 models were tested using grid search: 64 for bulk and 64 for shear. After finding the best set of parameters and standardizing our output data instead of our input data, we can see the RMSE for both the bulk and shear models, shown in Table 2.

	Learning Rate	**Layers**	**Layer Width**	**Batch Size**	**Dropout**	**Epochs**	**Training RMSE**	**Testing RMSE**
Bulk	0.001	3	128	32	0.0	1000	0.1196 ± 0.0046	0.4130 ± 0.0326
Shear	0.0001	3	64	32	0.0	1000	0.5006 ± 0.1665	1.1601 ± 1.001

Table 2: Best parameters for bulk and shear moduli models after a grid search

It can also be noted that they both converge around 200 epochs, shown in Figures 3A and 3B, for the bulk and shear models, respectively.

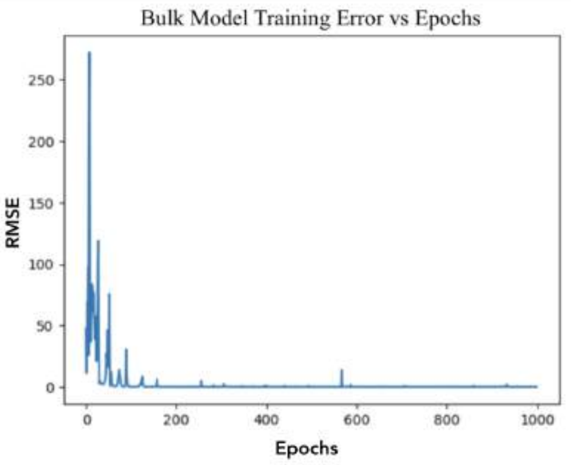


FIG. 3A. Bulk Model's training error vs. epochs upon standardizing output data

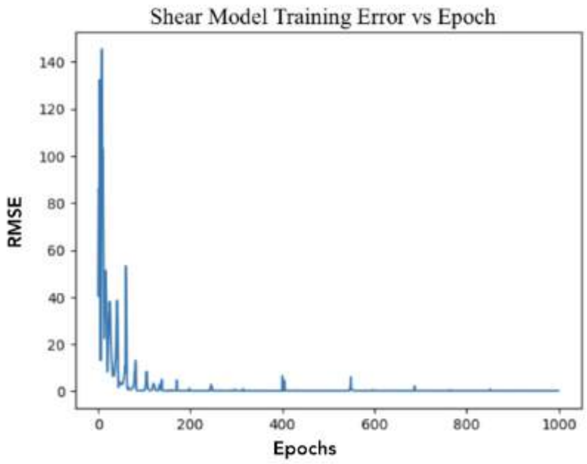


FIG. 3B. Shear Model's training error vs. epochs upon standardizing output data.

III. RESULTS AND DISCUSSION

A. Model Finalization

Our final hyperparameters, for both bulk and shear modulus models, are a learning rate of 0.001 and 1000 epochs. Our final architecture, for both bulk and shear modulus models, are:

Activation Function	Layers	Nodes per layer	Dropout	Batch size
Relu	3	128	0.0	32

Table 3: Final model parameters

The bulk model, with this set of parameters, gave a training RMSE of 0.2414 ± 0.0811, and a testing RMSE of 0.3565 ± 0.0907, which means that our bulk modulus prediction is on average ~0.36 gigapascals off. Similarly, the shear model had a training RMSE of 0.4882 ± 0.4971, and a testing RMSE of 0.9212 ± 0.5557 - so our prediction of shear modulus is on average ~0.92 gigapascals off. Through these low percent errors, we can conclude that this model can accurately predict both the bulk and shear, and therefore Young's modulus of an alloy using just density and element composition, which can exponentially expedite the alloy discovery process.

B. Model Comparison

As previously mentioned, the linear regression model had a testing RMSE of 0.42 ± 0.09 for the bulk model and 2.01 ± 2.90 for the shear model. In comparison, the linear regression model had a 0.07 higher percent error for the bulk model, and 1.09 higher error for the shear model, showing that machine learning was a more accurate solution to predicting the bulk and shear modulus.

It's important to note that the standard deviation is greater than the average for the linear regression model because there was a large amount of split sensitivity. However, we can see that our shear model did not have this issue, indicating a great improvement in accuracy and more resistance to split sensitivity.

Finally, We can see that two models for bulk and shear work well but there's an error associated with each of them and we are unsure of how that error will add up when we calculate Young's modulus, so, using the same input parameters, we made one model with Young's modulus. This model had a training RMSE of 0.6605 ± 0.3166 and a testing RMSE of 0.9233 ± 0.6353. Although it had a lower error than the shear model, it had a significantly higher error than the bulk model. These results proved to be both promising and accurate, and we believe that further optimization of the input materials and fields will even further decrease the percent error for Young's modulus model.

C. Discussion

We believe that standardizing our output data made the target easier to interpret by the model or made it closer to a Gaussian curve, which allowed the model to learn patterns easier – the lack of standardizing the input data could have also contributed to this. Additionally, the output data could have a wider range or a more complex distribution (there was more split sensitivity). By standardizing the output data, we were able to remove or lessen the impact of the outliers and make the output range a more manageable range for the model. In order to know for certain, extensive experimentation and analysis with model inputs must be done, and given the time constraints, this was unfeasible.

IV. CONCLUSION

This paper presents a machine-learning approach to predict the elastic moduli of alloys. The ANNs we trained were able to predict the bulk, shear, and Young's moduli of alloys within the Materials Project dataset with an accuracy upwards of 98%. This high amount of accuracy indicates that our model can be employed as a reliable way to predict such elastic moduli, to identify alloys that are desirable for certain purposes, and to identify alloys able to withstand the additive manufacturing process. Furthermore, our work shows that machine learning has the potential to dependably predict important attributes of any material, which would allow for efficient identification and characterization of serviceable materials, greatly accelerating materials discovery in the future.

WORKS CITED

1. Hu, Y. (2021). Recent progress in field-assisted additive manufacturing: materials, methodologies, and applications. Materials Horizons, 8(3), 885-911. https://pubs.rsc.org/en/content/articlehtml/2021/mh/d0mh01322f
2. Gupta, Nayanee, Christopher Weber, and Sherrica Newsome, (2012). Additive manufacturing: status and opportunities. https://www.researchgate.net/profile/Justin-Scott-4/publication/312153354_Additive_Manufacturing_Status_and_Opportunities/links/59e786db458515c3630f917b/Additive-Manufacturing-Status-and-Opportunities.pdf
3. Bandyopadhyay, Amit, et al, (2022). Alloy design via additive manufacturing: advantages, challenges, applications and perspectives. https://www.sciencedirect.com/science/article/pii/S1369702121004314
4. Gil, F. J., & Planell, J. A. (1998). Shape memory alloys for medical applications. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 212(6), 473-488. https://journals.sagepub.com/doi/abs/10.1243/0954411981534231
5. Peters, M., Kumpfert, J., Ward, C. H., & Leyens, C. (2003). Titanium alloys for aerospace applications. Advanced engineering materials, 5(6), 419-427. https://onlinelibrary.wiley.com/doi/abs/10.1002/adem.200310095
6. Davis, J. R. (1993). Aluminum and aluminum alloys. ASM international. https://materialsdata.nist.gov/bitstream/handle/11115/173/Aluminum%20and%20Aluminum%20Alloys%20Davis.pdf
7. Beck, Lucile, et al. (2004). Silver surface enrichment of silver–copper alloys: a limitation for the analysis of ancient silver coins by surface techniques. 226, 1-2, 153-162. https://www.sciencedirect.com/science/article/pii/S0168583X04008316?ref=pdf_download&fr=RR-7&rr=7e6bcdb12cfd52c5
8. Stanciu, Mariana Domnica, et al. (2022). Mechanical and Acoustic Properties of Alloys Used for Musical Instruments. 15.15 5192 https://www.mdpi.com/1996-1944/15/15/5192
9. Tammas-Williams, S., and I. Todd, (2017). Design for additive manufacturing with site-specific properties in metals and alloys. 135: 105-110. https://www.sciencedirect.com/science/article/pii/S1359646216305255
10. Kaufmann, Kevin, and Kenneth S. Vecchio, (2020). Searching for high entropy alloys: A machine learning approach. 198: 178-222. https://www.sciencedirect.com/science/article/pii/S1359645420305814?casa_token=4HrKzBedW9oAAAAA:or_5vhdx8eW95icOGrTmsKo3o8txHhPzPo6c7-ejIdy767d6D7pBJZB1NMqXdkHVxAm6IGFxhJw
11. Mukherjee, Tuhin, et al. (2016). Printability of alloys for additive manufacturing. 6.1: 19717. https://www.nature.com/articles/srep19717

12. Li, Guohe, Meng Liu, and Shanshan Zhao, (2021). Reduced computational time in 3D finite element simulation of high speed milling of 6061-T6 aluminum alloy. 25.4: 558-584. https://www.tandfonline.com/doi/full/10.1080/10910344.2020.1855651
13. Goldsmith, Bryan R., et al. (2018). Machine learning for heterogeneous catalyst design and discovery. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/144583/aic16198.pdf
14. Wang, Juan, et al (2017). New methods for prediction of elastic constants based on density functional theory combined with machine learning. 138: 135-148. https://www.sciencedirect.com/science/article/pii/S0927025617303191
15. Yang, Kai, et al. (2019). Predicting the Young's modulus of silicate glasses using high-throughput molecular dynamics simulations and machine learning. 9.1: 8739 https://www.nature.com/articles/s41598-019-45344-3
16. Khan, Naseer Muhammad, et al. (2022). Application of machine learning and multivariate statistics to predict uniaxial compressive strength and static Young's modulus using physical properties under different thermal conditions. 14.16: 9901. https://www.mdpi.com/2071-1050/14/16/9901
17. Liu, Yue, et al (2017). Materials discovery and design using machine learning. 3.3: 159-177. https://www.sciencedirect.com/science/article/pii/S2352847817300515?
18. Singh, R., Kainthola, A., & Singh, T. N. (2012). Estimation of elastic constant of rocks using an ANFIS approach. *Applied Soft Computing*, 12(1), 40-45. https://www.sciencedirect.com/science/article/pii/S1568494611003899?
19. Mahmoud, Ahmed Abdulhamid, Salaheldin Elkatatny, and Dhafer Al Shehri. "Application of machine learning in evaluation of the static young's modulus for sandstone formations." Sustainability 12.5 (2020): 1880. https://www.mdpi.com/2071-1050/12/5/1880
20. Liu, X., Xu, P., Zhao, J., Lu, W., Li, M., & Wang, G. (2022). Material machine learning for alloys: Applications, challenges and perspectives. Journal of Alloys and Compounds, 921, 165984.. https://www.sciencedirect.com/science/article/pii/S0925838822023751
21. Rao, Z., Tung, P. Y., Xie, R., Wei, Y., Zhang, H., Ferrari, A., ... & Raabe, D. (2022). Machine learning–enabled high-entropy alloy discovery. Science, 378(6615), 78-85. https://www.science.org/doi/full/10.1126/science.abo4940?casa_token=m0_QvUWeE7AAAAAA%3Ai3y3KeUR-NblU3glu8Huqi3u2QFLUFryVH8c9NLYeeS8lKadYylRS1ZqC8zAsSi5sx1zzZE6IoHEEqQ
22. Kaufmann, K., & Vecchio, K. S. (2020). Searching for high entropy alloys: A machine learning approach. Acta Materialia, 198, 178-222. https://www.sciencedirect.com/science/article/pii/S1359645420305814?via%3Dihub
23. Roy, A., Babuska, T., Krick, B., & Balasubramanian, G. (2020). Machine learned feature identification for predicting phase and Young's modulus of low-, medium-and

high-entropy alloys. Scripta Materialia, 185, 152-158.
https://www.sciencedirect.com/science/article/pii/S1359646220302347

24. Liu, X., Peng, Q., Pan, S., Du, J., Yang, S., Han, J. et al. (2022). Machine learning assisted prediction of microstructures and Young's modulus of biomedical multi-component β-Ti alloys. *Metals*, *12*(5), 796. https://www.mdpi.com/2075-4701/12/5/796
25. Khakurel, H., Taufique, M.F.N., Roy, A. *et al.* Machine learning assisted prediction of the Young's modulus of compositionally complex alloys. *Sci Rep* 11, 17149 (2021). https://doi.org/10.1038/s41598-021-96507-0
26. De Jong, Maarten, et al. "Charting the complete elastic properties of inorganic crystalline compounds." *Scientific data* 2.1 (2015): 1-13.

    https://www.nature.com/articles/sdata20159
27. Makishima, Akio, and John D. Mackenzie. "Calculation of bulk modulus, shear modulus and Poisson's ratio of glass." *Journal of Non-crystalline solids* 17.2 (1975): 147-157.

    https://www.sciencedirect.com/science/article/pii/0022309375900472
28. Jain, Anubhav, et al. "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation." *APL materials* 1.1 (2013).

    https://pubs.aip.org/aip/apm/article/1/1/011002/119685
29. Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.
30. Standard, A. S. T. M. (2012). Standard terminology for additive manufacturing technologies. *ASTM International F2792-12a*, 1-9.
31. Blakey-Milner, B., Gradl, P., Snedden, G., Brooks, M., Pitot, J., Lopez, E., ... & Du Plessis, A. (2021). Metal additive manufacturing in aerospace: A review. *Materials & Design*, *209*, 110008.

    https://www.sciencedirect.com/science/article/pii/S0264127521005633
32. Leal, R., Barreiros, F. M., Alves, L., Romeiro, F., Vasco, J. C., Santos, M., & Marto, C. (2017). Additive manufacturing tooling for the automotive industry. *The International Journal of Advanced Manufacturing Technology*, *92*, 1671-1676.

    https://link.springer.com/article/10.1007/s00170-017-0239-8
33. Brajlih, T., Valentan, B., Balic, J., & Drstvensek, I. (2011). Speed and accuracy evaluation of additive manufacturing machines. *Rapid prototyping journal*, *17*(1), 64-75.

    https://www.emerald.com/insight/content/doi/10.1108/13552541111098644/full/html
34. Petrovic, V., Haro, J. V., Blasco, J. R., & Portolés, L. (2012). Additive manufacturing solutions for improved medical implants. *Biomedicine*, *2012*, 147-180.

35. Huang, S. H., Liu, P., Mokasdar, A., & Hou, L. (2013). Additive manufacturing and its societal impact: a literature review. *The International journal of advanced manufacturing technology*, *67*, 1191-1203. https://link.springer.com/article/10.1007/s00170-012-4558-5
36. Levämäki, Henrik, et al. "Predicting elastic properties of hard-coating alloys using ab-initio and machine learning methods." NPJ Computational Materials 8.1 (2022): 17. http://www.diva-portal.org/smash/get/diva2:1637546/FULLTEXT01.pdf

ACKNOWLEDGEMENTS

We thank Alexander (Sasha) Davydov, Caelin Muir, and Abed Musaffar for providing us invaluable information and guidance during this research project.

We also would like to thank Dr.Lina Kim and Dr.Teresa Holden for giving us this research opportunity at SRA, and UCSB for access to their facilities.

AUTHOR CONTRIBUTION STATEMENT

All authors performed a literature review. M.K. and M.R. obtained data and constructed the machine-learning model. All authors analyzed the results and reviewed the manuscript.

To what extent did the psychological inclination toward self-preservation play a role in the conformity or defiance of the Sonderkommando prisoners of Nazi Germany?

By Sarah O’Grady

Introduction

The extent to which the Third Reich perpetually altered the minds of those subjugated by the regime is debated by historians. While Adolf Hitler oscillated between structuralist and intentionalist policies throughout his time as Führer, the multi-faceted lens through which the consequences of his actions can be perceived is crucial to maintaining a well-rounded understanding of the corruption and human degradation evidenced by the Holocaust. The various concentration camps created and altered to carry out the “Final Solution”, or the systemic eradication of the European Jewish population, acted as a realm previously foreign to humankind. Therefore, the partition of the camps, particularly Auschwitz-Birkenau, into various sectors designated for “special” jobs was also a concept beyond what human beings had fathomed. The fabrication of the Sonderkommando unit of prisoners, or those who worked in the crematoria, directly coincided with the innate SS desire to subjugate Jewish prisoners to the same moral field they had regressed to and to force the victims themselves to perform as the *Geheimnistrager*, or “bearer of the secret”. The unabated isolation and corruption of those chosen to perform tasks associated with the crematorium act as a testament to the ulterior Nazi motive to disfigure the victim as a product of the Reich and to convince themselves that the Sonderkommando prisoners were too, perpetrators of systemic crime. The morally ambiguous depiction of the Sonderkommando generally arose from relatively uninformed groups of individuals, who neglected to attempt to grasp the complexity of the domain these particular prisoners are associated with. However, many overt critiques of the conformity of the Sonderkommando stem from Jewish survivors of the Holocaust, who condemn the Sonderkommando for willingly harming those with intersecting identities. The extent to which the inflected burden of guilt translated through to the Sonderkommando prisoners who survived Auschwitz is notable, yet nuanced in the sense that moral reprieve derived largely from individual experience and preconceived notions concerning humanity as a whole. The innate human partiality toward self-preservation permeates many of the written Sonderkommando testimonies and also serves as a layered phenomenon that explores the behavior of such individuals when placed into unimaginable yet realized circumstances.

Source review

Various interpretations of the overt motives of the Sonderkommando revolve around the notion that the prisoners sacrificed their moral integrity when submitting to the role. Timothy E. Pytell falls into this scope of thought in *Shame and Beyond Shame* as he examines various first-hand accounts of "privileged Jews" in an attempt to consolidate his perspective. Pytell focuses almost exclusively on the written testimonies of Primo Levi and Tadeusz Borowski, whose writings are contrastable but similar in the way that they emanate a sense of shame surrounding the moral positioning of the Sonderkommando. Pytell categorizes the writings of these authors as desperate attempts to reconnect to humanity and make sense of the world after it had been rendered void of meaning during their camp experience. The first-hand account of Borowski provides a crude yet useful perspective on life in Auschwitz as an Aryan who worked for one day as a Sonderkommando. As he was not psychologically affected in the same way that those who had been working the crematorium for months were, Pytell indicates that his cynicism and perceived collusion with the Nazi regime revealed valuable insight into the reasoning behind the Sonderkommando prisoners' ability to perform their tasks. Though the analysis of an unfiltered account of the Sonderkommando experience is crucial in understanding their frequent conformity, Pytell's reasoning is fragmented along the lines of obtaining a multi-faceted understanding of a delicate situation. Moreover, Pytell's assertation in his article that the human capacity to inflict pain on others was frequented in Auschwitz and that there was no such conception of "heroic survival" is presumptuous considering his perspective as a secondary evaluator. Pytell's emphasis on the position of privilege of those who survived the camp, and their conscious contribution to the deterioration of those who died coincides with other detatched opinions, such as Kate Lawless' "Memory, Trauma, and the Matter of Historical Violence: The Controversial Case of Four Photographs from Auschwitz ". Lawless falls into alignment with Pytell's overarching attestation to the compromised integrity that defined the Sonderkommando, although her argument is much more implicitly stated. The premise of Lawless' article explores the avenues of which the four existing Sonderkommando photographs have been taken societally, and how they can exist as both a historical document and an aesthetic object. She focuses on Didi-Huberman's book, *Images in Spite of All*, which allegedly fetishizes the pictures and overshadows the verbal testimonies of survivors. Further, Didi-Huberman categorizes the images as "'capable of disrupting, and reconfiguring, the relation habitually maintained by the historian of images" (Lawless 398). His insinuated abhorrence to the photos prompts readers to consider a subconscious disdain for the Sonderkommando. Didi-Huberman's stagnant discomfort, not taking steps to delve into the context of what makes these photos so shocking, provides insight into his perspective. However, apart from Lawless' critique of Didi-Huberman, she does not explicitly consider the lengths to which the individuals who took these photographs went to take them. Lawless consistently places emphasis on the notion that politics are derived from images, but acknowledgement of the stark opposite would make her argument all the more valuable, as both Pytell and Lawless maintain a relatively narrow scope of what is truly a layered matter.

Alternate perspectives regarding the psychological inclinations of the Sonderkommando center around the position of a "privileged Jew", and how it creates a multi-faceted detachment from one's ethical beliefs. In La "Zona Grigia", Adam Brown analyzes Primo Levi's "Gray Zone", and hones in on how Levi was unable to completely abandon his humanist ideology after he survived Auschwitz. Brown's paper provides insight into Levi's incessant introspection, which settles on the notion that humans are complex beings at their core, and are reduced to merely a "skeleton" when subjected to circumstances reminiscent of what he endured. The complexity

Brown continually highlights as he describes Levi's mental turmoil post-Auschwitz is a theme prevalent in Gideon Greif's "We Wept Without Tears: Testimonies of the Jewish Sonderkommando from Auschwitz". However, Greif and Brown differ in their views of what lies beyond the "discarded" humanity that frequented the external representation of the Sonderkommando. Configuring a cohesive and faceted representation of the Sonderkommando experience in Nazi Germany is a nearly impossible task, due to how individual experience determined perspective. Greif hones in on how to define the intentions behind the Sonderkommando prisoners is something that must be treated with utmost delicacy, and with consideration of the implicit prejudices one inevitably holds. She further emphasizes the extent to which deceit was thoroughly and purposefully employed throughout the process of selecting and consequently using Jewish prisoners in this "special unit". From the very moment SS guards selected various prisoners to partake in the Sonderkommando unity, they were not allowed time to consciously comprehend what they were experiencing before they had already immersed themselves in action. The violent psychological process employed by the Aryan guards, and the *Kommandoführer* in particular, was an intentionally disorienting strategy that reversely prioritized action over thought. Additionally, Greif's novel and Leah Christine Ingle's "Witness and Complicity: The Scrolls of Auschwitz and the Sonderkommando" heavily overlap as they explore the deliberate behavior of the SS in regard to the Sonderkommando. The Nazi guards wanted to rob the victims of their innocence and took systemic measures to isolate them from the other prisoners of the camp, as well as figuratively distance the Sonderkommando from each other. Both Greif and Ingle's consideration of this factor makes their respective arguments useful, as they consider the nuanced lens through which one can observe the Sonderkommando prisoners. Furthermore, Greif's adamant testament to the humanity that often lay "beneath" the characterization of a hardened Sonderkommando prisoner can be utilized to refute the uninformed belief that the prisoners were selfish or subhuman. Although the interminable physical labor enforced upon the Sonderkommando was intentionally industrialized to permeate the sense of detachment from the outside world, underneath this "layer", Greif supports the notion that the prisoners harbored normal human tendencies and did have a desire to live, though not in a way that intentionally degraded others. Though subtle at times, humanity was prevalent throughout the actions of the Sonderkommando and their sparse interactions with the living. The empathy noted in the secret Sonderkommando writings, or the Scrolls of Auschwitz, reveal the stark opposite of the callous façade imposed on the prisoners. Greif and Ingle take into consideration primary accounts of these writings and translate them into their overarching theses -- which ultimately conclude that the intentions of the Sonderkommando will continue to be best defined by those who experienced it themselves.

Concluding Thoughts

Although the psychological inclination toward self-preservation is certainly notable when considering the patterns of conformity that permeated the behavior of the Sonderkommando prisoners, the anomaly of their position provides crucial insight into the ways in which humanity too, was preserved. The blurred lines between the tangibility of life and death that confronted the Sonderkommando prisoners through their incessant work provide valuable insight into how the meaning of life itself became distorted for many. Despite the strings of moral decency that may have followed one individual or another into the camps, a variety of primary accounts testify to the developed inability to grasp onto the remains of one's principled conscience. Zalman Lewental described how, as a Sonderkommando prisoner, "our intelligence is subconsciously influenced by the wonderful will to live, by the impulse to remain alive; you try to convince yourself, as if you do not care about your own life, but want only the general good, to go through with all of this for this and that cause, for this and that reason; you find hundreds of excuses, but the truth is that you want to live at any price" (Greif 19). The visceral obligation toward life exists in tandem with the human hesitance to disrupt one's moral beliefs. As evidenced by Lewental, the position of the Sonderkommando is unique in and of itself, and the human partiality toward self-preservation should not be condemned in a delicate situation such as this. As Greif considered in her novel, one ulterior motive of the Sonderkommando to comply with the demands of the SS was the prospect of surviving to give witness to the events that occurred. The position thrust upon the Sonderkommando prisoners defined them as both victims and witnesses and though the potential of survival was slim, many felt obligated to utilize their position of "privilege". Despite the existence of primary sources, including the SCI Grapevine newspaper issue of 1972, that dispute the integrity of the Sonderkommando prisoners, the contextualization of their situation provides the most authentic evidence. The SCI Grapevine emphasizes how the Sonderkommando were not forced to submit to their role, and many volunteered. The diary of Rudolf Höss similarly recounts a certain eagerness of the Sonderkommando to lie to the prisoners, depicting them as unimaginably callous figures in a weak attempt to justify his own unscrupulous behavior. However, the various forms of resistance by the Sonderkommando prisoners directly dispute claims against their general moral standing. As Dan Stone describes in "The Sonderkommando Photographs", the mere existence of photographs serves as an exemplification of the lengths to which individuals subjected to this role were willing to go to maintain visibility in the face of such circumstances. As depicted in Exhibit 1, an individual who smuggled film into Auschwitz hid behind a dark corridor to capture a moment of action - the burning of naked bodies outside, when the crematoria became too full. The pictures demand an awareness of the situation itself, and as Stone asserts, the very point of them is to illustrate crude death. The claims some historians have made against the Sonderkommando photographs, asserting that they are voyeuristic at their core, entirely discredit the context of the images, and the value of the medium as evidence. Moreover, many Sonderkommando prisoners, through their secret writings, emphasized the sense of embarrassment that washed over them when confronted with naked bodies. Keenly aware of the sentimental value of clothing, and the separation it represents from the inevitability of what comes, combined with the disorienting nature of such a scene, provoked many to struggle telling the prisoners to undress before going into the gassing room. Furthermore, the pictures are characterized by motion, and Appendix 3 is at such an angle that the perceived focus remains in the lower left-hand corner of the frame. Appendix 4 is shot from the hip of a Sonderkommando prisoner, and both images translate a sense of urgency on the side of the photographer, which suggests that the point was not to objectify, and rather to provide substance to such an unreal scene. Examining the depth behind the delegation of prisoners to the Sonderkommando evokes uncomfortable reflection, yet is crucial to maintaining a balanced understanding of the individual position and the indexical motivations behind those forced to realize such a role.

Works Cited

Greif, Gideon. *We Wept Without Tears: Testimonies of the Jewish Sonderkommando from Auschwitz*, Yale University Press, 2005. ProQuest Ebook Central, https://www.proquest.com/legacydocview/EBC/3419874?accountid=3672.

Mark, Ber, editor. *The Scrolls of Auschwitz*. Translated by Steven Lehrer, Academic Studies Press, 2020

"SCI Grapevine." *SCI Grapevine*, vol. 3, no. 59, Feb. 1972. JSTOR, https://jstor.org/stable/community.32495024. Accessed 22 Apr. 2024.

Stone, Dan. "The Sonderkommando Photographs." *Jewish Social Studies*, vol. 7, no. 3, 2001, pp. 131–48. JSTOR, http://www.jstor.org/stable/4467613. Accessed 10 Apr. 2024.

Secondary Sources

Brown, Adam. "La 'Zona Grigia': The Paradox of Judgment in Primo Levi's 'Grey Zone.'" *Judging "Privileged" Jews: Holocaust Ethics, Representation, and the "Grey Zone*," 1st ed., Berghahn Books, 2018, pp. 42–75. JSTOR, https://doi.org/10.2307/j.ctt9qd04w.6. Accessed 10 Apr. 2024.

Ingle, Leah Christine, "Witness and Complicity: The Scrolls of Auschwitz and the Sonderkommando" (2019). Masters Thesis. 573.https://digitalcommons.liberty.edu/masters/573

Lawless, Kate. "Memory, Trauma, and the Matter of Historical Violence: The Controversial Case of Four Photographs from Auschwitz." *American Imago*, vol. 71, no. 4, 2014, pp. 391–415. JSTOR, https://www.jstor.org/stable/26305100. Accessed 22 Apr. 2024.

Pytell, Timothy E. "Shame and beyond Shame." *New German Critique*, no. 117, 2012, pp. 155–64. *JSTOR*, http://www.jstor.org/stable/23357069. Accessed 18 Mar. 2024.

Prediction of the Distribution of Invasive Plants in Chesapeake Bay Region through Data Visualization and Modeling by Kwanhee Lee

Abstract

Invasive plants, if left unchecked, can dominate the entire ecosystem. Knowing whether certain areas are showing growth or decline of a portion of invasive plants, can be an effective first step in taking measures. The methods used are data visualization and modeling. The data visualization method shows the distribution of three major invasive plants: Japanese Honeysuckle, Garlic Mustard, and Multiflora Rose over the past 5 years. The number of those plants is compared to that of the three major native plants: Eastern Redbud, Red Maple, and Eastern White Pine. Through the prediction modeling, it shows that for every one native plant there are more than one invasive plant mainly in the Chesapeake Bay region. This means that there must be increased efforts to allow the natives to take back the ecosystem taken by the invasive plants.

Introduction

Over the past years, it is without a doubt that more and more invasive plants have become present and made themselves at home. Sometimes people introduce invasive plants intentionally or accidentally through various methods. And once introduced, they are spread by wind, water, wild animals, or people (Swearingen, 2010). It is also without a doubt that the number of invasive plants has increased rapidly through their special abilities. Some invasive plants have root systems that are much stronger than those of native plants, allowing them to spread great distances from a single plant, and some others produce chemicals that prevent other plants from growing nearby (Invasive Plants). These invasive plants have a negative impact on the ecosystem, local biodiversity, and environmental quality we live in (Pejchar et. al., 2009; Kueffer 2017; Jones et al., 2017; Bartz et al., 2019). But how much have they increased? Is this a serious problem for our ecosystem and something we need to worry about right away? Answering these questions is the first step in accurately understanding the impact of invasive plants on the ecosystem and establishing a control plan for them. Data visualization can be one of the ways to research the trend of their distribution using general graphic tools like geographical plotting (What Is Data Visualization?). The choropleth map brings together two datasets: spatial data which represents a division of a geographic region and statistical data which are collected within that space (Pedriquez, 2022). This map visualizes variables that vary across a geographical area and shows the level of variability within the region. The moving average method in data modeling can also be one of the ways to make a prediction of the future growth or decline of variables. It predicts the long-term trend from the historic data by "smoothing" short-term fluctuation (Hyndmanet al., 2018).

Literature Review

To study the distribution and growth trends of invasive plants in the Chesapeake Bay region which covers parts of six states of New York, Pennsylvania, Delaware, Maryland, Virginia, and West Virginia, and all of Washington, D.C, three species of invasive and native plants that most abundantly thrived in the region are selected. The top three most populated invasive and native plants in the Chesapeake Bay region were chosen for this research. The three invasive plants are Japanese Honeysuckle, Garlic Mustard and Multiflora Rose. The Three native plants are Eastern Redbud, Red Maple, and Eastern White Pine.

The first invasive plant species studied in my research is the Japanese Honeysuckle. This plant, originated from East Asia, covers up everything in the way by twirling around small trees and shrubs (Japanese Honeysuckle (Lonicera Japonica, 2010). Because of its dense vegetation, young trees and shrubs do not stand a chance against Japanese Honeysuckle and collapses. This is how this plant grows and finds its way to bigger trees and takes over the entire ecosystem.

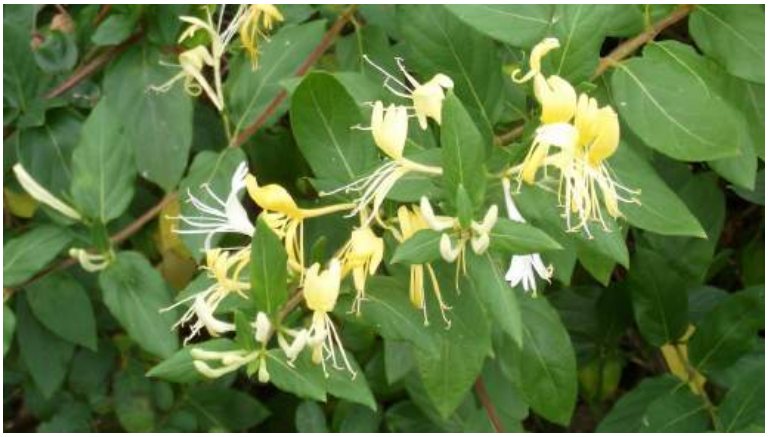


Fig 1: Image of invasive Japanese Honeysuckle photo by Richard Gardner (Japanese Honeysuckle2023)

Next invasive plant is Garlic Mustard. This plant was present in the U.S because of its herbal and medical qualities. However, because Garlic Mustard emerges earlier than most of the native plants, it prevents native plants from receiving sunlight. Also, Garlic Mustard starts its growth earlier, and thus it outcompetes the young sprouts for vital nutrients and moisture (Garlic Mustard: Invasive, Destructive, Edible, 2020).

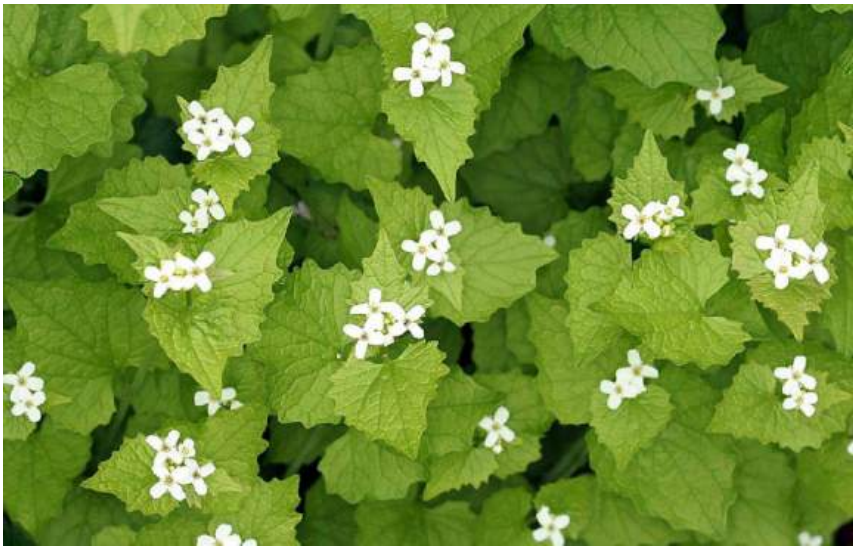


Fig 2: Image of invasive Garlic Mustard photo by David Cappaert (Garlic Mustard, 2022)

Last invasive plant is the Multiflora Rose. This shrub is notoriously difficult to kill. There is no predator that can effectively check its population nor any disease. It usually resides near sunlight, but absence of it does not hinder its growth. One plant can successively produce up to 500,000 seeds (Wenning, 2012).

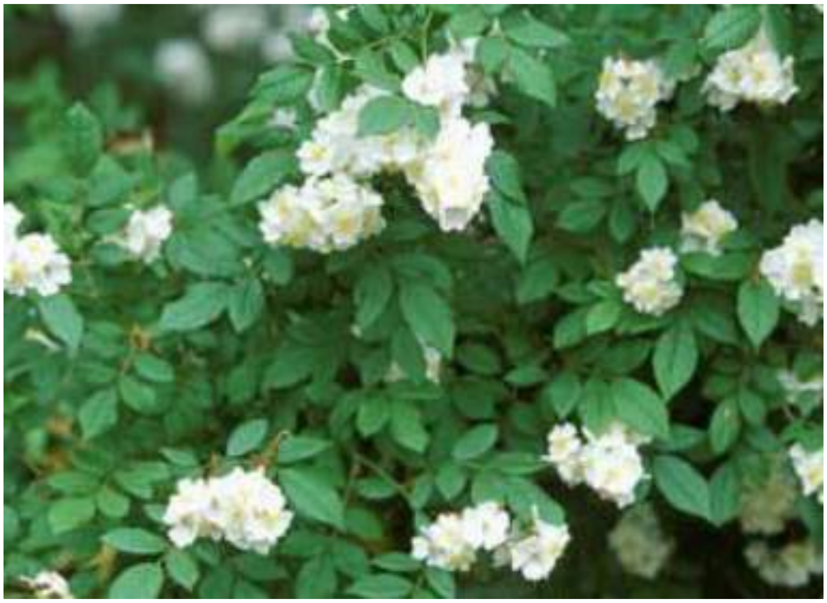


Fig 3: Image of invasive Multiflora Rose photo by James H. Miller (Multiflora Rose)

The first native plant species is called Eastern Redbud trees. These trees attract pollinators and beneficial insects. It also attracts native birds which can increase biodiversity. Another benefit of this tree is that it has a root system that can hold soil from erosion (Eastern Redbud).

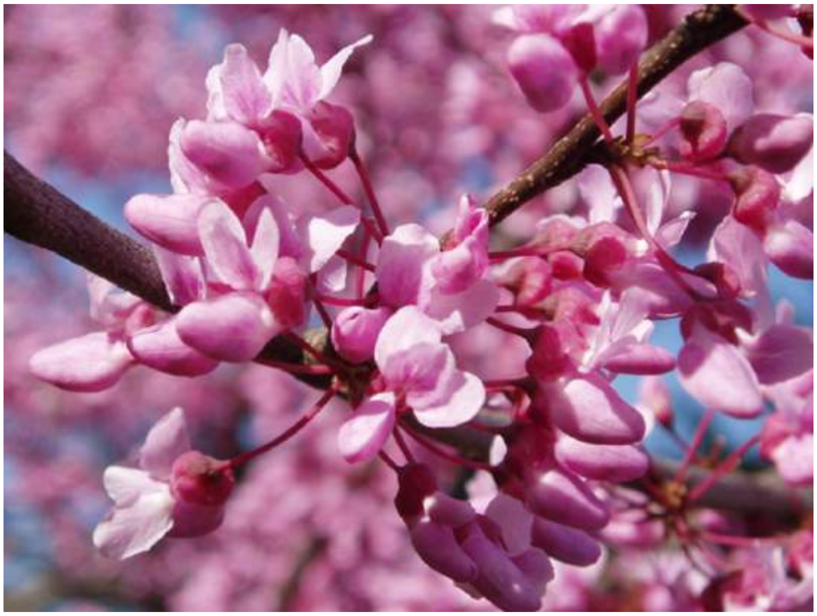


Fig 4: Image of native Eastern Redbud photo by Julie Markin (Cercis Canadensis)

Next native plant is Red Maple. Not to mention the numerous benefits it has on humans such as treating inflamed eyes and cataracts, the seeds, buds, and flowers offer a food source for squirrels, chipmunks, deer, mooses, elks and rabbits (Acer Rubrum).

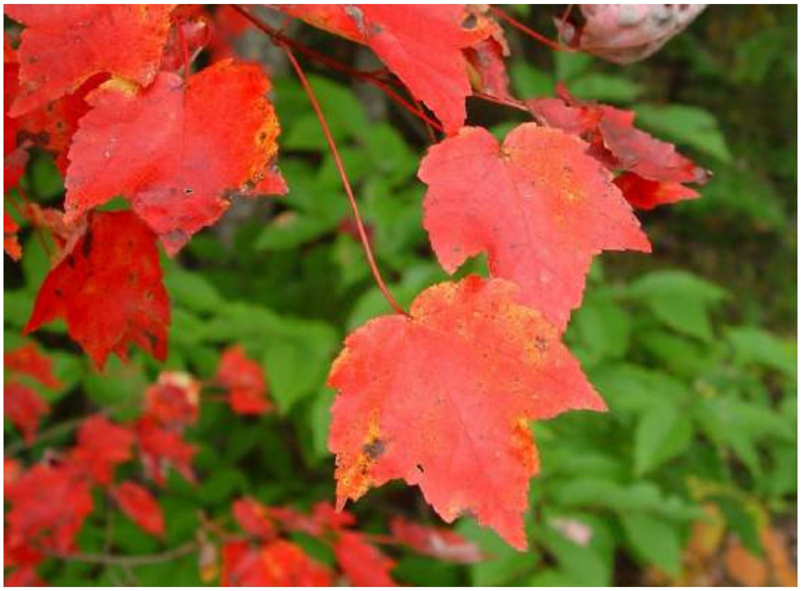


Fig 5: Image of native Red Maple photo by Arthur Haines (Acer Rubrum-Red Maple)

Last native plant is an Eastern White Pine. This tree offers a food source for many forest mammals such as squirrels, chipmunks, mice, and hare. In addition to that, this tree is a common breeding habitat for many birds such as owls, sparrows, and many types of warblers (Pinus Strobus).

Fig 6: Image of native Eastern Pine Tree photo by Ed Reschke (Nix, 2022)

The data used for this research was gathered from the iNaturalist, which is a social networking platform where people record and share the observations of biodiverstiy by taking pictures and uploading them (INaturalist, 2024). iNaturalist provides biodiversity data identified and verified to a high taxonomic resolution along with metadata for the users in citizen science such as myself (Callaghan et al., 2022). Simple moving average is one of the modeling tools applicable for the time-series data with no trend nor seasonality. This method simply uses an arithmatic average of the actual data over a set of time to predict the future trend (Polanitzer, 2022). The value of the moving average (MA) for a length of n is the summation of actual historical data (Yi) in time sequence i.

$$MA_n = \frac{\sum_{i=1}^{n} Y_i}{n}$$

In choosing the value of n, there is a trade-off between two effects: filtering out more noise vs. being too slow to respond to trends and turning points (Nau, 2014).

Contextualizing Research

During my volunteer work in the city of Rockville to remove invasive plants, I have learned about how the invasive plants have the ability to absolutely demolish the food web and destroy an entire ecosystem. These plants can have a devastating impact on human lives as well. One example of this negative impact is the loss of pollinators. We rely on pollinators which cooperate with native plants to live and assist the growth of crops such as almonds and apples. My motive for starting this project was because I wanted to visualize if the invasive plants were indeed taking over the native ecosystem and at what rate in order to fully understand how widespread and imminent the danger was in the Chesapeake Bay region where I reside.

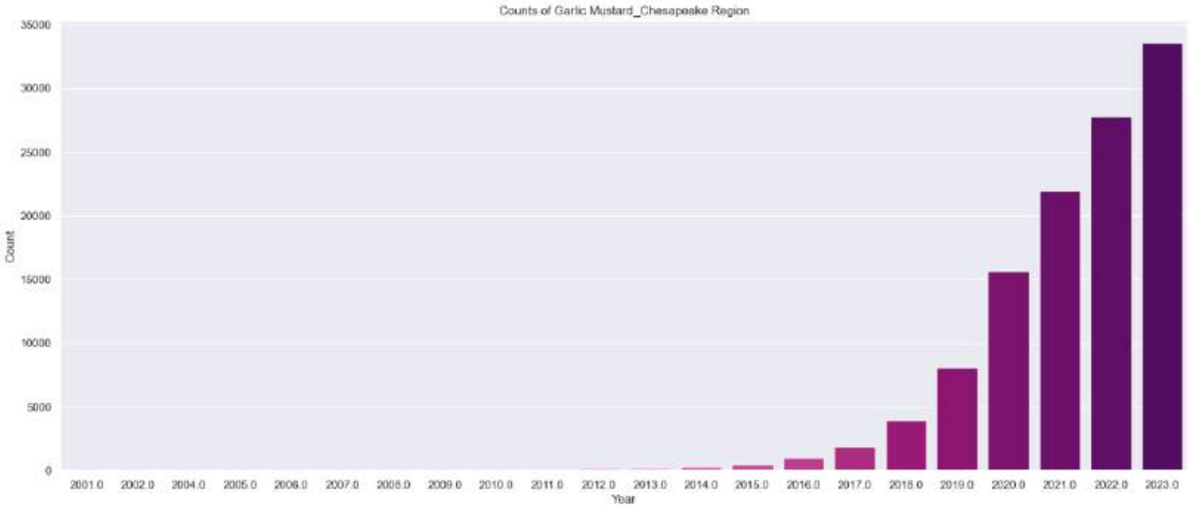
Experiment

The data gathered from iNaturalist was prepared to analyze according to the year (date of detection) and the location (place of detection). The information about the date came as “year-month-day” and “year” was separated in order to analyze the data by year. The same process was done to the data about the location. The data came as “Country-State-County” and “Country” and “State” were separated for the same purpose of analysis. From looking at the latitude and longitude of all the data, it was safe to assume that no two plants were the same plants. Because of this information and the hypothesis that plants do not disappear once introduced, the annual counts of the plants could be cumulative.

For the data visualization by location (state), the choropleth map was drawn in python by using plotly. The number of plants by state are plotted on the map of the United States from 2018 to 2023, and the distribution of the plant is studied by the comparison of the color of each state according to the colorbar scale.

The annual count of invasive plants is calculated by summing up all the counts of three invasive plants to minimize the abnormality of the data for each plant and the same work is done for the native plants. Then, the annual count of invasive plants is divided by that of native plants to calculate the ratio of invasive per native. Based on the historic data from 2001 to 2023, the simple moving average forecasting model is designed. By splitting the data into training and testing sets, the algorithm is assigned some portion of the data for training and uses the rest of the data to test the accuracy. For this model, the first 12 out of 23 data points were chosen to be the training data, and the rest as test data. The value of 3 is chosen for the simple moving averaging and the model gives the accuracy by calculating RMSE (Root Mean Squared Error) which is based on how well it predicted the test set.

Results



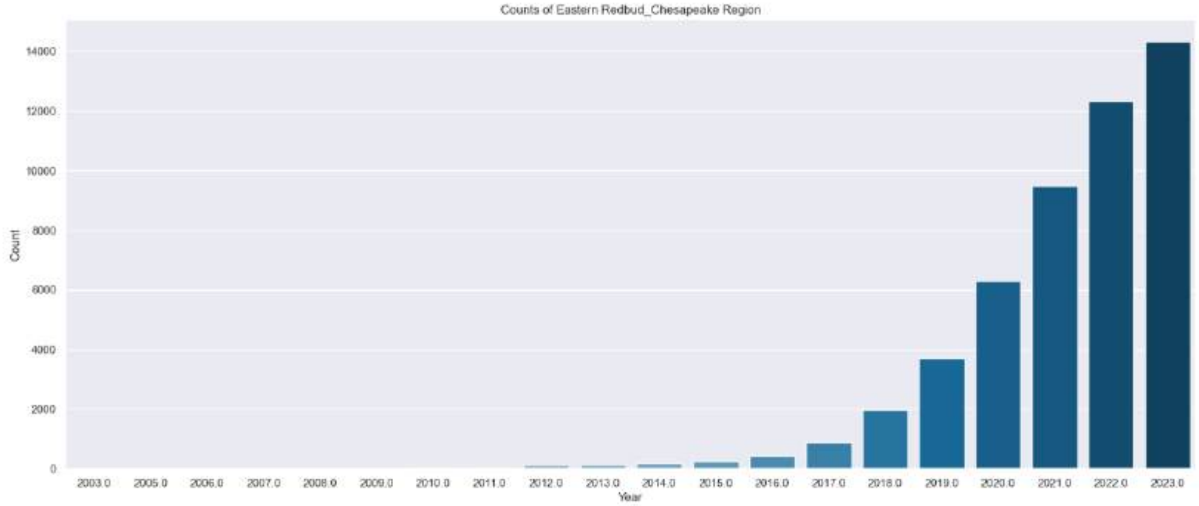
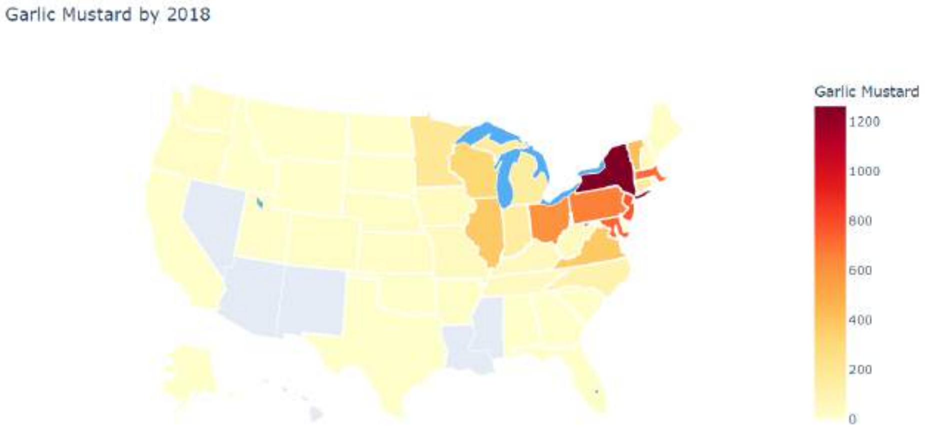
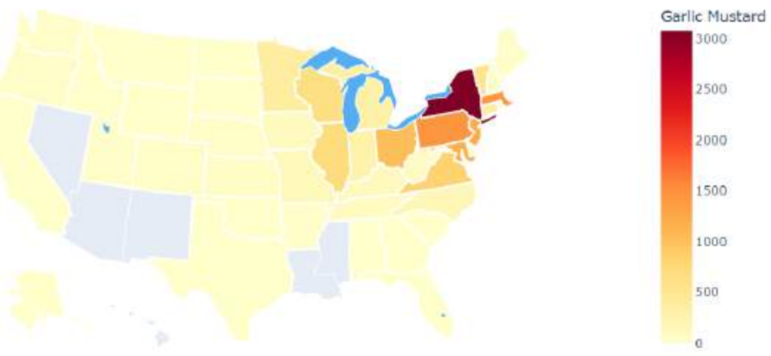


Fig 7: Annual counts of Garlic Mustard and Eastern Redbud in Chesapeake Bay region from 2001 to 2023

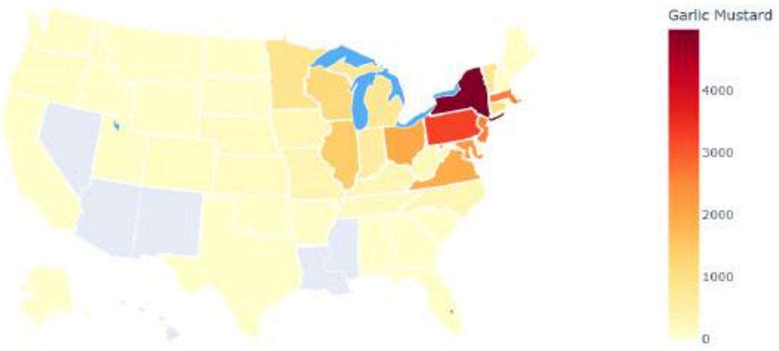
Because the data was gathered from a citizen science platform, there might have been a chance that the data was biased and some plants could have had excessively more data than the others. Garlic mustard and Eastern Redbud are one example each from the invasive plant and the native plant category, and they show that the number of data are well balanced. The graphs also show that in the Chesapeake Bay regions, the two plants are increasing and the other four plants show the same pattern.



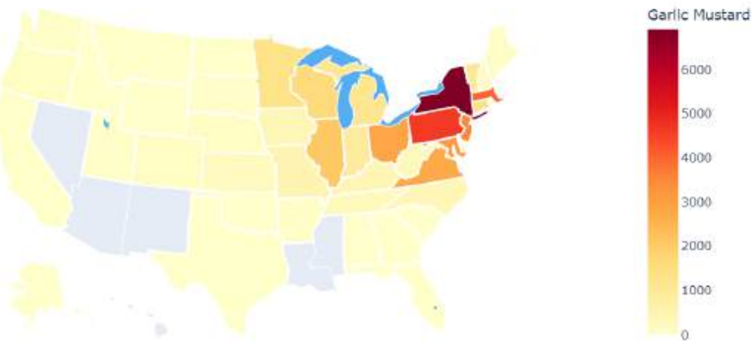
Garlic Mustard by 2019



Garlic Mustard by 2020



Garlic Mustard by 2021



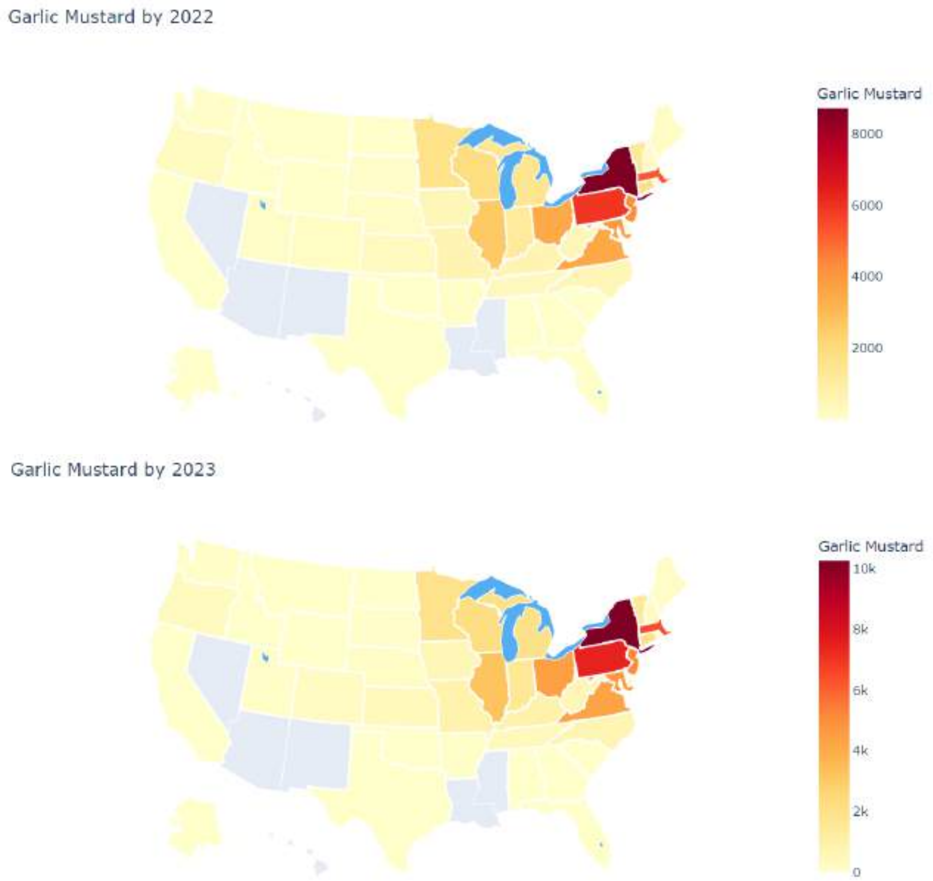
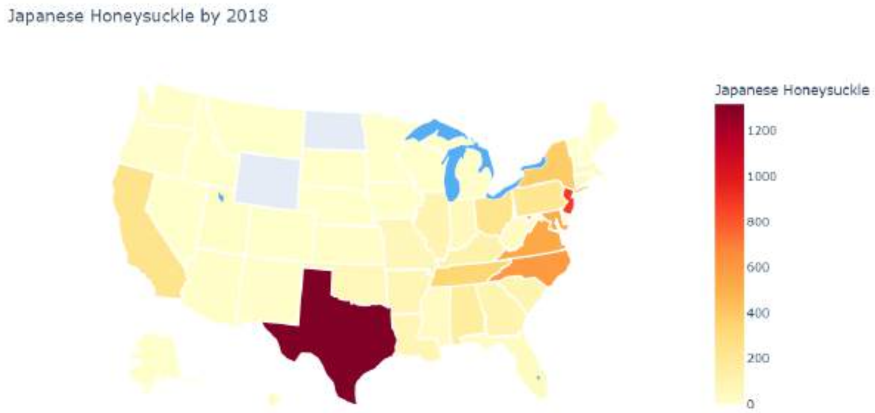


Fig 8: The Choropleth maps of Garlic Mustard across the United States from 2018 to 2023

Garlic Mustard was again used as one example to visualize the distribution and its growth patterns through the state. Shown in the images, the distribution of where Garlic Mustard thrives does not seem to change over the years in most states. However, as the year increases, the colors of the states of Pennsylvania, Virginia, and Maryland become darkened which means the number of Garlic Mustard increases rapidly in these states. Other two invasive plants, Japanese Honeysuckle and Multiflora Rose, show similar patterns.



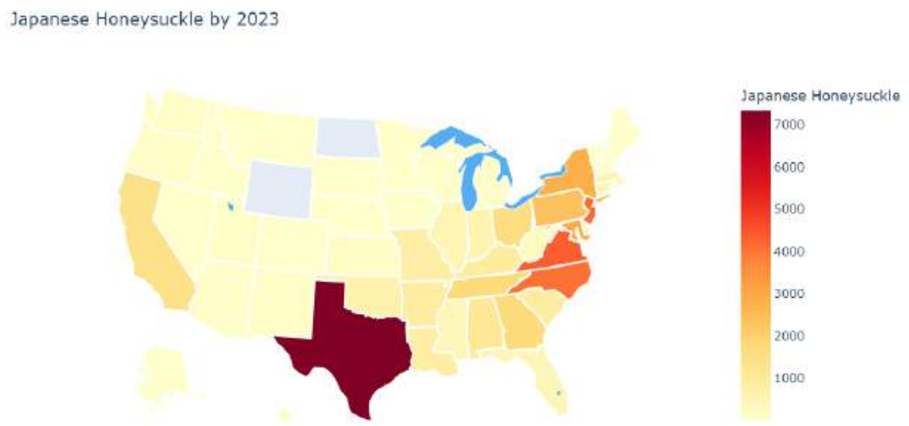


Fig 9: The Choropleth maps of Japanese Honeysuckle across the United States from 2018 to 2023

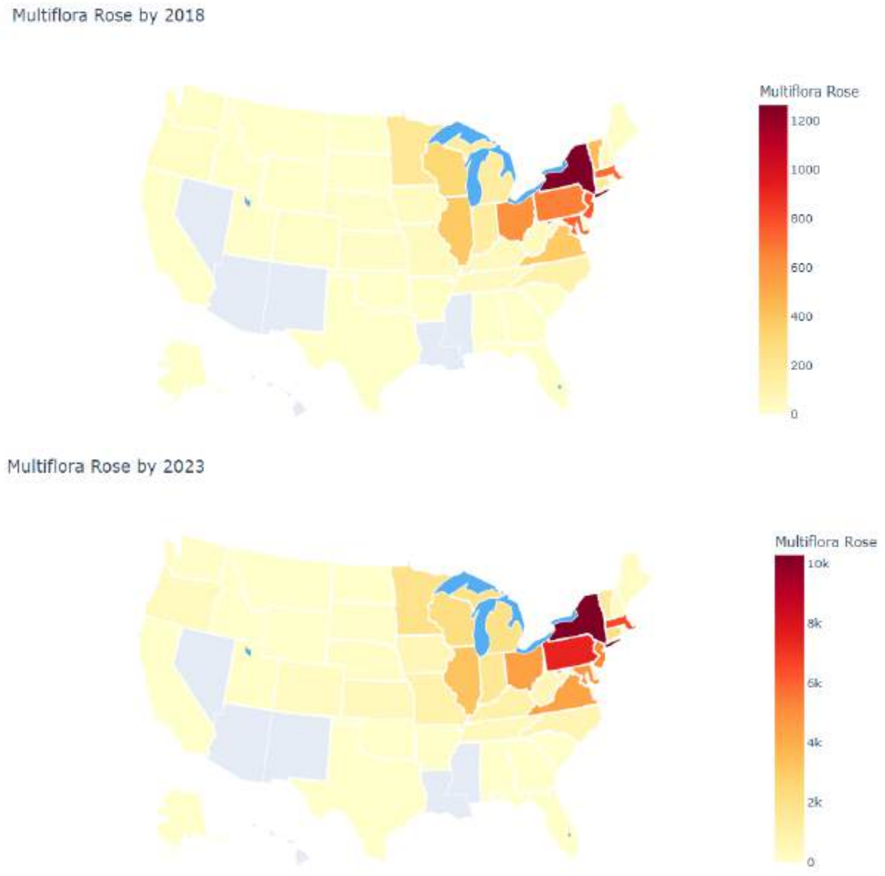


Fig 10: The Choropleth maps of Multiflora Rose across the United States from 2018 to 2023

These are the visualizations of 2018 and 2023 distribution of Japanese Honeysuckle and Multiflora Rose. Even though Texas is an outlier for Japanese Honeysuckle, it can be seen that the two other invasive plants demonstrate an increasing trend especially in states of Chesapeake Bay regions, similar to Garlic Mustard.

Based on the data gathered, the trend can be seen as shown below.

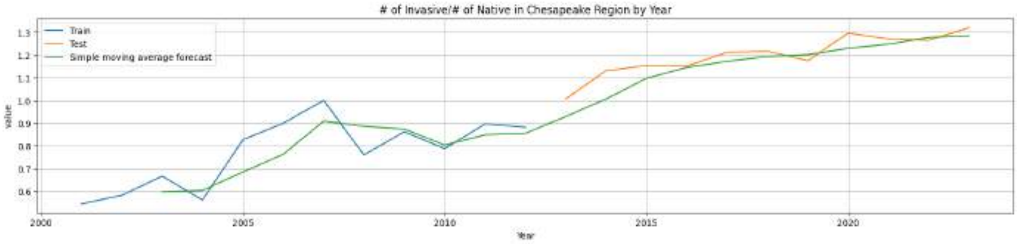


Fig 11: Time-Series plot of ratio of the number of invasive plants to the number of native plants with 3-year simple moving average forecast

This graph shows the ratio of the number of invasive plants to the number of native plants. The graph is obtained by dividing the total number of three invasive plants by the total number of three native plants. If that number is greater than 1.0, it indicates that there are more than one invasive plant for every one native plant. The number has surpassed 1.0 and has reached 1.3 in 2020. The RMSE of this model is 0.06. This number shows the difference between the actual value and the predicted value by squaring then square rooting the number. Since the error is 0.06, this predicted model can be seen as credible.

Discussion and Conclusion

The limitation of this research is the use of citizen science data. Since citizen science data are collected by people, there is a chance that some data may include bias in them. For instance, people may have found one type of plant more often than other just because it was present near a residential area, or because of its distinguishing features. In order to remedy the potential bias, this experiment incorporated three plants from invasive and native groups. The first set of graphs shows the increase of the population of the three invasive plants especially in the Chesapeake Bay region over the span of five years. By itself, it shows that invasive plants have increased over time, which might be a natural occurrence for any plant in this region. The forecast graph, however, shows that invasive plants have surpassed the number of native plants. Generally, an ecosystem starts out with more native plants than non-native plants. But in the Chesapeake Bay region, there are now more invasive plants than there are native plants. This means that the growth in the number of invasive plants has stagnated the growth in the number of native plants. As a result, there must be increased efforts to eradicate the invasive plants in order to ensure that native plants have a place to thrive in the ecosystem.

Acknowledgement

I acknowledge the co-leaders of the “Weed Warrior” program in the city of Rockville, Chuck Woolery and Monica Berger, for the motivation and interest to carry out this project

Works Cited

"Acer Rubrum." Lady Bird Johnson Wildflower Center, Accessed January 15, 2024. www.wildflower.org/plants/result.php?id_plant=acru.

"Acer Rubrum-Red Maple." Native Plant Trust Go Botany, Accessed January 15, 2024. https://gobotany.nativeplanttrust.org/species/acer/rubrum/.

Bartz, Robert, and Kowarik, Ingo. "Assessing the Environmental Impacts of Invasive Alien Plants: A Review of Assessment Approaches." NeoBiota 43 (2019): 69-99. Accessed December 01, 2023. https://neobiota.pensoft.net/article/30122/.

Callaghan, Corey T., et al. "The Benefits of Contributing to the Citizen Science Platform INaturalist as an Identifier." PLOS Biology 20, 11(2022). Accessed January 11, 2024. https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3001843.

"Cercis Canadensis." Lady Bird Johnson Wildflower Center, Accessed January 15, 2024. https://www.wildflower.org/plants/result.php?id_plant=ceca4.

"Eastern Redbud." Arbor Day Foundation. Accessed January 15, 2024. https://shop.arborday.org/eastern-redbud.

"Garlic Mustard." Cornell University Cooperative Extension. Last modified July 02, 2019, Accessed January 15, 2024. https://nyis.info/invasive_species/garlic-mustard/.

"Garlic Mustard: Invasive, Destructive, Edible." The Nature Conservancy. Last modified July 22, 2020, Accessed January 15, 2024. www.nature.org/en-us/about-us/where-we-work/united-states/indiana/stories-in-indiana/garlic-mustard/.

Hyndman, Rob J., and Athanasopoulos, George. "6.2 Moving Averages | Forecasting: Principles and Practice." Last modified June 20, 2024, Accessed January 15, 2024. https://otexts.com/fpp2/moving-averages.html.

"INaturalist." Wikipedia. Last modified May 02, 2024, Accessed July 01, 2024. https://en.wikipedia.org/wiki/INaturalist.

"Invasive Plants." US Forest Service. Accessed September 05, 2023. https://www.fs.usda.gov/wildflowers/invasives/index.shtml.

"Japanese Honeysuckle." University of Maryland Extension. Last modified February 22, 2023, Accessed January 15, 2024. https://extension.umd.edu/resource/japanese-honeysuckle/.

"Japanese Honeysuckle (Lonicera Japonica)." Invasive.org.. Last modified November 11, 2010, Accessed January 15, 2024. www.invasive.org/alien/pubs/midatlantic/loja.htm.

Jones, Benjamin A., and McDermott, Shana M.. "Health Impacts of Invasive Species through an Altered Natural Environment: Assessing Air Pollution Sinks as a Causal Pathway." Environmental and Resource Economics, 71 , 1 (2018): 23-43. Accessed December 01, 2023. https://link.springer.com/article/10.1007/s10640-017-0135-6.

Kueffer, Christoph. "Plant Invasions in the Anthropocene." Science 358, 6364 (2017): 724-725. Accessed October 13, 2023.

"Multiflora Rose." USDA National Invasive Species Information Center. Accessed January

15, 2024. https://www.invasivespeciesinfo.gov/terrestrial/plants/multiflora-rose.

Nau, Robert. “Forecasting with Moving Averages.” Fuqua School of Business, Duke University. Last modified August 01, 2014, Accessed March 05, 2024. https://people.duke.edu/~rnau/Notes_on_forecasting_with_moving_averages--Robert_Nau.pdf.

Nix, Steve. “13 Most common North America Pine Species.” Treehugger. Last modified July 03, 2024, Accessed July 05, 2024. www.treehugger.com/most-common-north-american-pine-species-1341866

Pedriquez, Daleska. “What Is a Choropleth Map and How to Create One.” Venngage. Accessed January 15, 2024. https://venngage.com/blog/choropleth-map/.

Pejchar, Liba, and Mooney, Harold A.. “Invasive Species, Ecosystem Services and Human Well-Being.” Trends in Ecology & Evolution 24, 9 (2009): 497–504. Accessed September 03, 2023.

“Pinus Strobus.” Lady Bird Johnson Wildflower Center, Accessed January 15, 2024. https://www.wildflower.org/plants/result.php?id_plant=PIST.

Polanitzer, Roi. “Time-Series Methodologies - Part 1: Single Moving Average and Error Estimation.” Medium. Last modified January 01, 2022, Accessed July 05, 2024. https://medium.com/@polanitzer/time-series-methodologies-part-1-single-moving-average-and-error-estimation-34b6518797a8.

Swearingen, Jil. “Plant Invasive of Mid-Atlantic Natural Areas, 5th ed.” (2014): 8-9, Accessed September 03, 2023. https://www.invasive.org/alien/pubs/midatlantic/midatlantic.pdf.

Wenning, Bruce. “Multiflora Rose: An Exotic Invasive Plant Fact Sheet.” Ecological Landscape Alliance. Last modified July 16, 2012, Accessed January 15, 2024. www.ecolandscaping.org/07/landscape-challenges/invasive-plants/multiflora-rose-an-exotic-invasive-plant-fact-sheet/.

“What Is Data Visualization?” IBM. Accessed January 15, 2024. www.ibm.com/topics/data-visualization#:~:text=Data%20visualization%20is%20the%20representation.

Unlocking the Versatility of Titanium Dioxide: Exploring Nanoparticle Structural Properties in Metal Oxides By Aashi Shah

Abstract

Nanoparticles, the fundamental building blocks of atoms, possess unique structural properties that profoundly influence the functionality of materials. This research study looks into the structural properties of metal oxides with a particular focus on titanium dioxide (TiO2). Understanding the determining factors that contribute to its wide range of applications had been the goal of the inquiry. A complete and succinct literature review highlighted many structural characteristics of metal oxides and the analysis of the chemical aspects of said characteristics. Few investigations have evaluated the singular structural property of titanium dioxide that allows its widespread functionality and that sets up the research on this topic. Resulting from an in-depth review of sources looking for recurrences of specific structural characteristics including Crystal Structure Phases (rutile, anatase, brookite), Surface Area/Morphology (shape), Defects & Doping, Interaction with Light & Electricity, and Structural Integrity, the investigation found the interactions with light and electricity as the most significant structural property, influencing applications ranging from sunscreen formulations to pigments in coatings to photocatalysis in concrete. Furthermore, this exploration underscores the integration of titanium dioxide's many structural properties and their collective impact on its multifaceted applications. While the research provides valuable insights, limitations such as the limited number of reviewed documents and the potential of human errors require the need for further investigation of the subject. Regardless, this study indicated the critical role nanoparticles' structural properties play in optimizing the utilization of titanium dioxide across its various industries, setting a path towards better goals of editing nanoparticles and maximizing their efficiency.

Introduction

Every object, material, and item in the world around us comprises a set of pure substances called elements. Within that, each element contains atoms. Now in between the scale of an element and an atom, there are what we call *nanoparticles.* Nanoparticles are the tiny building blocks of atoms all integrated to form the element.

What was found to be the most interesting in these building blocks is that the tiniest tinker in the shape of the nanoparticle would completely alter the functionality of the element. To go even further, investigating how these alterations in the structure would cause different functions was a good approach to looking further into the topic. Through that newfound knowledge, scientists would be able to better understand how to manipulate nanoparticles to create the optimal structure for an intended application.

After preliminary research, it was determined that the topic of nanoparticles is extensive, so the focus shifted to the genre of metal oxides and one particular substance, namely titanium dioxide (TiO2). Choosing this genre comes from the fact that metal oxides are singular in the world of nanomaterials due to their interesting structural properties including high surface area (how big they are), variable pore size (size of the holes in the material), and stability. The properties of certain elements and materials are highly dependent on the structures of these nanoparticles.

Focusing on titanium dioxides and understanding the optimal structural properties can allow engineers to better control their performance in industrial and commercial applications. Correspondingly, contradictory to public thought, TiO2, the metal oxide, is in the majority of everyday products. Whether it be the pigments in paints, the protective layer in sunscreen, or a green energy resource such as a battery, humans consistently come in contact with titanium dioxide which has those special properties allowing it to do special tasks. Finding a common, optimal structural property of these materials would thus allow these commercial products to be constructed optimally. With that in mind, this research was conducted in order to answer the question "What structural property of titanium dioxide contributes to its versatility across a wide range of applications?"

Literature Review

According to a research article from Chongqing University of Science & Technology, nanotechnology is a tiny technology that can be scaled to the millionth of a millimeter (Ou, 2021). The characterization of the nanotechnology properties lies in which types of materials these nanoparticles are. There is an ample number of the types of compounds in the world that make up these nanomaterials. In his discussion of nanomaterials, a Staff News Editor at Nanotechnology Weekly reported that a compound is described as an element of the periodic table bonded to another (Nanotechnology Weekly, 2020). Several nanotech researchers take the position that the strongest nanotech is metal, however, a new understanding portrays that metal oxide compounds are the prime nanotech material. These compounds, as described by Professor Zumdahl of the University of Illinois' Department of Chemistry, comprise a metallic element and the element of Oxygen (Zumdahl, 2011). On those grounds, studies indicate that certain characteristics allow metal oxides to succeed more than any other compound in nanotech.

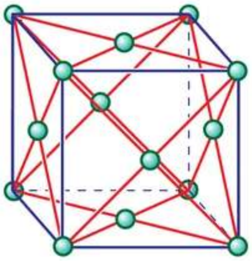
Characteristics of Metal Oxides

Several hypotheses concerning the nature of nanoparticle characteristics have been advanced by various authors but all have a consistent characterization of metal oxides. The collection of these characteristics demonstrates how metal oxides are superior to other compounds in the way they enable them to excel in a wide range of industrial, commercial, and scientific applications. Firstly, according to Professor Siva Prasanna and peers from the Department of Mechanical Engineering at the University of Petroleum and Energy Studies in Dehradun, India, it has a large surface which increases its reactivity and interaction with other substances (Siva Prasanna, 1905). With greater surface area, these particles have more space and a greater chance of reacting with other particles allowing them to form stronger bonds. Therefore, metal oxides are primary as they have enhanced reactivity. As Bawoke Mekuye and Birhanu Abera from the Department of Physics at Mekdela Amba University in Awuliya, Ethiopia further expand, it also exhibits chemical and thermal stability which means they can withstand harsh environmental conditions of high temperatures (Mekuye, 2023). That characteristic allows metal oxides to be widely used in many applications across different environments due to the way it ensures durability and reliability. Finally, as noted in an article by a Professor of Physics at Pennsylvania State University Gerald D. Mahan, they have crystalline structures that provide mechanical strength against strain and tension (D. Mahan, 2019). This quality is significant because crystalline structures are usually the characteristic of the strongest nanoparticles. To explain, take a look at the image to the left.

Nanostructures

The reasoning behind the structure of bonded nanoparticles is largely based on chemistry. Chemical bonds form between atoms through the sharing/transfer of electrons. A study by Professors Chaohai Ou and DeWei Wang at the University of South Carolina advances the notion that in the case of metal oxides, these bonds typically involve the transfer of electrons from the metal atoms to the oxygen atoms, leading to the formation of ionic bonds (Ou, 2021). According to Encyclopædia Britannica, ionic bonds can be defined as the type of bond "formed from the electrostatic attraction between oppositely charged ions in a chemical compound". Further investigated by Professor Zumdahl, this results in the metal atoms carrying a positive charge (cation) and the oxygen atoms carrying a negative charge (anion) (Zumdahl, 2011). The strong electrostatic attraction between these oppositely charged ions holds the structure together.

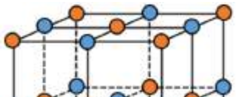
To further elaborate, compounds of metals and oxides make up close-packed structures where metal atoms and oxygen atoms are arranged in layers with the metal cations occupying the gaps between the oxygen anions and vice versa as seen in the image on the right. The alternating positions of metal and oxygen atoms create a repeating pattern, forming a crystalline structure (D. Mahan, 2019). Professor Ou and Wang elaborate on this pattern as this arrangement maximizes the packing efficiency, making the structure dense and contributing to the stability and strength of metal oxide nanoparticles (Ou, 2021).

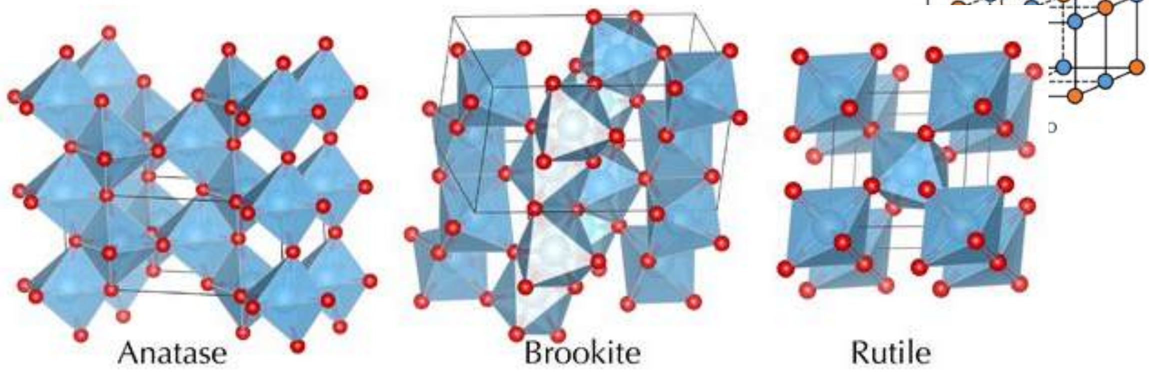


Characteristics of Titanium Dioxide

In the case of titanium dioxide, this compound has two Oxygen atoms bonded to one Titanium atom. After determining that metal oxides are significant due to their versatility, looking at the most versatile metal oxide, titanium dioxide, was indispensable. This powerful compound has several desirable structural properties giving it its high versatility but it can be narrowed into five categories.

1. Crystal Structure phases (anatase, brookite, rutile)





The crystal structure is characterized by various arrangements of atoms within a crystal lattice of the nanoparticle. There are three such structures a titanium dioxide nanoparticle can appear as, namely anatase, brookite, and rutile. The rutile and anatase phases have a tetragonal crystal structure and brookite is orthorhombic. Of the three phases, anatase has the lowest density while rutile has the highest. Anatase and brookite are less stable forms of titanium dioxide in comparison to rutile, the most stable phase. Moreover, titanium dioxides in the rutile phase have a higher refractive index compared to anatase and brookite.

2. Surface Area/Morphology(shape)

   Surface Area and Morphology refer to the outer layer of the nanoparticles and their physical appearance. Surface area is a unique quality to nano-scaled particles because, unlike bulk materials, they have a much larger surface area-to-volume ratio. For reference, a 1x1x1 cube's volume would be 1cm3 but its surface area would be 6cm2. With a higher amount of surface area per volume, a nanoparticle's reactivity increases. Similarly, the morphology- or shape of the particle- determines many of the particle's properties. The three structural phases from the first category denote the shape of the nanoparticles, however, morphology denotes the shape of the *collection* of nanoparticles otherwise known as the nanomaterial. Different shapes may include nanotubes, nanowires, spherical, oval, cubic, prism, helical, or pillar.

3. Defects and Doping

   Defects are described by irregularities and imperfections in the structure of nanomaterials that can happen during the manufacturing process. These may come in the form of missing or misplaced atoms which have the effect of manipulating the nanoparticle's strength and conductivity. Similarly, doping is the intentional creation of such impurities as explained in an article by medical researcher Aniket Koley in 2018 (Koley, 2018). Scientists and engineers use the process of doping in order to fine-tune certain properties of titanium dioxide with the use of equipment that manipulates these nanoscale materials.

4. Interaction with Light and Electricity

   There are two main aspects under this category: Ultra Violet (UV) light reactions and conductivity. Titanium dioxide can reflect UV light like a mirror which helps protect materials from the Sun's rays. Additionally, it absorbs some of the UV light and converts it to harmless energy. The reflectivity and absorbing ability of the nanoparticle is determined by the refractive index which corresponds to a study conducted by students at the University of Missouri-Kansas's Department of Physics claiming that rutile has the highest index (Praveen, P., 2019). As for electricity, titanium dioxide isn't a good conductor of electricity but when the structure is doped as Koley mentioned previously, the nanoparticle can become a semiconductor or essentially can conduct electricity better than before (Koley, 2018). Moreover, TiO2 becomes more conductive when exposed to light thus generating an electric current. Overall, titanium dioxide's structural properties protect materials from UV light and enhance their electrical conductivity.

5. Structural integrity

Titanium dioxide nanoparticles are valued for their mechanical strength which comes from their crystalline structural lattice and their strong ionic bonds between the atoms. These structural properties allow them to resist deformation or fractures when undergoing pressure and stress.

Several studies point out that many of these structural properties are akin to titanium dioxide, and they reflect metal oxide's structural properties as well. These five structural properties of TiO2 will continue to be mentioned throughout the course of the paper since they are the primary focus of the compound. All five of the categories are integrated and connected which collectively makes titanium dioxides the most superior titanium dioxide. However, there was a lack of identification of a specific significant property. Many of the structural properties were sought for, but they lacked a certain property that was better than the rest. With that information, scientists and engineers would be better able to manufacture the ideal titanium dioxide for its various solutions. Per this issue, it was sought to determine which quality of metal oxides' structures causes such multifaceted abilities in their application.

Research Methodology

To determine the optimal structural property that allows titanium dioxide nanoparticles their versatile applications, it was vital to analyze previous studies. Considering the lack of access to equipment that can manipulate materials as small as 1 nanometer in width, studying credible articles on nanoparticles through the Ex Post Facto Method was the only viable solution.

This approach can be traced to another nanotechnology research study conducted by a Professor of Mechanical Engineering at the University of Texas at Dallas specializing in nanotechnology, Mr. Rodrigo Bernal Montoya. He conducted research regarding the mechanical and electromechanical properties of Piezoelectric Nanowires by reviewing both experimental and computational studies. Focusing on both experimental and computation works would be crucial to obtaining unambiguous data which creates an objective stance on the topic. Considering the effective and credible approach used to attain data, those data points will be applied in this study as well.

Common instruments of research in this method included document analysis, observations, and synthesis of these observations, however, no participants are to be used. By analyzing and reviewing these studies, recurring structural attributes across these experimental and computational studies were found. After that, conclusions to define the bigger picture in the world through further readings and document analysis could be made. Although this approach is very similar to Professor Montoya's, it differs in the fact that this study analyzes the occurrences of structural properties rather than analyzing the common themes of diverse sources. This is because this exploration is characterized by specific categories of data and aims to locate them in the text whereas the alternative method used by Professor Montoya is looking for an unknown answer. Additionally, with numerous categories to find data for, it would be difficult to quantify themes and label them accurately. That being the case, instead of looking for themes across various sources, this research focuses on finding trends of specific data in varying sources.

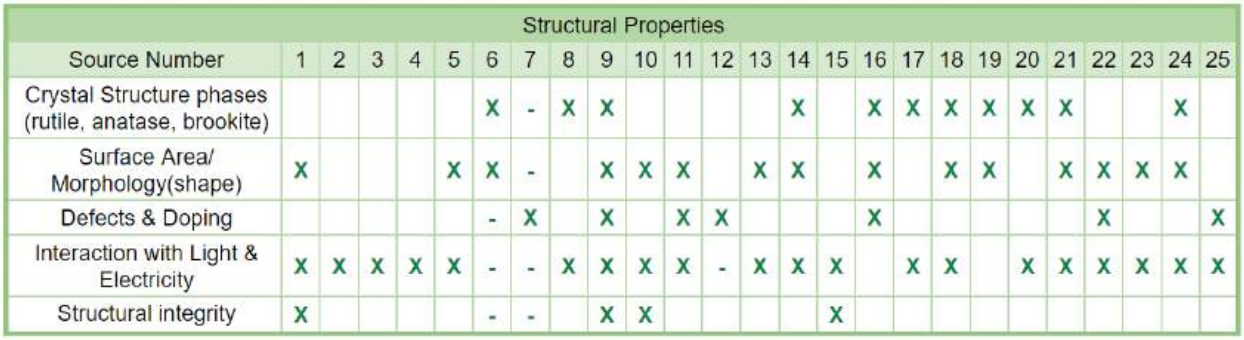
To organize this expansive data, the tables were utilized to organize observations. Two in-depth tables with observations of the structural properties and observations of the applications were essential to creating a claim, and thereby answering the question. The Structural Properties table codes the sources into the five categories of varying properties. These include Crystal Structure Phases (rutile, anatase, brookite), Surface Area/Morphology (shape), Defects and Doping, Interaction with Light and Electricity, and Structural integrity. Similarly, the Applications table codes the sources into six categories of applications as follows: concrete, grease and lubricants, plastic and rubber, pigments and coatings, use in the food industry, and sunscreen formulations. Tallying each category with the number of sources mentioning them had been an option for organizing the data, but this was not chosen because by looking at each source individually and the attributes mentioned, there was a possibility of finding trends across both tables with specific sources. Moreover, it would allow room to look at the range of the data and evaluate if certain applications correspond to particular structural properties.

A collection of 25 experimental and computational studies was used to work towards revealing information about the connection between the structural properties of titanium dioxide and its applications. While evaluating each source, any mention of the five structural properties in the article was tallied in the table. The same would be done for the Applications table. The frequency of a structural property being mentioned in studies determines the relative importance of the property for the marked application category on the Application table.

Observing and analyzing the effects of naturally occurring conditions of nanomaterials through Ex Post Facto allows one to gather existing data through research papers and draw correlations and conclusions. To ensure the credibility of the collected data, extra steps must be taken to make sure the sources are credible and peer-reviewed. Another factor playing into the credibility would be to ensure the publishing dates are within the 21st century. That would ensure that modern and accurate information tangential to the topic exists. Additionally, the large number of evaluated sources would allow for a diverse theme rather than biases. The results of the analysis are reported in the next section.

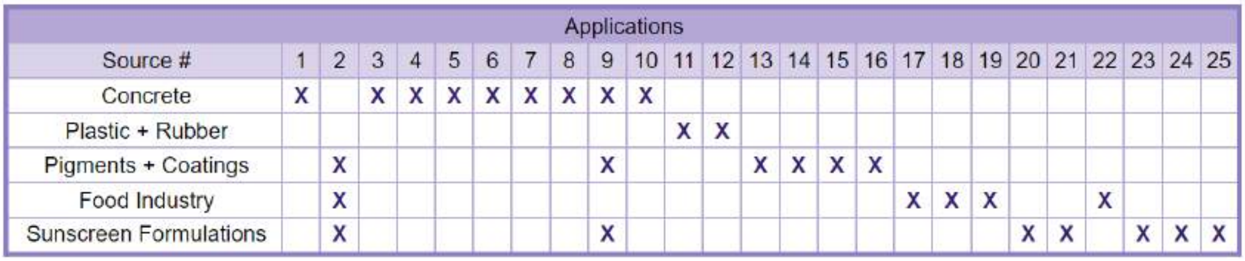
Data Collection

Data Table 1

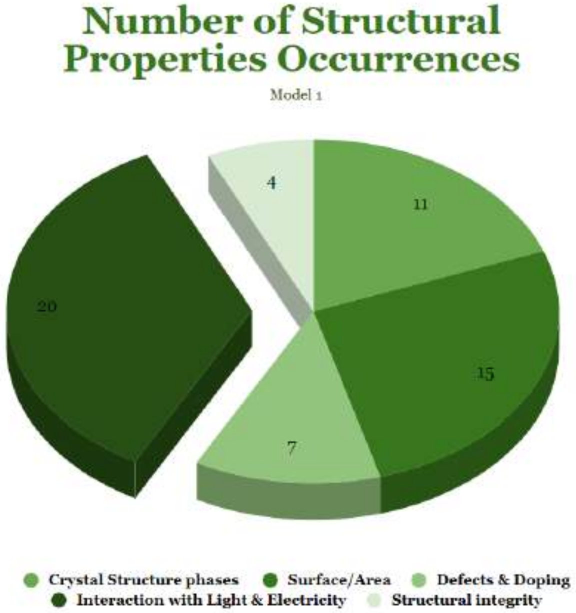


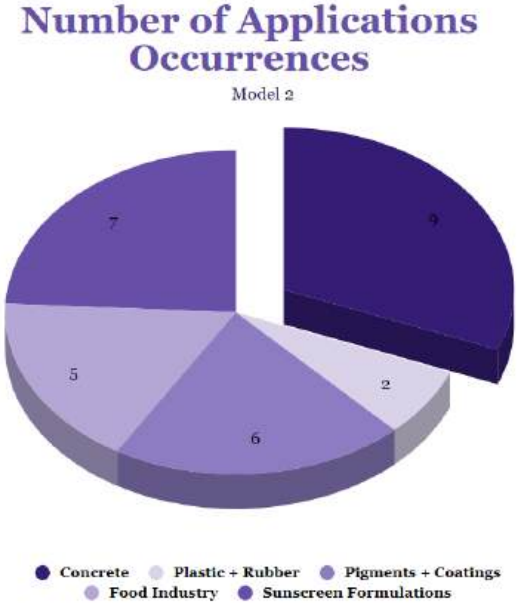
Structural Properties																									
Source Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Crystal Structure phases (rutile, anatase, brookite)						X	-	X	X					X		X	X	X	X	X	X			X	
Surface Area/ Morphology(shape)	X				X	X	-		X	X	X		X	X		X		X	X		X	X	X	X	
Defects & Doping						-	X		X		X	X				X						X			X
Interaction with Light & Electricity	X	X	X	X	X	-	-	X	X	X	X	-	X	X	X		X	X		X	X	X	X	X	X
Structural integrity	X					-	-		X	X					X										

Data Table 2



Applications																									
Source #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Concrete	X		X	X	X	X	X	X	X	X															
Plastic + Rubber											X	X													
Pigments + Coatings		X							X				X	X	X	X									
Food Industry		X															X	X	X			X			
Sunscreen Formulations		X							X											X	X		X	X	X





Analysis

Data Table 1, Structural Properties Occurrences Chart, displays the analysis of the 25 sources and which of the five structural properties each source highlighted. In a more digestible format, Model 1 indicates the net number of mentions for each structural property category. The Xs indicate an in-depth look at that structural property and the dashes represent a slight mention of the property. With a total of 23 Xs and dashes combined, the structural property of Interaction with Light & Electricity gained the most mentions throughout the sources. These mentions refer to the refractive qualities of the titanium dioxide nanoparticles which reflect and thus obscure UV light. The second-most recurring structural property had been the Surface Area and Morphology property with 16 mentions across all 25 databases. This property was marked when the article noted changes in the shape and size of the titanium dioxide nanoparticles. For example, Source 22, an open-access journal written by a professor at Iowa State University's Department of Ecology, Evolution, and Organismal Biology, details how the shapes of the titanium dioxide nanoparticles in products are sometimes spherical while others are shaped cylindrically indicating the importance of the structure of the nanomaterial. Conversely, nanoparticles' least recurring structural property had been the structural integrity category earning a total of 6 mentions. The sources implied that titanium dioxides weren't outstanding from the mechanical strength perspective, regardless of being the most powerful metal oxide nanoparticle.

Data Table 2 examines the different categories of applications of titanium dioxide nanoparticles. When searching for sources, the category of application was used to organize the content. That would explain the staircase pattern of the data collected. As indicated in Model 2, the most frequent applications had been in concrete with 9 mentions, and sunscreen formulations with 7 mentions. In both of these applications, titanium dioxides would be used as protectors and durability empowerers in contact with photons. On the other hand, the least common application category was plastic and rubber on account of its use as a pigment in those applications and not serving as an essential structural component.

Discussion

Per the findings illustrated in the previous section, mentions related to the fourth structural property category (Interactions with Light and Electricity) were prevalent indicating that the most recurring structural qualities of titanium dioxides prioritize the reflective and absorbing nature of its nanoparticle. That being the case, it can be concluded that the refractive qualities of titanium dioxide nanoparticles play an integral role in their versatility. This structural property makes titanium dioxide highly sought after in applications such as sunscreen formulations, where protection against harmful UV radiation is crucial to the product. An important aspect to uncover is understanding the science behind how titanium dioxides achieve these the property of interacting with light and electricity in a unique way and that can be attributed to photocatalysis. Photocatalysis, which will be addressed later, is vital to the property of interacting with light and electricity. This structural property is not limited to exclusively sunscreen, however. With mentions in articles discussing 16 other applications aside from sunscreen, the nanoparticles' interactions with light and electricity are a unique attribute of titanium dioxide regardless of the application. Source 4, with interaction with light and electricity marked in Table 1 and Concrete in Table 2, highlighted the process of photocatalysis which is when the TiO2 nanoparticles are activated by light energy. Through this exposure, TiO2 nanoparticles absorb photons (light quantum) and generate electron-hole pairs. This is essential to concrete because these pairs can then react with molecules on the surface including pollutants which results in the conversion of pollutants into harmless compounds. In a broader definition as explained by students at the University of Science and Technology of China, photocatalysis is a chemical reaction where the catalyst, in this study the titanium dioxide, absorbs photons from light and generates electron-hole pairs which ultimately amplify the catalytic process's efficiency (Low, 2020). Indicatively, the property of the nanoparticle's interactions with light and electricity plays a versatile role on its own. This fact regarding titanium dioxide and its structure makes it a valuable compound.

As mentioned before, this desirable characteristic integrates all the other five structural properties as well. The structural phases of rutile, anatase, and brookite determine the wide bandgap energy specifying what portion of ultraviolet (UV) light from the solar spectrum would be absorbed. This absorption promotes the generation of electron-hole pairs instantly when light touches it, thus allowing photocatalytic reactions. With a larger surface area of the nanoparticle, there'd be a greater number of reactions with photons. defects and doping play a role in the generation of electron-hole pairs. The titanium dioxide nanoparticle must be mechanically strong with good structural integrity in order to withstand prolonged exposure to light. Though all the structural properties are distinctive of each other, they each play a role in the most favored property of interactions with light and electricity. Therefore, the fourth structural property has prominence while the other four properties play a role in ensuring it. Overall, the flexible role of titanium dioxide nanoparticles stems from their interactions with light and electricity and accordingly underscores their magnitude of importance across diverse applications, as demonstrated through this investigation.

Limitations

Though these are the collective findings of experts in the field, only 25 documents have been reviewed. If a few of the documents had been biased, the data would have been skewed. That decreases the validity of the claim to a certain extent. Therefore, for future research, it would be best to look at many more randomized sources: possibly hundreds. Additionally, there is room for human error. It is unavoidable that some specific mentions might've been overlooked since half of the sources had been dozens of pages long. To mitigate this issue, possibly in future studies, there could be a set of keywords pertaining to each structural property and only those words are allowed to be used as indicative mentions of the property. That would ensure consistency across various sources.

Conclusions

The inquiry of nanoparticles began when there was a question of the strongest compound or element in the world. The answer was carbon nanoparticles. That became the prominent element overpowering all other candidates of strong compounds. When stepping away from pure mechanical strength, however, and looking at a new type of strength- strength in versatility- metal oxides came out on top. In the same way that carbon nanotubes find their strength from their optimal crystal structure, it was sure there was a particular structural property that metal oxides, particularly titanium dioxides, had that allowed for their versatility. Throughout the study, that property had been found and evaluated. Titanium dioxide is one of the most investigated metal oxide nanoparticles due to its versatility and looking at its crystalline and compact structure contributed to an understanding of the most optimal aspects. All in all, this study broadened an understanding of nanostructures and the special distinction of metal oxides. The findings emphasize the importance of the property of interaction with light & electricity in determining the functions of titanium dioxide nanoparticles. With a deeper understanding of these structural properties, researchers and industry professionals can optimize the design and utilization of titanium dioxide nanoparticles for a diverse multitude of applications, from sunscreen formulations to pigments in coatings to construction materials. This newfound understanding of nanoparticles allows an opportunity for manufacturers of these applications to produce optimized titanium dioxides- and metal oxides in general- in a way that they hadn't been aware of before. By knowing the best structure of these nanoparticles, products containing them and available to human consumers would be stronger, tougher, and more resilient.

Moving forward, further research into increasing the mechanical strength of titanium dioxide nanoparticles. Though metal oxides aren't the strongest compound, researching how they could be could unlock even more possible uses in diverse industries, later leading to innovations that benefit society. Additionally, there is room for researching more into other applications. This study focused on five categories, but possibly looking at more would allow for more trends across the varying data to appear. At the same time, looking at different properties of titanium dioxide and not exclusively structural would allow for a better understanding of the nanoparticle's desirability. Another way to research further into the topic is to look at another metal oxide, Zinc oxide perhaps since that is also used in sunscreen applications, and see how the structural properties compare which may cause certain characteristics to become more prevalent in one over the other. There is a great amount of room for growth in this topic and each step in that direction of curiosity would allow for a more cohesive understanding of the various properties of metal oxides.

Works Cited

ACS Publications: Chemistry Journals, Books, and References Published ..., pubs.acs.org/doi/pdf/10.1021/cr500422r. Accessed 30 Apr. 2024.

Al Mutairi MA;BinSaeedan NM;Alnabati KK;Alotaibi A;Al-Mayouf AM;Ali R;Alowaifeer AM. (n.d.). Characterisation of Engineered Titanium Dioxide Nanoparticles in Selected Food. Food Additives & Contaminants. Part B, Surveillance. Retrieved from pubmed.ncbi.nlm.nih.gov/37255019/

Alhalili, Z. (2023). Metal Oxides Nanoparticles: General Structural Description, Chemical, Physical, and Biological Synthesis Methods, Role in Pesticides and Heavy Metal Removal through Wastewater Treatment. Molecules (Basel, Switzerland), 28(7), 3086. doi:10.3390/molecules28073086

Ayman Hijazi, et al. (2018, May 2). Solar Pyrolysis of Waste Rubber Tires Using Photoactive Catalysts. Waste Management. Retrieved from www.sciencedirect.com/science/article/abs/pii/S0956053X1830268X?via%3Dihub

B. Mekuye, B. Abera. (2023). Nano Select, 4, 486. https://doi.org/10.1002/nano.202300038

Ceramic Composition and Properties. (n.d.). Encyclopædia Britannica. Retrieved from www.britannica.com/technology/ceramic-composition-and-properties

Chen, J., et al. (2009). Photocatalytic Activity of Titanium Dioxide Modified Concrete Materials – Influence of Utilizing Recycled Glass Cullets as Aggregates. Journal of Environmental Management, 120(1), 24–31. https://doi.org/10.1016/j.jenvman.2009.02.015

Chin, H. S., et al. (2010, November 15). Review on oxides of antimony nanoparticles: synthesis, properties, and applications. Journal of Materials Science, 45(22), 5993+. Gale In Context: Science. Retrieved from link.gale.com/apps/doc/A365455224/SCIC?u=j043905001&sid=bookmark-SCIC&xid=14f989bd

Cozart, C. R. (2010, April 30). Lack of Significant Dermal Penetration of Titanium Dioxide from Sunscreen Formulations Containing Nano- and Submicron-Size TIO2 Particles. ScienceOpen. Retrieved from www.scienceopen.com/document?vid=a7d33668-3830-4adc-b17d-7b7e0b20c4d1

Crystal. (2019, August 9). Britannica School, Encyclopædia Britannica. Retrieved from school.eb.com/levels/high/article/crystal/110297

Diniz RR;Paiva JP;Aquino RM;Gonçalves TCW;Leitão AC;Santos BAMC;Pinto AV;Leandro KC;de Pádula M. (n.d.). Saccharomyces Cerevisiae Strains as Bioindicators for Titanium Dioxide Sunscreen Photoprotective and Photomutagenic Assessment. Journal of Photochemistry and Photobiology. B, Biology. Retrieved from pubmed.ncbi.nlm.nih.gov/31434036/

Dr. Ananya Mandal, MD. (2023, July 20). Morphology of Nanoparticles. News. Retrieved from www.news-medical.net/life-sciences/Morphology-of-Nanoparticles.aspx#:~:text=High%20aspect%20ratio%20nanoparticles%20include,prism%2C%20helical%2C%20or%20pillar

Encyclopædia Britannica. (2011). Oxide. Britannica School. Retrieved from school.eb.com/levels/high/article/oxide/57833

Harrison, D. J. (n.d.). Titanium Dioxide Nanoparticle Integrated Concrete: An Assessment of Nanoparticle Release When Exposed to UV Radiation and Wet Weather Conditions. NTIS. Retrieved from ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/AD1036718.xhtml

He, R., et al. (2019, July 7). Preparation and Evaluation of Exhaust-Purifying Cement Concrete Employing Titanium Dioxide. Materials (Basel, Switzerland). Retrieved from www.ncbi.nlm.nih.gov/pmc/articles/PMC6650934/

Ionic Bond. (n.d.). Encyclopædia Britannica. Retrieved from www.britannica.com/science/ionic-bond

Jianrong Song, et al. (2014, July 16). The Effects of Particle Size Distribution on the Optical Properties of Titanium Dioxide Rutile Pigments and Their Applications in Cool Non-White Coatings. Solar Energy Materials and Solar Cells. Retrieved from www.sciencedirect.com/science/article/abs/pii/S0927024814003444?via%3Dihub

Koley, A. (2018, October 24). Doping in Semiconductors - Article. ATG. Retrieved from https://www.atg.world/view-article/3739/doping-in-semiconductorsp0

Low, J., & Jiang, C. (2020). Photocatalysts. In Photocatalysts - an Overview. Retrieved from www.sciencedirect.com/topics/materials-science/photocatalysts

Lu, P. J., Cheng, W. L., Huang, S. C., Chen, Y. P., Chou, H. K., & Cheng, H. F. (2015). Characterizing Titanium Dioxide and Zinc Oxide Nanoparticles in Sunscreen Spray. International Journal of Cosmetic Science, 37(1), 70–75. https://doi.org/10.1111/ics.12218

Lu, H., et al. (2018). Enhanced Diffuse Reflectance and Microstructure Properties of Hybrid Titanium Dioxide Nanocomposite Coating. ScienceOpen. https://doi.org/10.14293/S2199-1006.1.SOR-CHEM.A47B43C.V1

Lu, Y., & Lu, Z. (2024). Synthesis, characterization and thermal behavior of plasticized poly (vinyl chloride) doped with folic acid-modified titanium dioxide. Scientific Reports. https://pubmed.ncbi.nlm.nih.gov/35233000/

Morphology of Nanostructured Materials. (n.d.). Retrieved from www.saha.ac.in/surf/satyajit.hazra/papers/pdf/PAC02_MNM.pdf

Mostafa, F. E.-Z. M., et al. (2023). Analyzing the Effects of Nano-Titanium Dioxide and Nano-Zinc Oxide Nanoparticles on the Mechanical and Durability Properties of Self-Cleaning Concrete. Materials (Basel, Switzerland), 16(22), 4538. https://doi.org/10.3390/ma16224538

Nanotechnology Weekly. (2020). Recent Findings from LUT University Highlight Research in Nanoparticles (Structure of Manganese Oxide Nanoparticles Extracted via Pair Distribution Functions). Nanotechnology Weekly, 13(4), 1045. Retrieved from link.gale.com/apps/doc/A620215966/GPS?u=j043905001&sid=bookmark-GPS&xid=9ed32b04

Naseem, T., & Durrani, T. (2021). The Role of Some Important Metal Oxide Nanoparticles for Wastewater and Antibacterial Applications: A Review. Environmental Chemistry and Ecotoxicology, 3, 59–75. https://doi.org/10.1016/j.enceco.2020.12.001

National Nanotechnology Initiative. (n.d.). What is so special about "nano"? Retrieved April 29, 2024, from https://www.nano.gov/about-nanotechnology/what-is-so-special-about-nano#:~:text=Nanoscale%20materials%20have%20far%20larger,have%20phenomenally%20high%20surface%20areas.

Nineta Majcen, et al. (1998). Linear and Non-Linear Multivariate Analysis in the Quality Control of Industrial Titanium Dioxide White Pigment. Analytica Chimica Acta, 366(1), 81–90. https://doi.org/10.1016/S0003-2670(97)00137-2

Nunes, D., et al. (2018). Metal Oxide Nanostructures: Synthesis, Properties and Applications. Netherlands: Elsevier Science.

Ou, C., & Wang, D. (2021). Structural Performance Characteristics of Nanomaterials and Its Application in Traditional Architectural Cultural Design and Landscape Planning. Advances in Civil Engineering, 1–15. https://doi.org/10.1155/2021/5531679

Pacific Northwest National Laboratory. (2021). Novel Titanium Dioxide Structures Enhance Photoactivity. PNNL. Retrieved from www.pnnl.gov/publications/novel-titanium-dioxide-structures-enhance-photoactivity

Powell, B. J. (2000). Determination of Titanium Dioxide in Foods Using Inductively Coupled Plasma Optical Emission Spectrometry. ScienceOpen. Retrieved from www.scienceopen.com/document?vid=105436fd-db8f-4c4a-800f-369430b00754

Prasanna, S. (n.d.). Http://Www.Sciencedirect.Com/Science/Article/Pii/S0016508516300543 ... Chapter 4 - Metal Oxide Based Nanomaterials and Their Polymer Nanocomposites. SNFGE. Retrieved from www.snfge.org/content/httpwwwsciencedirectcomsciencearticlepiis0016508516300543

Praveen, P. (2019, January). Structural, Functional and Optical Characters of TiO2 Nanocrystallites: Anatase and Rutile Phases. Retrieved from https://www.researchgate.net/publication/13311548_Electronic_and_optical_properties_of_three_phases_of_titanium_dioxide_Rutile_anatase_and_brookite

Samontha, A., Chaleawlert-Umpon, S., Santaladchaiyakit, Y., Pon-On, W., Vongpromek, R., & Grudpan, K. (2011). Particle Size Characterization of Titanium Dioxide in Sunscreen Products Using Sedimentation Field-Flow Fractionation-Inductively Coupled Plasma-Mass Spectrometry. ScienceOpen. https://www.scienceopen.com/document?vid=949a0103-39e2-423a-b997-97fddc1b6851

Shi, H., Magaye, R., Castranova, V., & Zhao, J. (2013). Titanium dioxide nanoparticles: A review of current toxicological data. Particle and Fibre Toxicology, 10, 15. https://doi.org/10.1186/1743-8977-10-15

Society of Automotive Engineers. (2017). Quantification of Reduction of Nitrogen Oxides by Nitrate .... SAE International. Retrieved from journals.sagepub.com/doi/10.3141/2290-19

University of Copenhagen Reports Findings in Nanoparticles. (2021, December 13). Nanotechnology Weekly, 5218. Retrieved from Gale Academic OneFile database.

Wang, Y., Li, Y., Zhang, W., & Wang, D. (2014). Photocatalytic degradation and reactor modeling of 17α-ethynylestradiol employing titanium dioxide-incorporated foam concrete. Environmental Science and Pollution Research International. https://pubmed.ncbi.nlm.nih.gov/25242591/

Wellcomeopenresearch.Org. (n.d.). Retrieved April 29, 2024, from https://wellcomeopenresearch.org/articles/6-56/v2

Wu, D., Chen, S., & Feng, P. (2023). Study on the constitutive relationship between ordinary concrete and nano-titanium dioxide-modified concrete at high temperature. Materials (Basel, Switzerland). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10381539/

Zumdahl, S. S. (2011). Oxide. Britannica School. http://school.eb.com/levels/high/article/oxide/57833

Chronic Stress Influence on Autonomic Recovery Speed after Exercising in High School Students by Rachana Perachi Raja

Abstract

Mental stress from daily responsibilities or struggles is common throughout the general public, especially high school students. Over time, this stress develops into harmful chronic stress. The speed of autonomic recovery (AR) after a mental or physical stressor can be used to evaluate the impact this chronic mental stress has on the body during recovery from a stressful period. There was a lack of studies exploring the influences of chronic stress on AR specifically after the physical stressor of exercising. A perceived mental stress evaluation form was utilized to split up participants into higher or lower mental stress groups. Heart Rate Variability (HRV) was used to measure levels of sympathetic and parasympathetic nervous system activity. A baseline HRV was measured before exercise and then HRV was measured at additional resting minutes spanning 20 minutes after exercise. It was evaluated to see how long it took for HRV to return to its baseline levels from before exercise. It was found that AR did not fully occur within the given rest time. According to additional analysis, the higher stress group recovered around 63% and the lower stress group recovered around 68% within the given rest time. No statistical differences were found between these values. The results suggested that the amount of mental stress, specifically chronic stress, did not play a role in the speed of AR after exercising.

Introduction

A year ago, the American Stress Institute found that around 83% of US workers suffer from work-related stress (“WORKPLACE STRESS - the American Institute of Stress”). Additionally, 76% said they had experienced health impacts due to stress in the prior month, including headaches, fatigue, and changing sleep habits (“Stress in America 2022”). Feeling stress is prevalent among the US population, yet within that population, there is still a significant amount of people experiencing the harmful side effects of it. These damaging side effects underlie the physiology of the human body’s nervous system. The involuntary nervous system has two essential parts: the parasympathetic nervous system (PNS) and the sympathetic nervous system (SNS). The PNS is responsible for the “rest and digest" part of the body. It is activated during times of rest and in the absence of threat. The SNS, however, is activated when the body is active or stressed (Thayer et al. 748). Both the PNS and SNS play a role in stressors. A stressor can be anything that activates the SNS. The body is more physically or mentally stressed during a stressor, depending on whether the stressor is physical or mental. Therefore, the body switches from its rest phase to its active/stress phase, which is the switch from the PNS to the SNS (Michael et al. 2). Then, once the stressor is over, the body is no longer under that stress, causing the PNS to be reactivated again (Jia et al. 1-2). Switching from the SNS to PNS after a stressor is a vital function called autonomic recovery. Autonomic recovery protects the heart by preventing strain even when not under stress, preventing wearing over time (Jia et al. 1-2). The speed of this recovery is crucial; according to Jia et al., (2016) a faster autonomic recovery indicates more cardiovascular health, while a slower recovery indicates increased cardiovascular risk (Jia et al. 2). In the current literature, studies mostly explore how mental stress-related disorders can affect autonomic recovery after a mental or physical stressor; however, more research is required on the effect general perceived chronic stress has on autonomic recovery, specifically after exercising. So, this gap generated the research question/goal: How does perceived, chronic stress influence the speed of autonomic recovery after aerobic exercising in high school students?

Literature Review

As previously mentioned, there are two parts of the involuntary nervous system: the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). The PNS is the part of the nervous system that is activated whenever the body is at rest and not in danger. The activation is characterized by decreased blood pressure, increased blood circulation to digestive organs, decreased heart rate, and increased digestion (Noyes & Barber-Westin, 1128). Consequently, the SNS is the part of the nervous system that activates during a fear response or when the body requires more exertion (Thayer et al. 749-750). The activation of the SNS is characterized by an increase in blood pressure, increased blood circulation to skeletal muscle, increased heart rate, and decreased digestion (“Sympathetic Nervous System”).

While the SNS is important during stressful and fearful situations, it has been proven by numerous studies that increased chronic stress/SNS activation and less PNS activity can lead to “cardiovascular complications such as hypertension, increased risk of angina, myocardial infarction, ventricular arrhythmia, and acute heart failure” (Daniela et al. 5). A standard indicator of chronic SNS activity is having generally lower heart rate variability values. Heart rate variability (HRV) measures the variability in time between R-R intervals on an ECG (Thayer et al. 748). R-R intervals are defined as the space between the peaks of ventricular depolarization measured in an ECG. An ECG represents an Electrocardiogram, which is an electrical measure of the heart activity or the time between heartbeats (Jia et al. 4). HRV measures the extent to which the SNS or PNS is activated. A higher value of HRV would signify more PNS activity due to increased variability between heartbeats, indicating that the body is more flexible in adjusting to quick changes in the body. A lower HRV value would indicate more SNS activity, meaning the body is not as adaptable to changes (Thayer et al. 748).

A prominent circumstance where there can be increased SNS activity is during a job, where there is more work/job-related stress. Increased SNS activity due to a job can be seen in Borchini et al.’s (2018) study, which examined the effect prolonged stress in healthy nurses had on their HRV levels during the workday. The study split the nurses into groups of “prolonged high stress (PHS)” and "stable low stress (SLS)” (Borchini et al. 2). The results demonstrated that “the lowest HF values were in the “PHS (geometric mean: 76.3 ms^2) and the highest in the SLS group (139.1 ms^2)” (Borchini et al. 5). HF stands for high frequency. HF is a way to categorize HRV in ms^2 where a high value indicates higher HRV values and vice versa (Borchini et al. 1). Nurses who felt less stress overall (the SLS group) had higher HRV values, indicating their bodies had more PNS activity, which was more protective of the heart. On the contrary, nurses with PHS had lower HRV values and, therefore, more SNS activity. Their work-related stress is related to their increased SNS activity, indicating that their body has more cardiovascular risks (Borchini et al. 8-9).

Another way to measure cardiovascular risk is through one's recovery period after a stressor. The body has a mechanism involving the PNS and SNS activations. This mechanism is autonomic recovery (AR) after a stressor. As mentioned before, a stressor can be physical or mental if it invokes the SNS in the body through physical/mental stress. Then, after the stressor, AR occurs, which is the switch from the SNS back to the PNS. AR is essential for maintaining a healthy balance between the SNS and the PNS activity. AR prevents overstraining the body systems when the SNS is activated by switching back to the PNS, where the body systems are more relaxed (Jia et al. 1-2). The faster the AR occurs, the less the body is under strain, ultimately preventing sudden cardiac death (Jia et al. 2). Furthermore, a slower AR speed signifies more cardiovascular risk (Jia et al. 2). Because of the health indication AR speed can provide, there is a significant amount of research surrounding factors that can affect the speed of AR after physical stressors and mental stressors.

Specifically, mental stress-related disorders are explored through various studies to see their effects on AR speeds. Majewski et al.'s (2023) study, for example, compared the AR speeds of those with PTSD and those without it. PTSD (posttraumatic stress disorder) is characterized by high amounts of stress (Majewski et al. 1). Their study included 27 PTSD patients and 15 controls. A mental stressor test triggered the SNS activity, and AR was measured after the stressor. To do this, the researchers measured HRV through RMSSD (Root mean square of successive differences), which indicates the time between normal heartbeats: A way to measure HRV, a higher value indicating more PNS activity and a lower value indicating more SNS activity (Thayer et al. 748). The researchers found that the RMSSD values of those with PTSD were, in general, lower than the controls. Additionally, the speed of AR in those with PTSD was slower than in those without PTSD. This slower AR conveys an adverse effect of chronic stress associated with PTSD on the body (Majewski et al. 5). The lower HRV values found in the PTSD group agree with Borchini et al.'s (2018) study that more stress is associated with lower HRV.

Another study by Liu et al. (2021), similar to Majewski et al.'s (2023) study, examined schizophrenia's effects on AR speeds using a mental stressor test. The study had 45 schizophrenia patients and 45 controls. The mental stressor was performing arithmetic. Like Majewski et al.'s (2023) study, the researchers examined the HRV values of the participants after the mental stressor. The results of the study depicted that the schizophrenia patients' AR did not occur within the given time, while the controls' AR did. Liu et al. suggested that the significantly increased SNS activity and lower HRV values found in schizophrenia patients might have been the mechanism behind their conclusions (Liu et al. 8-9).

While both Majewski et al.'s (2023) and Liu et al.'s (2021) studies look into mental stress-related disorders' effects on AR, they only focus on AR after mental stressors specifically. When examining the impact of mental stress-related illnesses on AR after physical stressors, there are significantly fewer studies, but not completely zero. One study, for example, assessed scores on the Hospital Anxiety Depression Scale (HADS) and their effects on AR speeds after exercise. HADS is a method of identifying “probable cases of psychosomatic illness” (Santana et al. 295). The physical stressor used in this study was aerobic exercise on a treadmill. Like Majewski et al.’s study, the researchers allowed the participants to rest for 30 minutes after exercise to measure AR. The evidence indicated that those with increased scores on the HADS had a slower AR than those with lower scores. (Santana et al. 301).

Research Gap

The studies presented have examined mental stress-related illnesses on both mental and physical stressors. However, there has been a lack of studies researching the effect general chronic stress, independent of mental disorders, has on AR after a physical stressor. Borchini et al.’s (2018) study established that increased general work-related stress correlates with more SNS activity and lower HRV values. The studies that researched AR of those with mental stress-related illnesses, such as Majewski et al.’s (2023) and Liu et al.’s (2021) studies, found that the mental stress-related illnesses groups had a higher SNS activity similar to Borchini et al.’s study. They also found that those same groups had a slower AR after the mental stressor and associated the slower AR with increased SNS activity. In Santana et al.’s (2020) study, however, AR after a physical stressor was measured. The results showed the higher amounts of stress groups had slower AR than the lower stress group, similar to Majewski et al.’s (2023) and Liu et al.’s (2021) studies. Santana et al.’s (2020) study also had higher SNS activity for the higher mental stress group. Since SNS activity has been associated with slower AR in those with mental stress-related illnesses, it might also play a role in those with more general, chronic stress from responsibilities, as described in Borchini et al.’s (2018) study. Mental stress’s influences on AR speeds specifically after physical exercise are already largely underexplored, so specifically examining AR after physical exercise could provide more insight into the field of mental stress and physical exercise. Additionally, exploring a younger population instead of the older populations that the previous studies researched could provide more insight into the early-life effects of chronic stress. All these gaps in the research ultimately led to the research question: How does perceived, chronic stress influence the speed of autonomic recovery after aerobic exercising in high school students?

Studying the general chronic stress that many people experience daily can be beneficial in evaluating whether this stress threatens cardiovascular health through AR in the general population, especially the younger population. Knowing the results of this study could provide insight into the nuanced effects of stress and possible ways to mitigate these effects.

Hypothesis

It is hypothesized that those with higher levels of perceived stress will have a slower AR than those with lower perceived stress. Thayer et al. (2012) concluded that higher stress levels cause increased SNS activity. Additionally, Majewski et al.'s (2023) and Liu et al.'s (2021) studies found that patients with mental illnesses had increased SNS activity and slower AR. Therefore, an increased SNS due to perceived stress should correlate with a slower AR. The null hypothesis would be that amounts of perceived stress levels do not affect the speed of AR after aerobic exercising in high school students.

Methods

The research topic explores whether mental stress affects the speed of autonomic recovery (AR) after exercising. As mentioned, AR is the speed at which the body switches from the SNS to the PNS after exercise.

A quasi-experimental study was conducted to answer the research question, where the independent variable was the perceived level of mental stress the participants had measured by a perceived stress questionnaire. The dependent variable was the autonomic recovery speed for each participant, which was measured using heart rate variability (HRV). HRV was used to measure this speed of AR because it is an established indicator of the SNS and PNS activity in the body (Thayer et al. 748). HRV was measured by heart rate in the form of an ECG so that the variability in the time between heartbeats could be quantified in milliseconds. The time between heartbeats was used in the RMSSD formula to calculate HRV values.

$$\text{RMSSD} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}((RR)_{j+1} - (\overline{RR}_j))^2}$$

RMSSD is a reliable formula used by many studies to calculate HRV to show the activity levels of the PNS and SNS in the body, where lower values show SNS activity and higher levels show PNS activity (Thamm et al. 3).

Overall, the AR speeds of those with higher mental stress levels were compared with those with lower mental stress levels to see if there were any significant differences between them.

For this study, the participants' ages ranged from 15-18 years old, and they specifically attended my high school as students. This pool of participants was chosen because there are differences in the participants' school workload depending on the classes they take, so stress levels would vary. Additionally, access to my high school's cardio room, where the treadmills are, was given.

All participants were required to provide signed consent forms and were told what was expected of them from the study before they signed up and after they were selected.

Procedure

The participants were first sent a 30-item Perceived Stress Questionnaire (PSQ) to evaluate their perceived mental stress levels. In the form, there were 30 statements regarding perceived stress in the last month. The participants rated the statements based on how true they were using a 1-4 Likert scale, with one being the least true (or false) and four being the most true. The PSQ is a questionnaire first developed by Levenstein et al. (1993) and is validated by other studies. The PSQ measures factors like "overload, irritability, lack of joy, fatigue, worries, and tension" (Levenstein et al. 26). School counselors' contacts were given at the beginning of the form to ensure mental health and safety. Each participant's responses for each statement were calculated using the PSQ-specified formula and were the assigned mental stress levels for those participants. A lower and higher mental stress group was created by splitting participants by the median number to have adequate comparison groups for the study.

Before the experimental protocol, participants visited the designated cardio room a week earlier to establish their target heart rates and speeds at which they would perform the chosen physical stressor test. Participants were asked to jog 0.25 miles on the treadmills (referred to as the "trial jog"), simulating the physical stressor test for the actual experimental protocol. During the trial jog, they jogged 0.25 miles at grade 0 at a speed they felt matched a Borg scale rating of 13, described as somewhat challenging but manageable (Brockmann & Hunt 4; Williams 404). Heart rates were measured using an Apple Watch to confirm whether each target heart rate matched 70%-80% of that age group's estimated maximum heart rate. This entire trial jog session was conducted because participants had varying fitness levels, which can influence the speed of AR and serve as a confounding variable (Seiler et al. 1371). To minimize this confounding variable, the differing fitness levels of the participants were determined through target heart rates and speed on the treadmills.

A week after evaluating the target heart rates and speeds, participants returned to the cardio room for the actual experimental protocol. Only two participants could be tested daily because only two Apple Watches were available. The experimental protocol occurred between 10:30 and 11:00 a.m. every day to avoid circadian influences. Participants were informed that they must not eat at least an hour before arriving to avoid feeling any discomfort during exercise and to not consume any caffeinated beverages at least 10 hours before arriving to ensure reliable results.

Before they started jogging, I had participants measure their HRV using an Apple Watch. Apple watches were used because they were non-invasive, accessible, and did not require any payment. Additionally, Apple watches were a validated method of measuring HRV during rest (Hernando et al. 9). With the Apple watches on, participants were asked to open the "ECG+ | Analyzer for QTc & HRV" app on the Apple Watch and place their finger on the button for 30 seconds as the app instructs them to do ("ECG+ | Analyzer for QTc & HRV"). The Apple Watch provided me with an ECG and HRV value. This HRV value would be the baseline HRV for that participant, aka their resting HRV.

Then, the participants jogged 0.25 miles on the treadmill at grade 0 and had their Apple watches on, constantly measuring their heart rates. Participants were asked to monitor their heart rate and keep it similar to their target heart rate by changing their treadmill speeds accordingly. Jogging 0.25 miles was chosen because it is enough for the body to produce changes in the PNS and SNS activation since the body's SNS activates even with slight exercise (Jia et al. 2). Additionally, jogging is a standard method used by many studies examining AR speeds after a physical stressor (e.g., Santana et al. 299; Oliveira et al. 2-3).

Then, immediately one minute after participants were done jogging, they retook their HRVs. After that, the participants rested for around 20 minutes, at which point they measured their HRVs six more times in intervals. Along with minute 1, additional HRV measurements were taken at resting minutes 3, 5, 7, 10, 15, and 20 (M3, M5, M7, M10, M15, and M20, respectively). A diagram of the exercise protocol is given below in Figure 1.

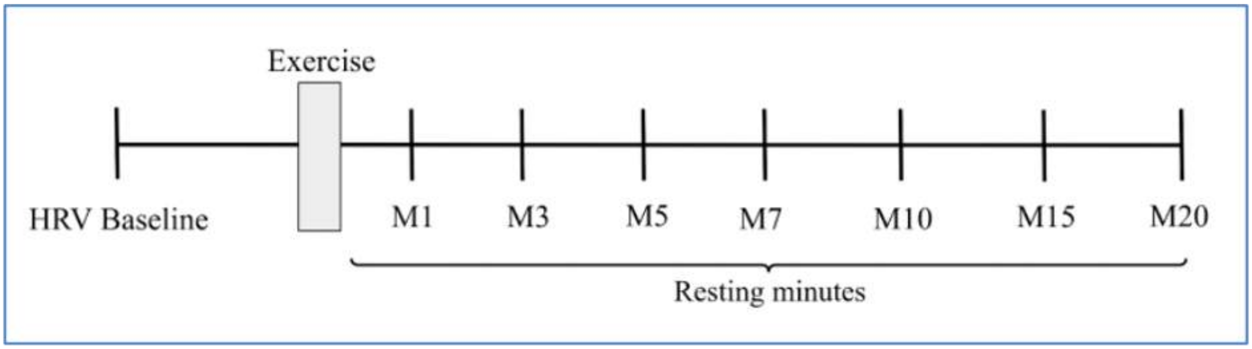


Figure 1. Minutes at which ECGs were taken. These measurements include the HRV baseline from before exercise and the following resting minutes M1, M3, M5, M7, M10, M15, and M20 after exercise.

Participants were then debriefed and were allowed to leave.

Analysis Methods

These intervals helped determine the speed of AR by examining at which interval HRV returned to its baseline HRV levels. This method was validated by Santana et al. (2020). To see whether the HRV values returned to baseline levels, I examined if there was a non-statistical difference between the HRV level for each resting minute and the baseline HRV, as well as non-statistical differences in the consecutive resting minutes. A paired t-test with an alpha value of 0.05 was used to determine significant differences. The method of increasing intervals and evaluating AR speeds was validated by studies by Santana et al. (2020) and Oliveira et al. (2020), which also examined AR speeds.

In addition, I used the percentage of the HRV from M20 out of the baseline HRV to find out which group recovered faster in the given rest time. This method of assessing the speed of AR is validated by Seiler et al.'s (2007) study, which measured AR speeds in athletes.

Lastly, a linear regression model was made of only the resting minute HRV values from after exercise to observe a trend in AR speeds between the higher and lower mental stress groups. The linear regression models were calculated using Google Sheets. Using linear regression models to evaluate AR speeds was validated by Santana et al. (2020), who also compared ARs between those with increased and decreased stress.

Overall, I compared the AR speeds and percentage of recovery of those with higher mental stress with those with lower mental stress. All t-tests (paired and unpaired) were performed by the Social Science Statistics website (“T-Test Calculator for 2 Dependent Means”; “T-Test Calculator for 2 Independent Means”)

Delimitations

Participants who smoked or vaped were excluded. Participants were also excluded if they had any conditions that could put them at risk when jogging.

Apple Watches were used instead of a more accurate device to measure HRV because they were non-invasive and accessible. While Apple Watches were validated as accurate when measuring HRV at rest, they have yet to be validated for measuring HRV during physical activity or when the heart rate is higher. The mechanism Apple watches use to measure HRV is photoplethysmography, which uses light to detect changes in blood. Photoplethysmography “is very sensitive to motion artifacts, which may lead to poor HRV estimation if false peaks are detected” (Alqaraawi et al. 136). So, limited resources and costs restricted the accuracy of this study.

Additionally, only 20 minutes of rest was given after exercise. Initially, 30 minutes of rest time was going to be given, like the model study by Santana et al. (2020); however, due to unforeseen circumstances and limitations in time, only 20 minutes of rest was given.

IRB Approval

The local IRB approved all methods and procedures for ethicality and feasibility. However, this IRB did not have HHS approval.

Results

After the study, there were 17 participants (n=17), all were females ages 15-18. Eight participants were in the high mental stress group (HMS), and nine were in the lower mental stress group (LMS). Each participant's baseline HRVs were recorded along with their resting-minute HRVs in an Excel spreadsheet. Each participant had one baseline HRV value, M1, M3, M5, M7, M10, M15, and M20.

Figure 2 shows the general AR trend in the HMS group. It uses an Excel graph to average the HRVs from each time interval.

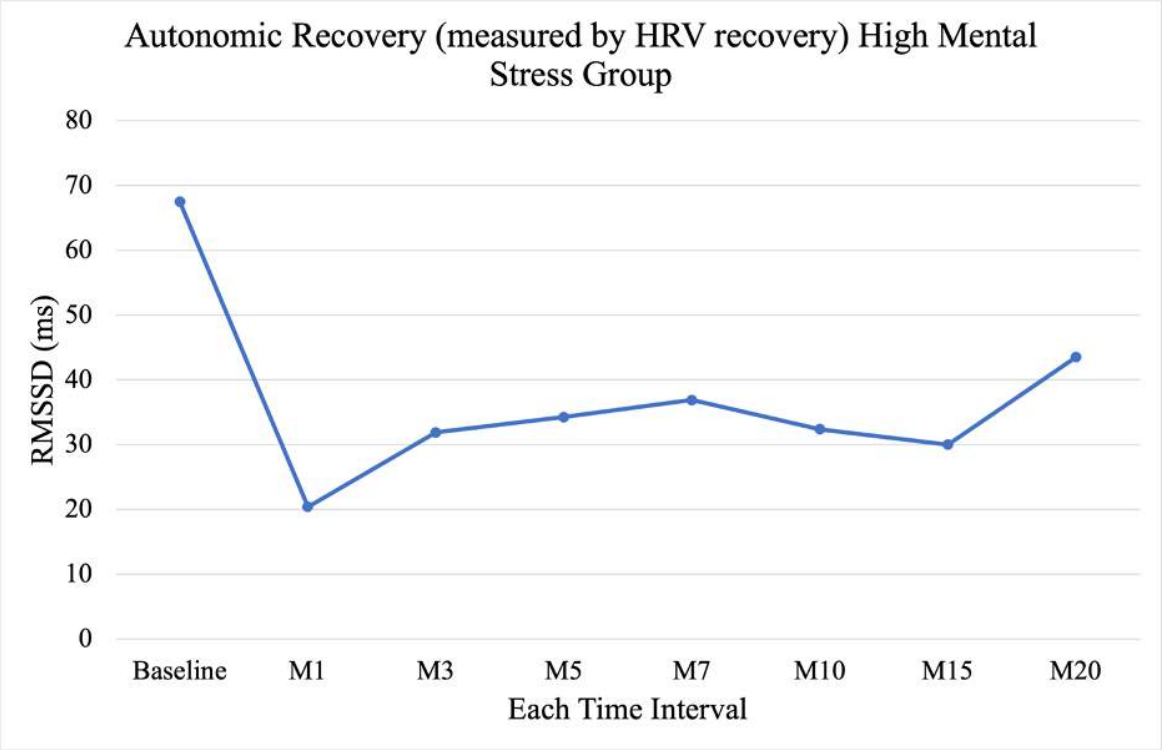


Figure 2. Autonomic recovery in HMS. The graph contains all the intervals at which HRV was taken, including the baseline HRV from before exercise, and the resting minutes after exercise. HRV values were measured through RMSSD in milliseconds.

After performing the paired t-tests, statistically significant differences were found between each resting minute and baseline HRV, as seen in Table 1. This indicates that AR did not fully complete within the given resting period.

Table 1. P-values of each resting minute compared to baseline HRV in the HMS group. There were no non-statistical differences between the baseline HRV and the resting minutes, suggesting that autonomic recovery did not occur.

Each Resting minute compared to the Baseline HRV	**p-value**	**Statistically or Non-statistically Different based on alpha value p<0.05**
M1	p= 0.00497	Statistically Different
M3	p= 0.01259	Statistically Different
M5	p= 0.00086	Statistically Different
M7	p= 0.00438	Statistically Different
M10	p= 0.00216	Statistically Different
M15	p= 0.00087	Statistically Different
M20	p= 0.02377	Statistically Different

Then, by averaging the LMS group's HRVs from each time interval, the general AR trend in the LMS group can be seen below in Figure 3. An Excel graph was used.

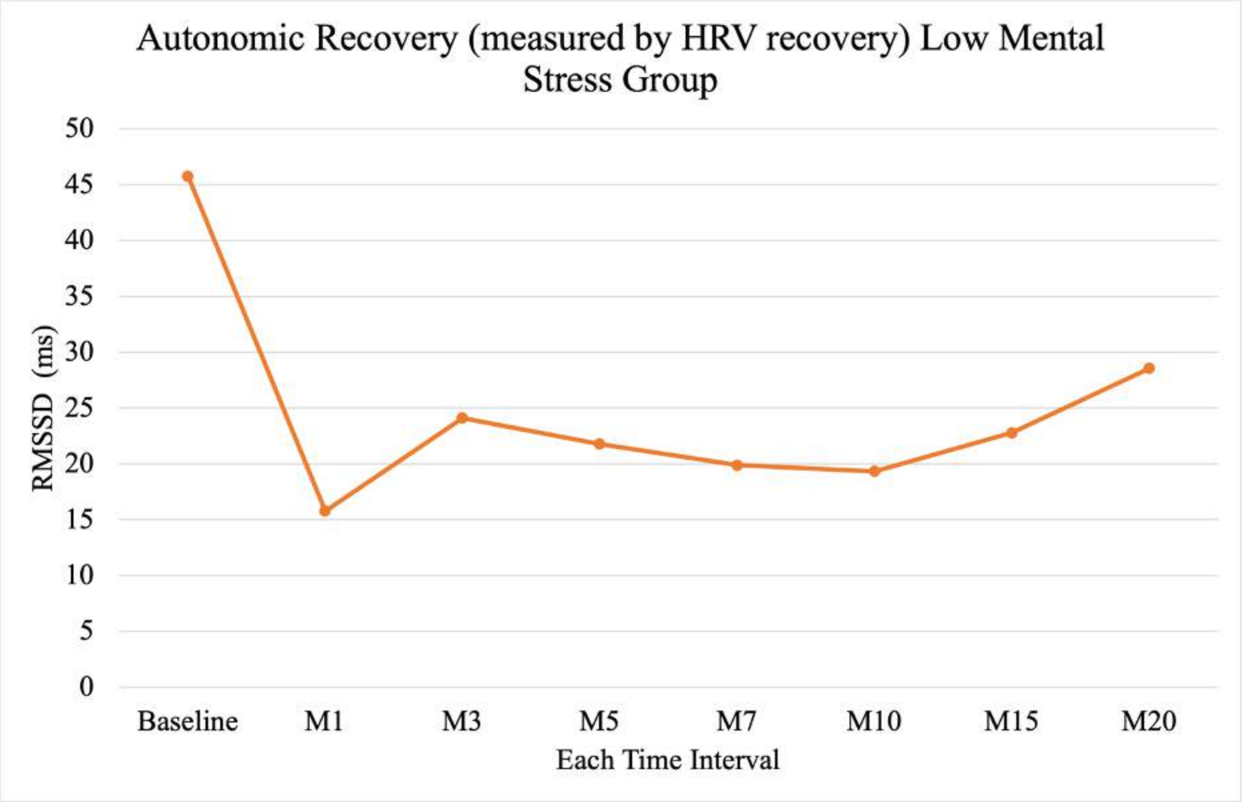


Figure 3. Autonomic recovery in LMS. The graph contains all the intervals at which HRV was taken, including the baseline HRV from before exercise, and the resting minutes after exercise. HRV values were measured through RMSSD in milliseconds.

When the LMS group was analyzed using paired t-tests, statistically significant differences were found between each resting minute and baseline HRV except in M20, as seen in Table 2.

Table 2. P-values of each resting minute compared to baseline HRV in the LMS group. There was only one non-statistical difference between the baseline HRV and M20.

Each Resting minute compared to the Baseline HRV	p-value	Statistically or Non-statistically Different based on alpha value p<0.05
M1	p= 0.0013	Statistically Different
M3	p= 0.00588	Statistically Different
M5	p= 0.00157	Statistically Different
M7	p= 0.00561	Statistically Different
M10	p= 0.00255	Statistically Different
M15	p= 0.00558	Statistically Different
M20	p= 0.05179	Non-Statistically Different

This suggests that AR might have occurred within the given time; however, it cannot be guaranteed because there was no non-statistical difference between resting minutes after M20 and baseline HRV, primarily because there were no additional resting minutes after M20 to observe whether there were any non-statistical differences. According to Santana et al. (2020), AR only occurs when a resting-minute HRV and consecutive resting minutes have non-statistical differences with the baseline HRV. Therefore, AR cannot be said to have occurred within the 20 minutes of given rest time.

After using the second method of evaluating the speed AR, where the percentage of recovery both groups had achieved within the given time was examined, there were some notable differences. After using the percentage of the HRV at M20 and baseline HRV in both groups, the HMS group recovered 63%, while the LMS group recovered 68%. Descriptively, the LMS did recover more than the HMS did, indicating that the autonomic recovery speed for the LMS was faster. However, when using an unpaired t-test with an alpha value of 0.05, it was found that the percentage recovery from both groups was not statistically different, suggesting that mental stress did not play a role in the speeds of autonomic recovery.

Additionally, a linear regression model involving only the resting minutes was made for both groups to see the overall trend of AR through HRV values. The equation for the trendline is given for both groups in Figure 4.

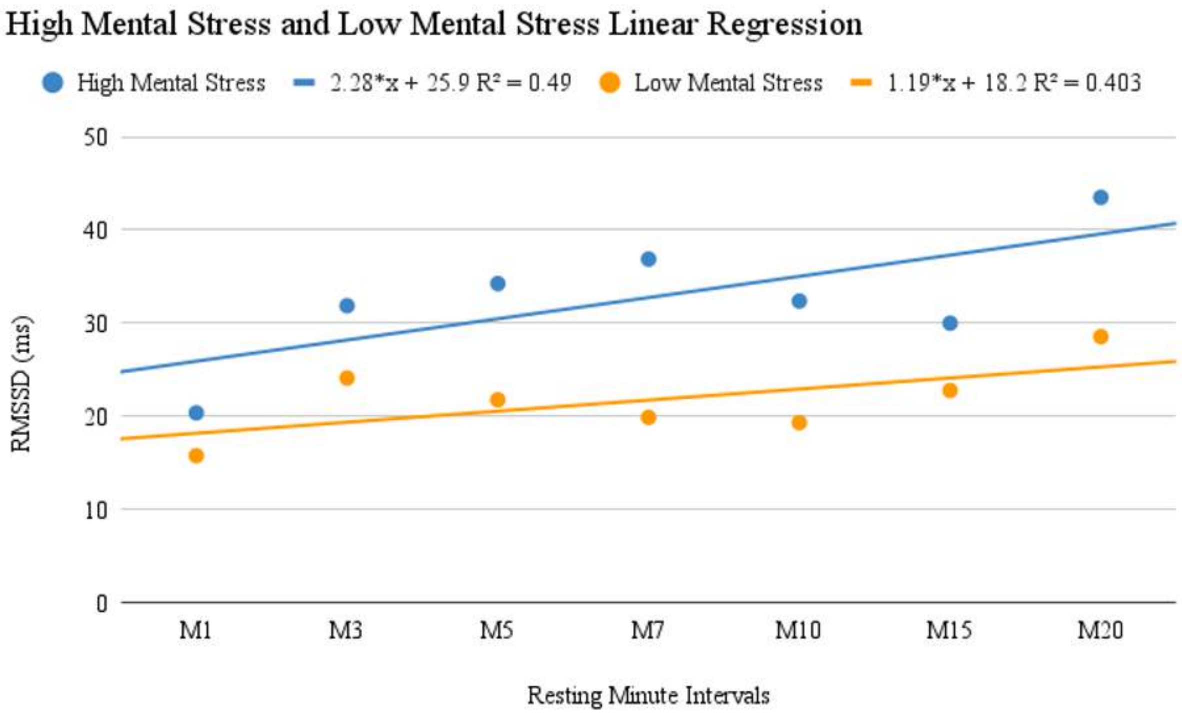


Figure 4. Linear regression model of both LMS and HMS groups.

The slope for the HMS group is 2.28, while the slope for the LMS group is 1.19. Unexpectedly, the slope for the HMS group is greater than the LMS group's slope. These results contradict Santana et al.'s (2020) study, which found that the higher-stress group had a lower slope of HRV than the lower-stress group. However, when doing an unpaired t-test, the calculated p-value was 0.359, indicating a non-statistical difference between both groups' linear regression slopes. The non-statistical difference suggests that mental stress did not affect the speeds of autonomic recovery after aerobic exercise.

By averaging the HRVs from both groups for each time interval, the general trend of HRV before and after exercise between HMS and LMS can be seen in Figure 5. An Excel graph was used.

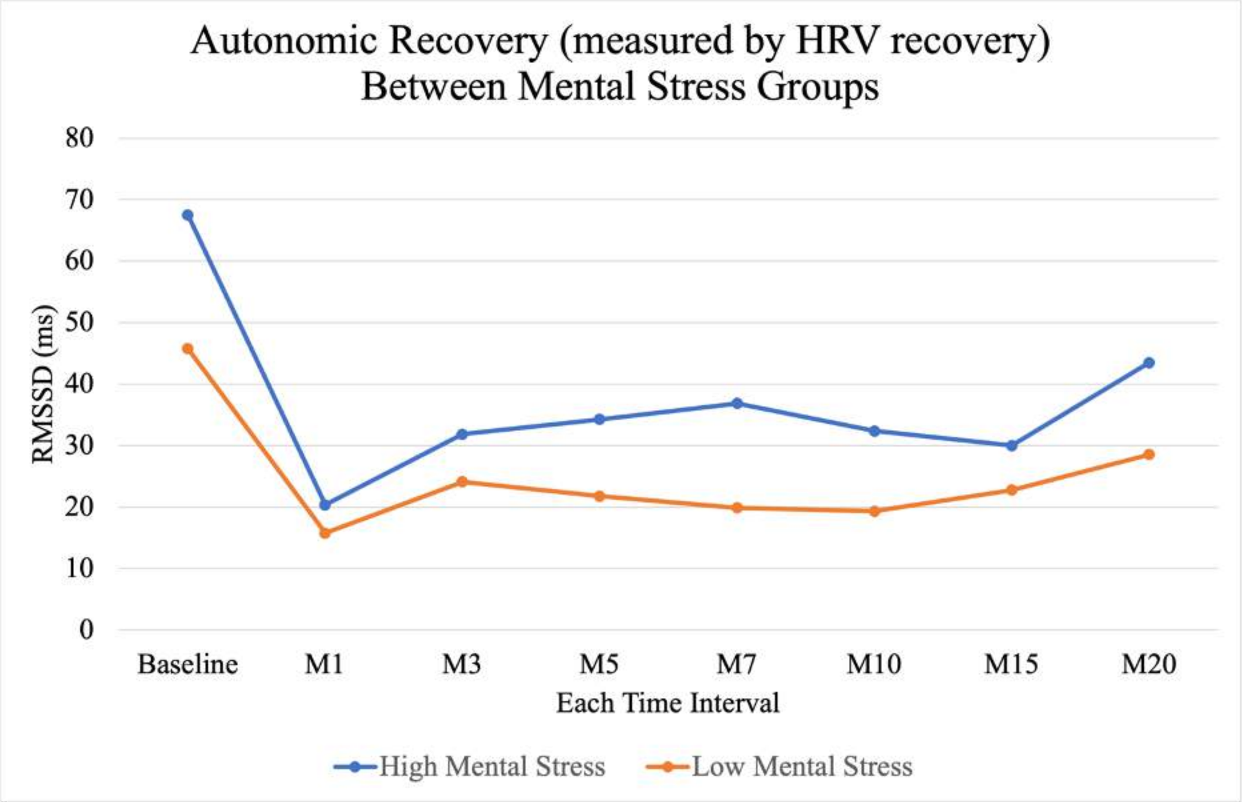


Figure 5. Autonomic recovery in LMS and HMS on the same graph

Both groups' HRV values could be compared to see the overall trend of higher and lower HRV values. According to the current literature, it has been widely established that more chronically stressed individuals have a lower resting/baseline HRV because their SNS has more activity (Thayer et al. 751; Borchini et al. 8-9). However, unexpectedly, in the results, the HMS had an average baseline HRV of 67.5 ms while the LMS had an average baseline HRV of 45.8 ms. Despite the HMS baseline-HRV being higher than the LMS baseline-HRV, they both had no statistically significant difference in their baseline ($p=0.05001>0.05$), suggesting the difference might have been due to chance alone. These results contradicted Thayer et al. (2012) and Borchini et al. (2018) studies that observed lower HRV values in those with more stress. Significance was determined by the unpaired t-test with an alpha value of 0.05.

Additionally, the overall HRVs for the HMS group were descriptively higher than the LMS group. However, when using a paired t-test, both groups had no statistical differences in HRV. P-values are displayed in Table 3.

Table 3. P-values of each resting minute HRV compared between LMS and HMS group

Each LMS and HMS Time Intervals' HRV	p-value	Statistically or Non-statistically Different based on alpha value p<0.05
M1	p= 0.276876	Non-Statistically Different
M3	p= 0.256922	Non-Statistically Different
M5	p= 0.113608	Non-Statistically Different
M7	p= 0.104065	Non-Statistically Different
M10	p= 0.130682	Non-Statistically Different
M15	p= 0.195842	Non-Statistically Different
M20	p= 0.140422	Non-Statistically Different

Discussion

This study aimed to determine whether perceived mental stress affected how fast the body went through AR after exercise in high school students. It was initially hypothesized that those with higher perceived stress would have a slower AR after aerobic exercise. The hypothesis was built off of Liu et al.'s (2021) and Santana et al.'s (2020) studies that found subjects with mental stress-related illnesses had a slower AR after a stressor and associated their results with the increased stress those patients exhibited. In this study mainly, the primary results presented that neither the HMS group nor the LMS group's resting minutes reached a non-significant difference with their respective baseline HRV and also did not have consecutive resting minutes that did the same. These results suggest that AR did not fully occur within the given resting time after exercise. Due to the lack of data in this method of analysis, there was no concrete speed of AR that could be compared between groups. The other two analysis methods were used as the primary evidence for the conclusion. When comparing the percentage of AR both groups achieved, the LMS group recovered a greater percentage of their baseline HRV than the HMS group. However, a non-statistical difference was found between both groups' recovery percentages, indicating that neither groups' AR occurred faster than the other. This implies that perceived mental stress differences did not play a role in the percentage of AR after aerobic exercise in high schoolers. Similar results were found in the linear regression model method of analyzing AR speeds. The HMS group descriptively had a greater slope of 2.28, while the LMS group had a slope of 1.19. Nonetheless, both values were not significantly different, indicating that both groups' rate of recovery was not faster than the other. Both the percentage of AR recovery and linear regression models have no significant differences between data from both groups, suggesting that any of the general differences seen in the HMS and LMS groups were probably due to chance alone. The results from this study accepted the null hypothesis, which states that perceived stress levels do not affect the speed of AR after exercising in high school students. The results that support the null hypothesis reject the initial hypothesis that high schoolers with higher perceived mental stress levels would have a slower AR speed after exercising. Additionally, the null results contradict Liu et al.’s (2021) and Santana et al.’s (2020) studies, which found that AR was slower in subjects with mental stress-related illnesses.

Other Analyzed Data

The already established literature indicates that more stress, especially chronic stress, leads to more prolonged SNS activation, causing generally lower HRV values among individuals (Thayer et al. 751). However, in this study's results, the evidence showed the opposite results, where the higher mental stress group (HMS) had generally higher HRVs than the lower mental stress group (LMS). Although no statistical difference was found between both groups’ baseline HRVs and resting-minute HRVs, these results were unexpected. It is unknown as to why this could have occurred. Future studies could research such unexpected results and what their cause is.

Fulfillment of Research Gap

The original and primary gap in the research was derived from the need for AR-speed-related studies investigating general chronic stress from daily life, independent of mental illnesses. Studies from Santana et al. (2020), Majewski et al. (2023), and Liu et al. (2021) studied AR speeds after a stressor, comparing those with mental illnesses and those without. However, general stress, such as the chronic stress Borchini et al.’s (2018) study examined, had not been previously explored with AR speeds. Additionally, studies measuring mental stress usually applied mental stressors to measure AR; however, a few studies utilize a physical stressor to analyze mental stress’s effects on AR speeds, such as Santana et al.’s (2020) study. In this study, general perceived mental stress was analyzed with a physical stressor of aerobic exercise to examine AR speeds, allowing the primary gap to be explored. The additional specificity to high schoolers was added to examine the effects of stress in a younger population, which other researchers have generally not analyzed as much.

Conclusion

The overall conclusion of this study is that the amount of perceived, general chronic stress did not affect AR speeds after aerobic exercise in high school students. Due to the previously mentioned limitations, such as only having female participants, these findings can not be generalized to the general high school student population.

The implications of this study include the idea that psychological stress can physiologically affect an individual's body in nuanced ways. Mental stress, especially as it becomes more common in modern times, opens the door for increased research on mental stress itself and its chronic, harmful effects on the body.

The results seen between the mental stress groups in this study might have suggested no correlation between perceived mental stress and AR speeds after exercise; however, future research can look into this topic except possibly doing some things differently, such as allowing more resting time after exercise to measure AR or using more reliable devices to measure HRV.

Overall, future research on mental stress and its harmful effects on the body will ultimately allow more understanding of protective physiological processes. More information can also raise awareness of the dangers of chronic mental stress and possibly push for new ways to prevent its development. Mental stress is a significant part of many people's lives, and finding out more about such a common occurrence can help those suffering from it.

Implications

The results of this study attempt to provide a possible way chronic mental stress might affect the body's recovery after a stressor in a high school population. AR is an essential physiological process that protects the heart in the long term. Exploring factors that can inflict harmful effects on AR is significant in studying cardiovascular diseases/disorders. Chronic mental stress has already been determined to be a significant cause of cardiovascular harm. It is especially prominent in the general public, so investigating the physiological effects of stress can provide more insight into general health. (Daniela et al. 5; "Stress in America 2022"). This study also examined a younger population of female high school students. This could provide more insight into the early-life effects of chronic stress, such as cardiovascular risks, in young females and possibly lead to research on ways to prevent them. Although this study led to null results, it could lead to more research into the same topic using better methods and more participants to see if these null results are consistent.

Limitations

Several limitations to this study's conclusions exist. The first is the gender of the participants. Only females participated in the entire procedure, making the data tailored only to them. The null results found in this study may be unique to females and can only be applied to females.

Another limitation was the number of participants in general. Out of a high school containing over 1,000 students, only 17 were participants in this study. Although other studies measuring AR speeds had around 40 participants, such as Majewski et al. (2023) and Santana et al. (2020), more than 17 participants are needed to generalize the conclusions of this study to the general high school population. The lack of participants was due to only having access to my high school's cardio room treadmills for two weeks for the exercise protocol and only possessing 2 Apple Watches.

Another limitation was the inadequate 20 minutes of rest after exercise, leading to incomplete AR. According to Seiler et al. (2007), increased exercise intensity can prolong autonomic recovery. Therefore, the exercise of 0.25 miles and the mild intensity that was chosen were too intense for only a 20-minute rest period. A possible improvement for future studies is to provide more resting time after exercise to allow AR.

The last but most prominent limitation was the use of photoplethysmography through Apple watches rather than a more accurate instrument to measure HRV. As previously stated, using photoplethysmography can be easily altered by outside motion, making it less reliable. While participants remained sitting still when measuring their HRVs, external movements could have still affected these measurements. This limitation reduces the accuracy of this study.

Work Cited

Alqaraawi, Ahmed, et al. "Heart Rate Variability Estimation in Photoplethysmography Signals Using Bayesian Learning Approach." *Healthcare Technology Letters*, vol. 3, no. 2, Institution of Engineering and Technology, June 2016, pp. 136–42, https://doi.org/10.1049/htl.2016.0006. Accessed 10 July 2024.

Borchini, Rossana, et al. "Heart Rate Variability Frequency Domain Alterations among Healthy Nurses Exposed to Prolonged Work Stress." *International Journal of Environmental Research and Public Health/International Journal of Environmental Research and Public Health*, vol. 15, no. 1, Multidisciplinary Digital Publishing Institute, Jan. 2018, pp. 113–13, https://doi.org/10.3390/ijerph15010113. Accessed 10 July 2024.

Brockmann, Lars, and Kenneth J. Hunt. "Heart Rate Variability Changes with Respect to Time and Exercise Intensity during Heart-Rate-Controlled Steady-State Treadmill Running." *Scientific Reports*, vol. 13, no. 1, Nature Portfolio, May 2023, https://doi.org/10.1038/s41598-023-35717-0. Accessed 10 July 2024.

Daniela, Matei, et al. "Effects of Exercise Training on the Autonomic Nervous System with a Focus on Anti-Inflammatory and Antioxidants Effects." *Antioxidants*, vol. 11, no. 2, Multidisciplinary Digital Publishing Institute, Feb. 2022, pp. 350–50, https://doi.org/10.3390/antiox11020350. Accessed 10 July 2024.

"ECG+ | Analyzer for QTc & HRV." *App Store*, 21 May 2021, apps.apple.com/us/app/ecg-analyzer-for-qtc-hrv/id1567047859. Accessed 10 July 2024.

Hernando, David, et al. "Validation of the Apple Watch for Heart Rate Variability Measurements during Relax and Mental Stress in Healthy Subjects." *Sensors*, vol. 18, no. 8, Multidisciplinary Digital Publishing Institute, Aug. 2018, pp. 2619–19, https://doi.org/10.3390/s18082619. Accessed 10 July 2024.

Jia, Tiantian, et al. "Music Attenuated a Decrease in Parasympathetic Nervous System Activity after Exercise." *PloS One*, vol. 11, no. 2, Public Library of Science, Feb. 2016, pp. e0148648–48, https://doi.org/10.1371/journal.pone.0148648. Accessed 10 July 2024.

Levenstein, S., et al. "Development of the Perceived Stress Questionnaire: A New Tool for Psychosomatic Research." *Journal of Psychosomatic Research*, vol. 37, no. 1, Elsevier BV, Jan. 1993, pp. 19–32, https://doi.org/10.1016/0022-3999(93)90120-5. Accessed 10 July 2024.

Liu, Ya, et al. "Altered Heart Rate Variability in Patients with Schizophrenia during an Autonomic Nervous Test." *Frontiers in Psychiatry*, vol. 12, Frontiers Media, Mar. 2021, https://doi.org/10.3389/fpsyt.2021.626991. Accessed 10 July 2024.

Majewski, Kristin, et al. "Acute Stress Responses of Autonomous Nervous System, HPA Axis, and Inflammatory System in Posttraumatic Stress Disorder." *Translational Psychiatry*, vol. 13, no. 1, Springer Nature, Feb. 2023, https://doi.org/10.1038/s41398-023-02331-7. Accessed 10 July 2024.

Michael, Scott, et al. "Cardiac Autonomic Responses during Exercise and Post-Exercise Recovery Using Heart Rate Variability and Systolic Time Intervals—a Review."

Frontiers in Physiology, vol. 8, Frontiers Media, May 2017, https://doi.org/10.3389/fphys.2017.00301. Accessed 10 July 2024.

Noyes, Frank R., and Sue D. Barber-Westin. "Diagnosis and Treatment of Complex Regional Pain Syndrome." *Elsevier EBooks*, Elsevier BV, Jan. 2017, pp. 1122–60, https://doi.org/10.1016/b978-0-323-32903-3.00040-8. Accessed 10 July 2024.

Oliveira, Leticia, et al. "Lower Systolic Blood Pressure in Normotensive Subjects Is Related to Better Autonomic Recovery Following Exercise." *Scientific Reports*, vol. 10, no. 1, Nature Portfolio, Jan. 2020, https://doi.org/10.1038/s41598-020-58031-5. Accessed 10 July 2024.

Santana, et al. "Association between Hospital Anxiety Depression Scale and Autonomic Recovery Following Exercise." *Journal of Clinical Psychology in Medical Settings*, vol. 27, no. 2, Springer Science+Business Media, Nov. 2019, pp. 295–304, https://doi.org/10.1007/s10880-019-09683-7. Accessed 10 July 2024.

Seiler, Stephen, et al. "Autonomic Recovery after Exercise in Trained Athletes." *Medicine and Science in Sports and Exercise*, vol. 39, no. 8, Lippincott Williams & Wilkins, Aug. 2007, pp. 1366–73, https://doi.org/10.1249/mss.0b013e318060f17d. Accessed 10 July 2024.

"Stress in America 2022." *American Psychological Association*, 2022, www.apa.org/news/press/releases/stress/2022/concerned-future-inflation. Accessed 10 July 2024.

"Sympathetic Nervous System." *Physiopedia*, 2019, www.physio-pedia.com/Sympathetic_Nervous_System#:~:text=eg%2C%20the%20sympathetic%20nervous%20system,sweating%20and%20raise%20blood%20pressure. Accessed 10 July 2024.

Thamm, Antonia, et al. "Can Heart Rate Variability Determine Recovery Following Distinct Strength Loadings? A Randomized Cross-over Trial." *International Journal of Environmental Research and Public Health/International Journal of Environmental Research and Public Health*, vol. 16, no. 22, Multidisciplinary Digital Publishing Institute, Nov. 2019, pp. 4353–53, https://doi.org/10.3390/ijerph16224353. Accessed 10 July 2024.

Thayer, Julian F., et al. "A Meta-Analysis of Heart Rate Variability and Neuroimaging Studies: Implications for Heart Rate Variability as a Marker of Stress and Health." *Neuroscience & Biobehavioral Reviews/Neuroscience and Biobehavioral Reviews*, vol. 36, no. 2, Elsevier BV, Feb. 2012, pp. 747–56, https://doi.org/10.1016/j.neubiorev.2011.11.009. Accessed 10 July 2024.

"T-Test Calculator for 2 Dependent Means." *Social Science Statistics*, 2024, www.socscistatistics.com/tests/ttestdependent/default2.aspx. Accessed 10 July 2024.

"T-Test Calculator for 2 Independent Means." *Social Science Statistics*, 2024, www.socscistatistics.com/tests/studentttest/default2.aspx. Accessed 10 July 2024.

Williams, Nerys. “The Borg Rating of Perceived Exertion (RPE) Scale.” *Occupational Medicine*, vol. 67, no. 5, Oxford University Press, July 2017, pp. 404–5, https://doi.org/10.1093/occmed/kqx063. Accessed 10 July 2024.

“WORKPLACE STRESS - the American Institute of Stress.” *The American Institute of Stress*, 10 July 2024, www.stress.org/workplace-stress/#:~:text=83%25%20of%20US%20workers%20suffer,stress%20affects%20their%20personal%20relationships. Accessed 10 July 2024.

Prosopagnosia: An Under-Studied Neurological Condition By Hannah Jacob

Abstract

Agnosia is a loss of a specific neurological ability, where a patient is unable to recognize and identify objects, persons, or sounds. A type of agnosia is prosopagnosia, or face blindness (Kumar and Wroten, 2023). This condition can present in two main forms; developmental (at birth), or acquired (gained from a traumatic brain injury or other disorder). Prosopagnosia is on a spectrum, and many individuals in the world end up having developmental prosopagnosia. This agnosia is when one cannot recognize faces, emotions, or visual cues, as a part of their brain; the fusiform face area (or FFA) is damaged or underdeveloped. This can either be developed as a baby when the fusiform face area is unable to develop or gain later on in life, after a traumatic brain injury, a health condition, or another setback that affects the brain. Acquiring a more in-depth understanding may help accelerate the awareness of this condition and can further accelerate the advancements in treatments to help individuals undergoing prosopagnosia, as it is a very understudied disorder, only coming to research recently, near 2015.

Introduction

When undergoing feelings of isolation or depression, one might be under the impression that there is not a person in the world who cares for them, even when an individual may have many people who care for them. An individual who has been identified with prosopagnosia, however, could feel like they are truly alone. The millions of faces they see throughout their life look the same to them; each unrecognizable, known or not. Prosopagnosia, a visual agnosia, is the inability to recognize faces (Sorger et al., 2007). These individuals cannot tell the difference between two separate people, whether it be a stranger, a loved one, or even their face in the mirror. The word Prosopagnosia comes from the Latin word *prosop,* meaning “face”, *-a-* for “not”, and *-agnosia* for “know” (Sundaram, 2023). This condition, developed at birth or acquired from an accident or larger impairment, happens when the ventral occipitotemporal regions in the brain are damaged (Sorger et al., 2007). This is more common than thought, as around 2.5% of the nation is affected by prosopagnosia, resulting in approximately one in fifty people thought to have prosopagnosia (Corrow). This condition is so common, yet not many people know about it. The impairment affects many, even well-known individuals such as Jane Goodall, Brad Pitt, and Steve Wozniak. Pitt reported in an interview that he had never been formally diagnosed, but had struggled for years to recognize faces, often leading him to isolate himself as a result. “If I see you tomorrow I won’t know I’ve ever seen you before,” Wozniak states, “unless you have strange hair, certain clothing, a voice that I can recognize.” (Sinclair, 2021) Many with this agnosia memorize attributes of a person, such as their hair, voice, gait, or another prominent feature, and use that to recognize the individual. However, this tactic can backfire, such as when the person gets a haircut or changes how they speak. Prosopagnosia can be seen as a symptom in larger conditions, such as Alzheimer's, schizophrenia, or the result of a stroke (Cabrero and Jesus, 2023). It also has often been linked to dementia. Prosopagnosia does not get as much recognition as it needs; it is a common impairment, and yet many aren't informed enough. This condition does not have to be developed as a child–it can happen to anyone at any given time. An accident, big or small, may leave one with an impairing disorder for the rest of their life. Since a cure has not yet been found, the most we can do right now is make sure that we and our loved ones are informed about this common chronic condition.

Prosopagnosia

History

In 1947, German neurologist Joachim Bodamer published a report on several patients who experienced selective difficulty in recognizing faces and created the term 'prosopagnosia' to describe patients with difficulty recognizing faces, but with no other visual processing difficulties. When Bodamer asked his patients to inspect their own faces in a mirror, it was clear that although he knew that they were looking at their own faces, there was no shown sense of familiarity. These patients were able to describe what a nose, mouth, and eyes were, but did not understand that their face was the one shown in the mirror, showing that the problem did not appear to be directly caused by an impairment in vision. Bodamer's report also described impaired face recognition in wounded soldiers, many of whom were prone to brain injuries while in combat (Robotham and Starrfelt, 2018).

Types

Prosopagnosia is a very common agnosia. There are two types of prosopagnosia —developmental (often called congenital) and acquired. Developmental prosopagnosia is obtained when developing in the womb as the brain fails to develop a part of the brain, while acquired prosopagnosia is present as a symptom in health conditions or brain injuries. (Kumar and Wroten, 2023)

Prosopagnosia is further broken up into two more variations of the disorder; apperceptive prosopagnosia and associative prosopagnosia (Kumar and Wroten, 2023). Apperceptive prosopagnosia describes an individual with the inability to recognize one's facial expressions, identity, or other non-verbal cues. The people still know what a nose, a mouth, and eyes look like. Associative prosopagnosia describes someone who can recognize a face, but cannot remember who the face belongs to, and cannot seem to place an identity to the said face. Both types can affect an individual with either developmental or acquired prosopagnosia. These individuals keep track of identities by characteristics such as gait, style of hair, and voice.

Symptoms

Symptoms besides a lack of facial recognition include difficulty recognizing emotions on people's faces, finding one's way around simple routes, and trouble identifying the age and gender of others. Moreover, individuals with prosopagnosia may have trouble identifying characters and following plots in TV programs or films. This may make it harder for the person to form relationships, which may affect their mental health and lead to functional impairment, social isolation, lack of fear of strangers, intense separation anxiety, behavioral issues, and the development of psychological disorders such as social anxiety or depression (Cabrero and Jesus, 2023). Thus, prosopagnosia can have both a psychological and social impact on the individual. Developmental prosopagnosia can be inherited as an agnosia, but acquired prosopagnosia mainly appears as a symptom in other health conditions. There is little to no data on acquired prosopagnosia, as this condition is difficult to diagnose, and has only been recently studied further (Armata, 2024).

This visual agnosia lies on a spectrum, as individuals can have different levels of the agnosia. As science advances, the future may lead to new diagnoses of millions who may have undiagnosed prosopagnosia.

Rarity

Although acquired prosopagnosia is not as common, affecting one in 30,000 individuals in the country, developmental prosopagnosia affects many more, affecting one in thirty-three individuals (Nealon, 2023). Joseph deGutis, an Associate Professor at Harvard University experimented on the rarity of prosopagnosia with 3,341 individuals getting tested for prosopagnosia. The results showed that thirty-one out of the 3,341 had major prosopagnosia, while seventy-two had a milder form (deGutis, 2023). DeGuitis' work sheds light onto topics as understudied as prosopagnosia, but this agnosia needs more attention than it gets, as the condition is so common, yet it can damage one's life.

It was only in 2014, that the National Health Service added prosopagnosia to 'NHS Choice', and in 2015 it was added to the 'Long-term conditions that can be coded onto patient' notes. (Face Blind UK, 2021) There is a form of awareness that is starting to form, but it will take quite a while before specialists gain enough insight into this agnosia. Thus, future research should explore this condition in more detail.

	Etiology	Rarity	Diagnostic	Appearance
Developmental	- Parts of the FFA and/or OFA not developed	- 1in 33 individuals	- Sensory tests (vision-related) - Cognitive & mental status tests - Memory tests - CFMT - Object recognition tests	- Developmental deficits
Acquired	- TBI - Other brain disorders - Toxins	- 1in 30,000 individuals		- Symptom along with another impairment

Table 1. The comparisons between developmental and acquired prosopagnosia

Developmental and Acquired Prosopagnosia

Overview

As stated before, prosopagnosia comes in two different components; developmental and acquired, both of which affect the facial recognition pathway (Table 1). Developmental prosopagnosia (DP) starts in the womb. The baby in the mother's abdomen had failed to develop part of the fusiform face area (FFA) and the occipital face area (OFA) (Collins and Olson, 2014). Developmental prosopagnosia is often found in children with autism spectrum disorder, epilepsy, and learning disabilities. Developmental prosopagnosia is much more common than acquired prosopagnosia, thus, there is more information on DP than acquired. Acquired prosopagnosia (CP; the C standing for congenital) is a little different, as this agnosia is not developed, but attained after a traumatic brain injury (TBI), when toxins enter your body like carbon monoxide poisoning, or through other health conditions such as Alzhiemer's (Cleveland Clinic, 2022). Acquired prosopagnosia shows up more as a symptom than an individual disorder.

Diagnosis

Prosopagnosia is diagnosed by using a few tests; the biggest one being the Cambridge Face Memory Test (CFMT) (Fig. 2 Geskin and Behrmann, 2017). This test is available for diagnosing prosopagnosia in other populations and is evolving more and more to be suitable and more precise in diagnosing prosopagnosia. The CFMT presents itself as six learned faces, and throughout the test, the participants need to recognize the learned faces from the unfamiliar, distracting faces. The test goes on to three stages; one for the recognition of the same images, another for the recognition of the same faces morphed into different images, and the faces in different viewpoints or lighting recognition of the same faces covered with heavy visual noise (Bowles et al., 2009).

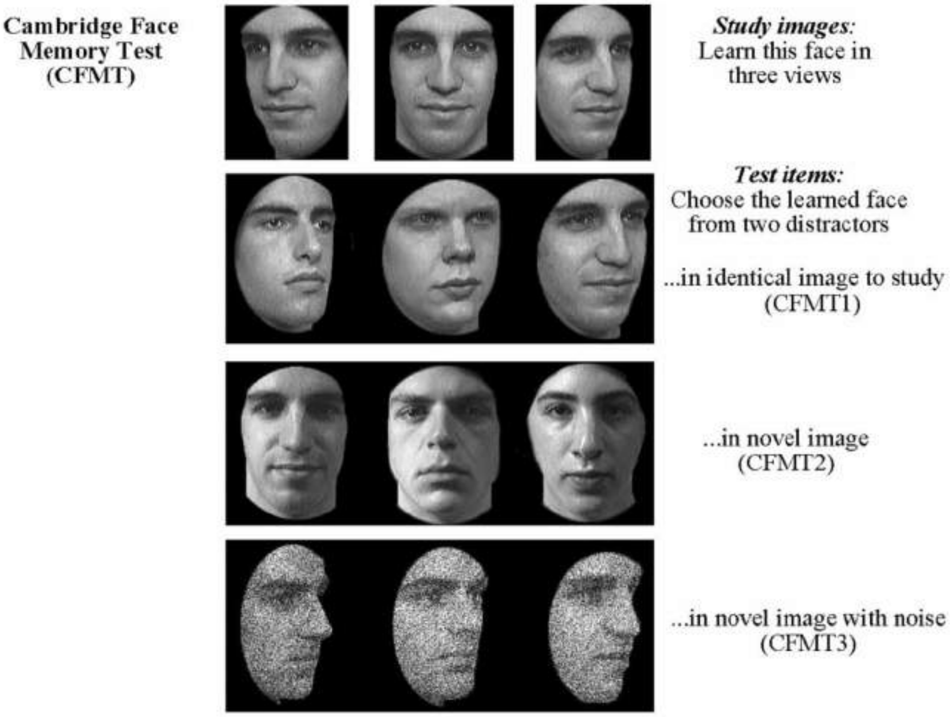

Cambridge Face Memory Test (CFMT)
Study images: Learn this face in three views
Test items: Choose the learned face from two distractors
...in identical image to study (CFMT1)
...in novel image (CFMT2)
...in novel image with noise (CFMT3)

Fig 1. The Cambridge Face Memory Test (CFMT)

However, there is no known cure for prosopagnosia, as this agnosia is still very newly studied. In the future, there is hope for one, so that these people overcome the obstacles placed in front of them.

Neuroimaging and Neuroanatomy of Prosopagnosia

Overview

The ventral temporal cortex is responsible for object recognition, which contains the FFA, a part of the brain being the facial recognition portion (Collins and Olson, 2014). The FFA is a specialized face-processing network located in the lateral middle fusiform gyrus (Fig 3). It responds more strongly to faces than to other objects. The FFA processes facial identification, emotion, and nonverbal cues. The occipital face area (OFA) is located upstream from the FFA (Fig 3), on the inferior surface of the occipital gyrus (Fig 4) (Zhao et al, 2022). The OFA likely contributes to an earlier stage of face analysis than the FFA, and both are activated automatically when viewing a face.

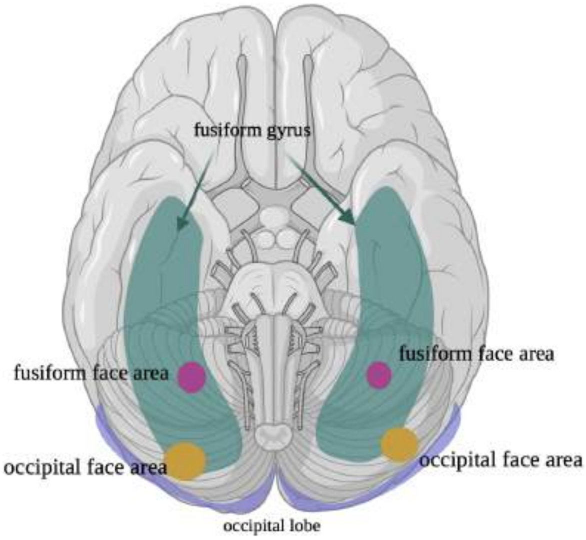


Fig 2. Locations of FFA, OFA, and fusiform gyrus in brain

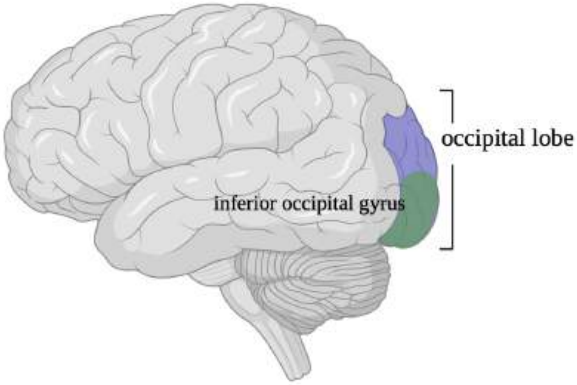


Fig 3. The location of occipital gyrus (where OFA is located)

Effects

Evidence from FMRIs of the FFA development when aging is prominent. The right hemisphere is shown to increase in size– the dominant hemisphere of the brain that processes faces. From the fetus to adulthood, the FFA slowly increases in size. Underdevelopment of the FFA causes problems such as prosopagnosia to arise. Individuals with developmental prosopagnosia showed decreased multi-voxel pattern stability for repeated faces in the bilateral medial temporal lobe (MTL), and its unstable mnemonic representation was associated with impaired face recognition performance in prosopagnosia. Many cases of acquired prosopagnosia have extensive damage to the temporal lobes and to the OFA, with both bilateral or unilateral lesions.

Previous research from the Department of Psychology and the Zlotowski Center

for Neuroscience in 2008 performed a study to understand the difference between an individual with acquired prosopagnosia versus one without by looking at their brain activity; more specifically, their ventral occipitotemporal cortex (VOTC), involved in the perception of visually presented objects (Avidan and Behrmann, 2009). To address their hypothesis, they used FMRI adaptation to view the difference in brain activity between individuals with CP and those without. Their experiment used six individuals with DP and twelve without, who were given the task of deciding if the two photos shown were the same person or not. They came to the conclusion that people with CP were slower to answer and were less accurate than the other group, but both groups were able to perform better when the face of a famous person was shown, and while looking at the VOTC, they found out that people without CP had a change in their VOTC from when viewing a familiar face to an unfamiliar one. CP individuals, however, did not. While there were no other differences to be found in the VOTC, it was to be concluded that these core VOTC areas don't properly activate these extended regions, preventing effective face recognition in CP.

Conclusions

Taken together, prosopagnosia is an agnosia that is both common and understudied, and when affected, this condition can change your life. It is on a spectrum and is presented when the OFA in the occipital lobe and the FFA in the temporal lobe are affected. Prosopagnosia is found as developmental (or congenital) and acquired, and further broken down as adaptive or associative prosopagnosia. As it is an agnosia, prosopagnosia is different from visual blindness, as this disorder is from a defect in the neurological aspect instead of visual. The field of studies in prosopagnosia should be further investigated, and more improvements should be made to both the research and the treatment of prosopagnosia. This condition has not been studied enough, leading to information that is not yet discovered and can be uncovered in the future. Furthermore, in the time to come, a cure will hopefully be found, benefiting many who have this agnosia for the better.

Works Cited

“Prosopagnosia or Prosopdysgnosia” *Aalbrog University*, https://vbn.aau.dk/ws/portalfiles/portal/267294058/PID5173079.pdf. Accessed 7 May 2024.

Corrow, Sherryse L. “Prosopagnosia: current perspectives.” *Taylor & Francis Online*, https://www.tandfonline.com/doi/full/10.2147/EB.S92838.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5398751/#b3-eb-8-165. Accessed 7 May 2024.

Gerlach, Christian, and Ro Julia Robotham. “Chapter 9 - Object recognition and visual object agnosia.” *Science Direct*, https://www.sciencedirect.com/science/article/abs/pii/B9780128213773000088?via%3Dihub. Accessed 7 May 2024.

Kumar, Anil, and Michael Wroten. “Agnosia.” *PubMed*, https://pubmed.ncbi.nlm.nih.gov/29630208/. Accessed 6 May 2024.

Lee, Andrew Go, and Seema Sundaram. “Prosopagnosia.” *EyeWiki*, 6 April 2023, https://eyewiki.aao.org/Prosopagnosia. Accessed 6 May 2024.

“Science Direct.” *Visual Agnosia*, 5 March 2024, https://www.sciencedirect.com/topics/neuroscience/visual-agnosia. Accessed 7 May 2024.

Sinclair, Allison. “My face blind life | The Channel.” *Ingenium*, 11 January 2021, https://ingeniumcanada.org/channel/blog/my-face-blind-life. Accessed 6 May 2024.

Sorger, Bettina, et al. “Understanding the functional neuroanatomy of acquired prosopagnosia.” *Science Direct*, Elsevier, April 2007, https://www.sciencedirect.com/science/article/pii/S1053811906009906. Accessed 6 May 2024.

"Prosopagnosia: Pathogenesis and Treatment." *NCBI Bookshelf*, U.S. National Library of Medicine, 2020, www.ncbi.nlm.nih.gov/books/NBK559324/.

Nierenberg, Cari. "How Common Is Face Blindness?" *Harvard Medical School News*, 13 Mar. 2023, hms.harvard.edu/news/how-common-face-blindness#:~:text=Published%20in%20February%202023%20in,for%20face%20blindness%2C%20or%20prosopagnosia.

"History." *Face Blind*, www.faceblind.org.uk/information/history/.

"Face Blindness (Prosopagnosia)." *NHS*, www.nhs.uk/conditions/face-blindness/#:~:text=The%20main%20symptom%20of%20prosopagnosia,they%20do%20not%20know%20well.

"Prosopagnosia (Face Blindness)." *Cleveland Clinic*, 15 Mar. 2022, my.clevelandclinic.org/health/diseases/23412-prosopagnosia-face-blindness.

Biotti, Federico, et al. "Congenital Prosopagnosia without Object Agnosia: A Literature Review." *Cerebral Cortex*, 14 Feb. 2020, dpl6hyzg28thp.cloudfront.net/media/Congenital_prosopagnosia_without_object_agnosia__A_literature_review.pdf.

Dalrymple, Kirsten A., et al. "The Effect of Facial Expression on Face Matching in Developmental Prosopagnosia." *Neuropsychologia*, vol. 163, 2022, www.sciencedirect.com/science/article/abs/pii/S0028393222002299?via%3Dihub.

Rezlescu, Cristian, et al. "Individual Differences in Prosopagnosia: A Study of Congenital and Acquired Cases." *Royal Society Open Science*, vol. 7, no. 200884, 2020, royalsocietypublishing.org/doi/10.1098/rsos.200884.

Biotti, Federico, et al. "Diagnosing Prosopagnosia: Effects of Ageing, Sex, and Participant-Stimulus Ethnic Match." *Cerebral Cortex*, 14 Feb. 2020, dpl6hyzg28thp.cloudfront.net/media/Diagnosing_prosopagnosia__Effects_of_ageing__sex__and_participant_stimulus_ethnic_match__lY8qdLi.pdf.

DeGutis, Joseph M., et al. "Dissociating Face Processing Skills: Insights from Developmental Prosopagnosia." *Behavior Research Methods*, vol. 44, no. 4, 2012, link.springer.com/article/10.3758/s13428-011-0160-2.

Barton, Jason J. S., et al. "What Else Is Different in Developmental Prosopagnosia?" *Neuropsychologia*, vol. 67, 2015, www.sciencedirect.com/science/article/pii/S0028393214001869.

Mannion, Damien J., et al. "Prosopagnosia." *EyeWiki*, American Academy of Ophthalmology, eyewiki.aao.org/Prosopagnosia#:~:text=Facial%20recognition%20is%20processed%20in,the%20patient%20complains%20of%20prosopagnosia.

"Face Blindness (Prosopagnosia)." *Osmosis*, www.osmosis.org/answers/face-blindness.

Haxby, James V., et al. "The Inferior Temporal Cortex and the Perception of Faces, Objects, and Scenes: A Review." *Frontiers in Human Neuroscience*, vol. 7, no. 7, 2013, www.ncbi.nlm.nih.gov/pmc/articles/PMC6404234/#:~:text=Inferior%20temporal%20cortex%20(IT)%20is,%2C%20face%2C%20and%20scene%20perception.

Duchaine, Bradley, et al. "Prosopagnosia." *International Encyclopedia of the Social & Behavioral Sciences*, 2nd ed., Elsevier, 2015, www.ncbi.nlm.nih.gov/pmc/articles/PMC10901275/.

Bukach, Cindy M., et al. "The Cognitive Neuroscience of Face Processing." *Frontiers in Psychology*, vol. 4, no. 756, 2013, www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2013.00756/full.

Rizzo, Matthew, and Mario F. Mendez. "Prosopagnosia." *Fundamentals of Cognitive Neuroscience*, 3rd ed., Elsevier, 2018, www.ncbi.nlm.nih.gov/books/NBK493156/#:~:text=Introduction,or%20unfamiliarity%20with%20the%20stimuli.

The Current State of Search and Rescue Robot Concepts, its Complications, and its Potential By Arya Shah

Abstract

Search and Rescue robotics is becoming a fast-growing field as more ideas and designs are being developed and published. Designs range from robots with snake-like chassis to help navigate terrain to tank-drive chassis with a rocker arm to climb difficult obstacles. These designs come with benefits and limitations, and it is important to analyze them in order to determine what is the best solution. Despite their potential, Search and Rescue robots come with plenty of complications. An ineffective robot could end up endangering the lives of people it is trying to save, and there is no guarantee that the robot will not malfunction or receive damage along the way. In summary, Search and Rescue robots have the potential to save lives more effectively than human rescuers and prevent human rescuers from risking their own lives; however, these robots can also be susceptible to many problems, and a viable design solution is important in order to make them effective.

Introduction

Search and Rescue is a key part of emergency response operations. It involves locating and assisting endangered individuals from a variety of possible emergencies including natural disasters, hostage situations, and more [1]. Search and Rescue teams typically contain specialized and trained personnel who work on extraction and communication. These operations are very important as they are the first response to big emergencies and are one of the biggest factors in reducing the death count of events such as natural disasters. Search and Rescue operations typically take place in hazardous conditions such as in the ocean, in wilderness, or in disaster zones [1]. These environments come with lots of dangers, including large fallen obstacles, difficult navigation due to rubble and destroyed roads, strong winds or large waves, limited visibility, dangerous fumes, and more. Because of this, rescuers on these teams often put their lives in danger in order to locate individuals. During the rescue period after the 9/11 attacks, over 91,000 workers and volunteers were exposed to hazardous conditions. Out of these, 343 firefighters and paramedics along with 23 NYPD officers were killed during rescue operations [2]. Many others died years after the attacks from health problems caused by injuries and exposure [3]. These deaths, along with the many lives that had to be risked saving others, can be prevented through the use of Search and Rescue robots. These robots would be able to replace Search and Rescue extraction workers by navigating the terrain and searching for individuals. They have the potential to be more effective in searching for people and reach survivors faster than human rescuers. They can also navigate terrain better through the use of scanners, sensors, camera vision, and a chassis that can adapt to terrain. These benefits, in addition to the ability to mass produce these robots make them a great alternative to human rescuers [4]. There are currently many different types of Search and Rescue robot prototypes that serve different purposes. These robots range from drones which are used to scout the area to ground vehicles which can find survivors. This review will be focusing on land-based robots in order to have a more direct and effective comparison. Additionally, we will be assessing these robots based on their ability to navigate terrain and find survivors, as most current Search and Rescue robot designs do not include systems for communication with survivors and survivor extraction. Search and Rescue robots are currently in an early stage, but four main prototype designs have begun to emerge: tanks, rocker arm tanks, wheeled robots, and snake robots. This research paper will review the different designs and prototypes for rescue robots and determine which will be the most effective by analyzing them in multiple categories.

Search and Rescue robot subsystems

For a Search and Rescue robot to be effective, there are certain subsystems that are required. These include some kind of drivetrain for mobility and navigating terrain, and a variety of sensors and cameras in order to detect survivors and terrain as well as sending information back to the operators. The drivetrains can come in many forms, which will be discussed later. These drivetrains are specialized for the types of terrain they will be navigating through. As for sensors, a variety are needed in order to effectively navigate and locate survivors. Standard cameras are useful for the operators who are driving the robot in order for vision. These cameras must be mounted in a position where they give the maximum visibility. In addition to these cameras, sensors which can help detect obstacles, such as ultrasonic sensors, and sensors which can detect survivors, such as infrared sensors, are also a must [5].

Search and Rescue robot requirements

There are also certain requirements that must be met for a Search and Rescue robot. These robots cannot be too large, as this will make them difficult to transport quickly to necessary locations and will also make deploying them much harder. A large robot also creates drawbacks with costs and adaptability. Durability is another key factor with robot designs. Because these robots are going to be deployed in unpredictable situations and environments, they must be able to withstand many conditions including hot and cold temperatures, rain or wet conditions, and unique terrain conditions. Because of these numerous factors, Search and Rescue robots need to be durable enough to survive falls, completely waterproof, and must be manufactured so that the parts are well-built and are able to cover the electronics. One final requirement for Search and Rescue robots is ease of operation. A well-built robot will not be effective if the operators have a hard time controlling it, which means that an easy to use robot is essential. In order to make a robot easy to operate, it must not have too many subsystems or moving parts which the operators have to control, and certain processes should be automated. Communication systems also play a role in this. These robots should be able to send back information to operators effectively, as this information is what allows them to operate the robot. Having a strong wireless connection is essential for this [6].

Tank Robots

Tank drive search and rescue robot designs were some of the first designs ever created, and they continue to be used because of the unique benefits they provide. Tank designs consist of two to three wheels connected together by tread. The tread allows the tank drive to easily move over rough terrain and small obstacles. Larger treads are often used with these robots in order to provide more mobility over rough terrain.

Tank robots are effective search and rescue designs due to their simplicity, durability, and size flexibility. Tank drive designs are fairly simple because there is only one main feature in the drivetrain. Additionally, tank drive designs have a lot of pre-existing research behind them as they have been used for many other purposes. These two elements make tank drives simpler and easier to build than many other designs. Tank drive designs are also simple when it comes to operation, as the tank is the only moving element. A tank drive only needs one operator, and the controls for a tank are very simple. This means that a tank design will take minimal operator training time. Tank drives designs are also very flexible, allowing for a variety of shapes and sizes to be effective depending on the task. For search and rescue robots, smaller tank drives provide greater maneuverability through tight spaces and are cheaper while bigger tank drives allow the robot to climb over bigger and more unconventional obstacles, such as stairs. The lack of complexity in tank robots also allows for a more durable design. Tank designs have minimal weak points as the design does not consist of any joints or other areas susceptible to damage. This allows tank robots to survive big falls and big impacts with terrain better than most other designs. These advantages allow for tank robots to be a viable search and rescue robot design.

However, tank robots also have some important disadvantages. One of these disadvantages is its ability to climb large obstacles effectively. Small tank robots are unable to get enough grip to climb such obstacles, and large tank robots have to trade off maneuverability through small areas in order to climb bigger obstacles. Tank robots also struggle with getting out of ditches or other places where they might get stuck. This is because tank robots can only rely on the grip present on their tread to climb out of steep areas, and if it is in a position where it can’t move, it has no moving parts which it can use to wiggle out.

Tank robot designs have been designed and used in competitions such as the Robocup Rescue League, where different Search and Rescue robots compete in multiple objectives simulating a real Search and Rescue scenario. Two examples of tank drive robots in the competition are Hector Darmstadt’s 2019 robot and ATR team’s 2019 robot [7, 8]. Hector Darmstadt’s robot is a great example of an extremely compact tank robot. The robot’s incredibly small dimensions allow it to maneuver very well and fit through gaps easily.

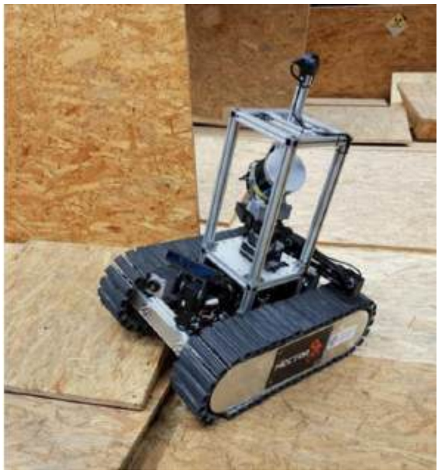


Figure 1: Hector Darmstadt 2019 robot [7]

The lack of complexity allows for a stable design which is less susceptible to damage than most other robots in the competition. These advantages allowed this robot to get third place in 2019, proving that a simple design can be very effective (Table 1). ATR team's robot is bigger, but it comes with its own advantages.

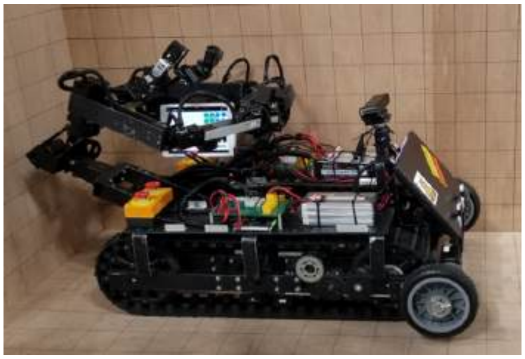


Figure 2: ATR Team 2019 robot

Firstly, it has a unique feature attached to the tank, which is the two tires in the back. These tires give the robot additional mobility by allowing it to climb over larger obstacles. For example, the robot is able to climb over steep surfaces such as stairs easier by using the back tires as support as the tank portion crawls up the wall [8]. This solves one of the biggest weaknesses of the tank robot, its ability to climb big obstacles. It also does not take away from the durability as the tires are securely integrated with the rest of the robot and do not contain an extra joint.

Outside of the Robocup Rescue League, Search and Rescue robots with tank designs are already being used. The Howe & Howe Thermite S1 is an example of a tank robot that can be used in Search and Rescue scenarios [9]. Although this robot was designed for firefighting purposes, it still contains features that enable it to cover rough terrain and search for survivors, which is needed for Search and Rescue scenarios. This robot is different compared to the robots in the Rescue League because of its enormous size. This allows it to climb over bigger obstacles including steep areas and stairs, as its large tread allows it to grip obstacles very well.

Additionally, this robot is built with far stronger materials than the Rescue League robots, making it very durable and able to withstand many weather conditions [9].

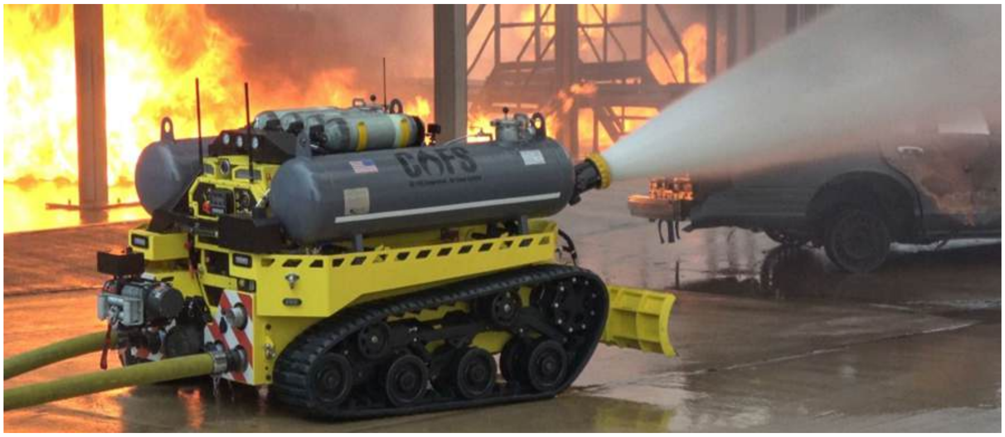


Figure 3: Howe & Howe Thermite S1 [9]

The QinetiQ Dragon Runner 10 is another Search and Rescue robot which appears at the opposite end of the spectrum when it comes to size. It is built to be incredibly compact and light, weighing just around 10 pounds. Its size gives it the advantage of being extremely portable, as it is designed to be carried on the backs of rescue workers [10]. This means that the Dragon Runner can be deployed anywhere rescue workers can reach, which is a big advantage compared to a robot like the Thermite S1, which requires multiple people along with a vehicle to deploy.

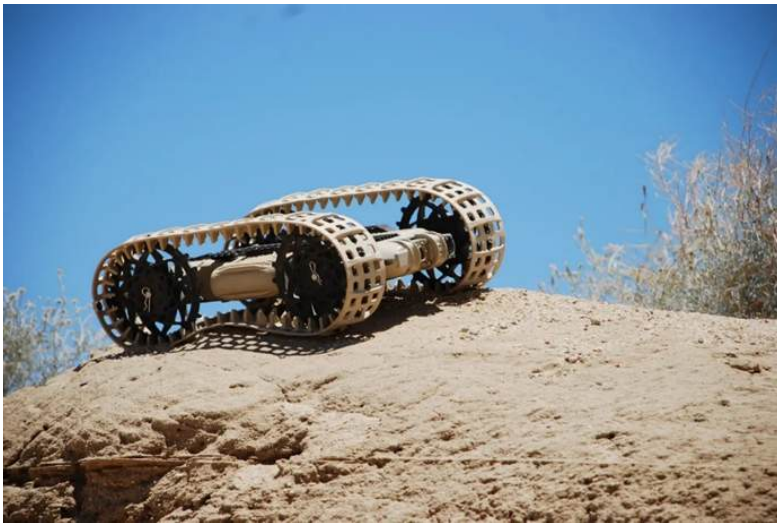


Figure 4: QinetiQ Dragon Runner 10 [10]

Overall, tank robots have proven to be an effective design in the Rescue League and in the real world, as the many shapes and sizes they appear in provide multiple different advantages.

Table 1 (Robocup 2019):

Team:	Max Speed:	Weight:	Cost:	Size:	Placement	Awards:

Hector Darmstadt [7]	0.6m/s	25kg	$4,000	0.6 x 0.42 x 0.6 m	Third place	Best-In-Class Mobility
ATR Team [8]	0.48m/s	56kg	$95,000	1.55 x 0.55 x 0.3 m	N/A	N/A

Rocker Arm Robot

The rocker arm design is one of the most researched designs, and there are many prototypes out there for this design. This design is similar to the tank drive in that it uses a track over the wheels which allows it to climb over small obstacles. However, this design also has a rocker arm which gives it extra mobility. The rocker arm is attached near the front wheel of the tank, and it is essentially an extension of the tank drive which can rotate on an axis. Rocker arm designs come with either two or four rocker arms. These designs are very effective at moving over large terrain, as the rocker arm gives them the ability to climb over obstacles with extra dexterity. The rocker arm can also be used for other things such as moving obstacles out of the way. Rocker arms are fairly simple to operate and only need two operators: one driving the robot, and one operating the rocker arm. Rocker arm designs are more complex than wheeled and tank robots, however, because they have an extra element of mobility added to them. This also tends to increase the costs as more motors and other parts are needed to create and power the rocker arm. The rocker arms on these robots add an extra weak point, as the joint that connects the rocker arm to the rest of the robot will not be as stable as the rest of the robot and is most likely to break off after a big impact. Rocker arm robots are by far the most popular in the Robocup Rescue league, and some examples of these robots are Shinobi and iRAP Sechzig's 2019 robots [11, 12]. These robots are great examples of well-designed rocker arm robots as they both were able to navigate terrain and complete objectives very successfully, earning them first and second place respectively along with mobility and dexterity awards. Shinobi's robot has an average weight despite being a larger robot. This combination allows for the main body of the robot, the tank portion, to be very large as the tread covers a large area. This aids the robot with moving over smaller rubble without getting caught, as the large tread is able to glide over these obstacles [11]. The four rocker arms on this robot give it great mobility. Despite the rocker arms each being much smaller than the main body, the rocker arms are still very powerful as they cover the weaknesses of the main tank body, which involve climbing over bigger obstacles. The rocker arms can adjust the angle at which they hit surfaces in order to gain a greater grip and climb easier. Additionally, this robot has great build quality as there are few openings and exposed wires where the robot could be damaged.

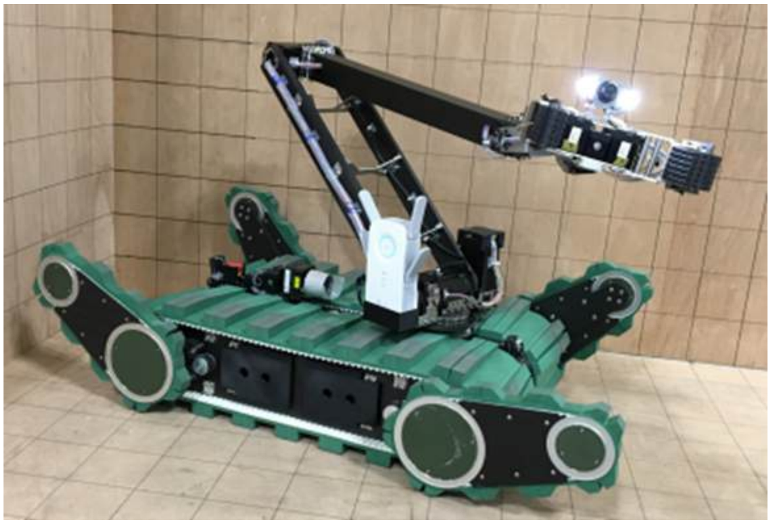


Figure 5: Shinobi 2019 robot [11]

iRAP Sechzig's rocker arm robot is very similar to Shinobi's. It also has four rocker arms which give it great mobility along with a large main body. However, this robot is much heavier at 77 kilograms, and this causes it to lose some dexterity as it is clunkier (Table 2). This would also be a drawback in a real-life situation as it would be much harder to deploy quickly by rescue workers. This design also features much larger rocker arms than Shinobi's robot, and this allows it to cover larger obstacles. There are also a few notable Search and Rescue robots outside of the Rescue League that use rocker arm designs. The iRobot 110 FirstLook is an ultra-mobile and lightweight rocker arm robot designed for Search and Rescue [13]. This robot weighs only 5 pounds, making it extremely easy to carry in hand or in a backpack. The robot is also extremely durable as it is designed to be thrown from a safe distance into dangerous areas that need to be searched. In order to allow for such a light design, this robot has a unique rocker arm that does not have any tread on it; instead, it is just a lever. The robot is able to use this lever to pull itself over obstacles and flip itself over, giving it extra mobility as well as the ability to right itself after big drops [13]. This robot design is far more durable than the robots in the Rescue League because it was built more securely in order to ensure it can survive being thrown from long distances. However, the lack of tread on the rocker arm means it has a harder time climbing obstacles, and the smaller rocker arm decreases the arm's effectiveness at pulling the robot over obstacles. These drawbacks come with the small design, and are necessary in order to achieve portability.

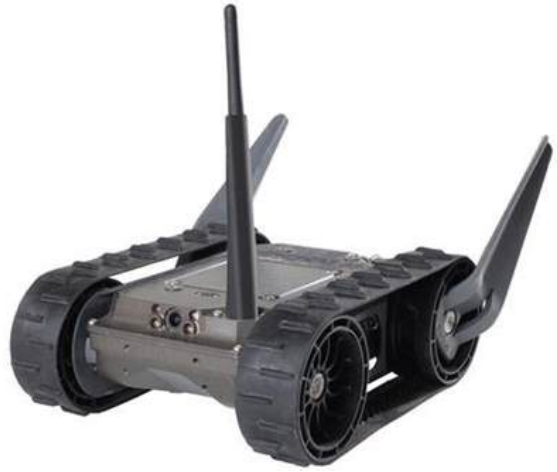


Figure 6: iRobot 110 FirstLook [13]

The iRobot 710 Kobra is another rocker arm Search and Rescue robot in the industry[14]. It is much bigger than the iRobot 110 FirstLook, and it is clearly not designed to be carried by hand as it weighs 367 pounds [14]. However, it includes a much bigger rocker arm that has tread, allowing it to climb bigger obstacles much easier. Additionally, it comes with a payload area for loading supplies and other necessities [14]. This feature is very useful as it can be used to send supplies to survivors once they are found. Rocker arm robots are a proven Search and Rescue robot design, as it is the most used in the Rescue League and is very common in the corporate world as well.

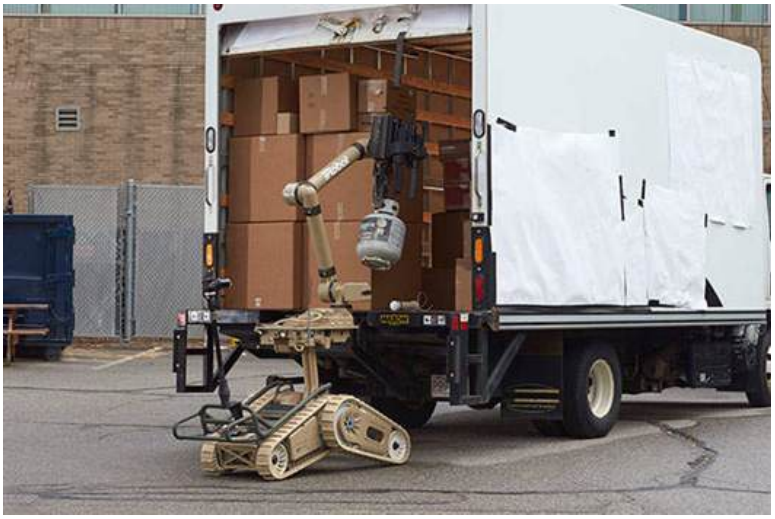


Figure 7: iRobot 710 Cobra [14]

Table 2 (Robocup 2019):

Team:	Max Speed:	Weight:	Cost:	Size:	Placement	Awards:

iRAP SECHZIG [12]	2m/s	77kg	$23,000	0.6 x 1.2 x 0.6 m	Second place	Best-In-Class Mobility
SHINOBI [11]	2m/s	35kg	$32,700	1.0 x 0.5 x 1 m	First place	Best-In-Class Dexterity
NuBot [15]	2m/s	28.41kg	$37,000	0.6 x 0.6 x 0.5 m	N/A	N/A
MRL [16]	0.85m/s	95kg	$35,000	0.8 x 0.6 x 0.6 m	N/A	N/A
MARS-Rescue [17]	0.28m/s	4.2kg	$3,640	0.29 x 0.24 x 0.13 m	N/A	N/A

Table 3 (Robocup 2022):

Team:	Max Speed:	Weight:	Cost:	Size:	Placement:	Awards:
DYNAMICS [18]	1m/s	65kg	$27,500	0.8 x 0.5 x 0.4 m	N/A	N/A
Hector Darmstadt [19]	1.2m/s	58kg	$30,000	0.72 x 0.51 x 0.6 m	Second Place	Best-In-Class Autonomy

Wheeled Robot:

Wheeled robots are a classic design that is present in lots of robots. The standard wheel design includes either four or six wheels, with two or three on each side of the robot. For Search and Rescue robots, tires are used for the wheels since the flexibility of the rubber makes it easier to travel over rubble. Additionally, the tires come in very large sizes so that the robot can climb bigger obstacles while not getting the tires stuck in small spaces. Wheeled robots sometimes include suspensions for the wheels as well. Suspension allows for smoother travel over rough areas and makes the robot less susceptible to damage when moving fast over rough terrain. Wheeled robots come with many advantages because of its simple design. Firstly, wheeled robots tend to be very durable. Similar to tank robots, wheeled robots do not have many points of weakness because of their simple design. There are no joints in these robots, and the inflated tires combined with suspension provides a lot of cushion when landing from big falls. Another advantage is the portability of wheeled robots. Smaller wheeled robots are very easy to carry and hold in hand, making them easy to deploy. Even larger wheeled robots that can't be picked up can still be moved easily by pushing them and taking advantage of the rolling wheels. This means that larger robots can easily be moved into and out of transportation vehicles by just one or two people, and it can be done very quickly. One final advantage of wheeled robots is how easy it is to operate them. Wheeled robots are simple to operate because only one operator is needed, and the only controls are moving each set of wheels forward and backward, similar to the tank robot. This means that operators can be trained quickly and achieve a high level of accuracy. Wheeled robots also have some notable drawbacks that need to be addressed. The first drawback is the lack of grip wheels get on surfaces. Compared to tank and rocker arm robots, wheeled robots make a lot less contact with the ground because each tire only has a small section in contact with the ground. This means that the tires have a lot less grip on the ground, preventing them from climbing steep or slippery surfaces and obstacles bigger than the tires. An example of a notable wheeled robot in the Robocup Rescue League is XFinder's 2019 robot [20]. This robot features a four-wheel design, with two wheels on each side. The wheels have large tires along with a suspension system, allowing the robot to move over small obstacles easily. Additionally, the bottom of the robot is a good distance away from the ground, meaning it won't get caught over large bumps. These advantages allow the robot to climb over a variety of obstacles, including small stairs and bumpy terrain [20]. However, the robot suffers from the same weaknesses as other wheeled robots, as it does not generate much grip with the ground.

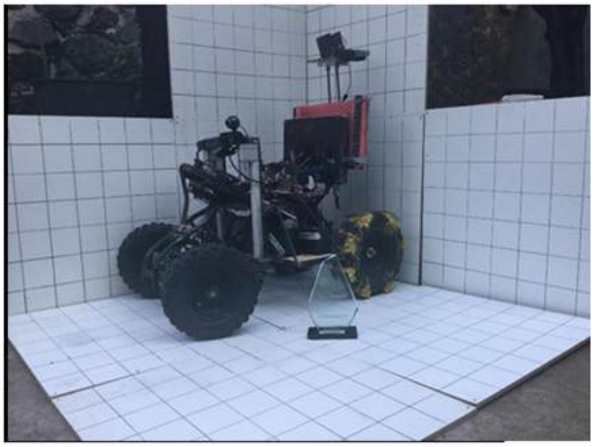


Figure 8: XFinder 2019 robot [20]

The Husky UGV is an example of a wheeled Search and Rescue robot in the industry, and it has a traditional wheeled robot design [21]. The Husky features a four wheel drivetrain with two wheels on each side. The tires are 13 inches in diameter, which allow the robot to easily traverse rough terrain. The size and build of the Husky is incredibly simple and compact, and it is built with durable materials, allowing it to survive any climate. The Husky's most unique feature, however, is its customizability. The Husky has many attachments that can be added to the robot, from movable claws to simple storage, which can be very useful in different Search and Rescue situations [21]. In general, wheeled robots tend to struggle with large obstacles and slippery surfaces because of the lack of grip wheels create. However, they also come with many advantages because of their durability and simplicity, making wheeled robots a viable design.

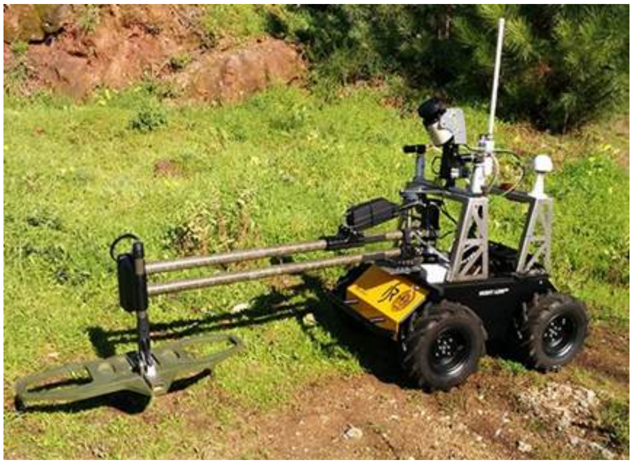


Figure 9: Husky UGV [21]

Table 4 (Robocup 2019):

Team:	Max Speed:	Weight:	Cost:	Size:	Placement	Awards:
XFinder [20]	0.8m/s	85kg	$4,000	1.0 x 0.6 x 0.78 m	N/A	N/A

Snake Robot

Snake Search and Rescue robot designs are some of the most unique, as they have a very unconventional design which solves prominent problems with other designs. Although snake robots come in many forms, there are some systems that are present in all types. Snake robots use multiple segments of moving parts to maneuver around obstacles and climb in tight spaces, similar to real snakes. Some advantages of snake robots are their ability to move through many different type of terrain. One of the snake robot design's greatest strengths is its ability to crawl through tight spaces. Snake robots are able to do this effectively because they can be made to be very thin, and the ability of each segment to twist and turn allows the robot to move through small passages easily. This flexibility also helps increase the surface area of the robot in contact with the ground, which helps with climbing a variety of obstacles as the robot has much more grip. The complexity of the snake robot design creates a lot of disadvantages. The segmented design creates a lot of different moving parts that need to work together in order for the robot to be functional. Additionally, snake robots involve a lot of different parts and have an unconventional style, making the building process more complicated. The multitude of joints between each section of the robot create lots of weak points. Finally, operating a snake robot is very difficult as it contains many degrees of freedom, meaning there are a lot more controls for an operator to manage.

Snake robots traditionally come in three different forms. The first is a segmented snake robot, which contains either wheeled or tank drive segments connected together. The second design relies mostly on joints to maneuver. This type of design does not have any wheels or tank to move itself, but instead relies on moving joints and therefore maneuvers similar to a real-life snake. The final snake robot design are soft snake robot designs. The body of these robots is made out of flexible materials that can easily change their shape, allowing for ultimate mobility through any space. Since most of these designs are still experimental and have not been put out for commercial use, statistics for these robots are unavailable. However, they can still be compared in categories such as maneuverability through large obstacles and small spaces, operability, and the usefulness of the unique abilities these robot designs have. One example of a segmented snake robot is the OmniTread OT4, which uses tank treads on each segment of the robot [22]. These tank segments are unique because they have treads on all four sides of the robot. This allows the robot to gain more grip in tight spaces where the sides and top of each segment come in contact with surfaces. The tank drive that is present on each segment gives it superior speed and traction compared to other designs. However, having treads on all four sides means the robot has to be fairly large in order to fit all of the systems, which can prevent the robot from fitting in very tight spaces.

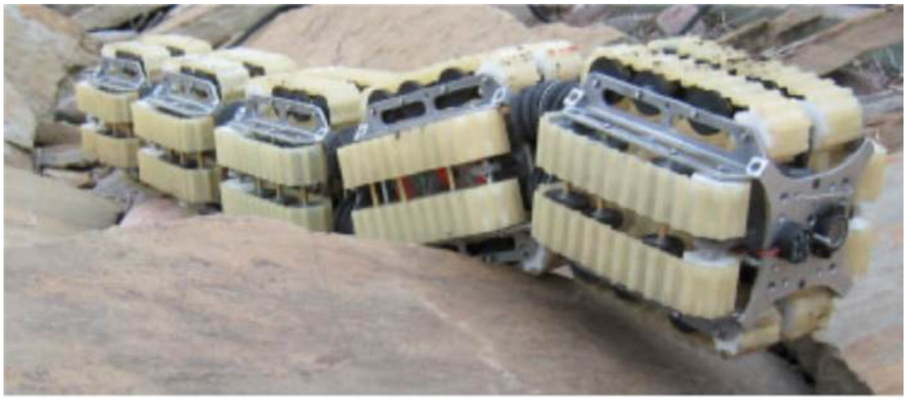


Figure 10: OmniTread OT4 [22]

Robots that rely on joints to maneuver can solve this, as they can be built very small due to the lack of large subsystems. The Carnegie Mellon Snake robot is an example of a snake robot that uses joints to maneuver [23]. This robot is much thinner and more compact than all the other robot designs, allowing it to climb in small spaces. The Carnegie Mellon robot was tested in small steam pipes, where it was able to successfully maneuver around multiple bends and valves while sending back a video feed [23]. This snake robot design can be extremely useful in use cases like this, where traditional robots would not be able to fit and maneuver around the small pipes. However, this snake robot design would not be good at covering large distances, as it does not have the ability to travel fast or climb slippery surfaces.

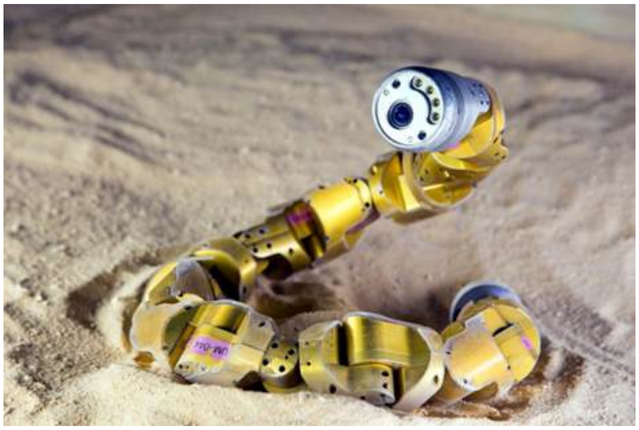


Figure 11: Carnegie Mellon Snake Robot [23]

Soft snake robots are the final and most unique design. This design manipulates air, allowing it to expel air in order to squeeze in spaces and expand by pumping air in. The Vinebot from Stanford University is an example of a soft snake robot which uses flexible cylinders and air to traverse terrain [24]. This robot is able to squeeze through any space by controlling the air pressure inside its body, and can also move or lift heavy objects by expanding while underneath it. This provides a variety of use cases in search and rescue scenarios, and gives this robot a lot of potential. Tests with this robot showed it could move across sticky glue and nails, climb an icy wall, and navigate the space above a dropped ceiling. This robot has a similar weakness to the previous design, as it cannot reach very fast speeds with its slow crawl. It also is prone to damage from sharp objects, as any puncture will result in the pressure in the robot to be released. Overall, snake robots come in many different variations and have unique characteristics compared to the previously mentioned designs due to their flexibility. However, they are still very experimental, and will require a lot of work to be ready to be used in the field.

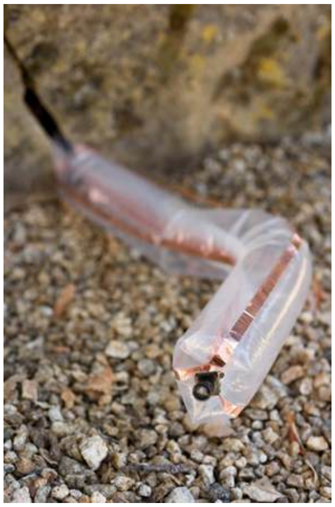


Figure 12: Stanford University Vinebot [24]

Conclusion:

Search and Rescue robots have come a long way, and there are many designs and prototypes that are ready to be used in the real world. However, determining which design is the most effective is much more complicated. Different situations call for various features that only certain designs have. However, a good way to compare these designs is by analyzing them in different categories where they would excel. These categories include durability, portability, maneuverability, cost, and complexity.

Robots that can survive large drops, impacts, and hazardous conditions are considered the most durable. Robot designs that avoid large and exposed joints and have minimal moving parts can excel in this category, as it reduces the number of areas where the robot could be damaged. Because of this, segmented snake robots, joint snake robots, and rocker arm robots are generally less durable. All of these designs rely on joints for mobility, creating weak points that are susceptible to damage after large falls. Between tank and wheeled robots, tank robots are more durable as they have tread protecting the drivetrain, while wheeled robots are susceptible to sharp objects that can puncture the tires or large divots in which tires can get stuck.

Portability is the next factor to consider for these robots. Portable robots are ones that can be transported and deployed easily in any situation. In order to achieve this, robots need to be lightweight, easily carried by one person, and deployable from a distance. Tank robots and soft snake robots can be lightweight, as they have many variations including large and small robots. Example of lightweight robots for these designs are the Stanford Vinebot and the QinetiQ's Dragon Runner 10. These robots are light due to the materials used, and they are small enough to easily be carried by a single individual. Rocker arm robots and tank robots are generally not lightweight because rocker arms add a significant amount of weight and size to the robot and the large wheels needed to climb obstacles make wheeled robots large and heavy as well. However, these two robots defer when it comes to deployment. In it current state, the vinebot can't easily be deployed from a distance, while the QinetiQ Dragon Runner 10 can be thrown into situations and can operate no matter which side it lands on. This makes tank robots in their current state the most portable. Soft snake robots obviously have lots of room to improve here, as these robots can also be designed to be thrown into situations or deployed from a distance due to their soft body being able to absorb impact. This is a common trend with snake robots in general, as these robots are relatively new and have a lot of potential.

Maneuverability is important in rescue situations where navigating terrain is a big factor. Robots that can manever through various obstacles will be able to search more effectively and will be less prone to getting stuck in a way which requires human assistance. Snake and rocker arm robots are the best in this category since they both have additional moving parts which allow them to navigate terrain. Rocker arm robots are better on rough and rocky terrain as they can use the rocker arm to get around terrain, while wheeled and tank robots do not have any moving parts to assist with this. Snake robots are best in tight caves or underground situations, as all three robot designs are slim and can easily make their way through narrow and twisty caverns. These two robots designs are best in these respective situations when it comes to maneuverabiltiy.

The cost of a search and rescue robot is important as it impacts various aspects of a search and rescue robot design. A high-costing robots will be more difficult to mass-produce, meaning it won’t be as readily available. Additionally, it will also impact the usefulness of these robots, as a robot that is very expensive must be effective in order to make up for the price spent. Finding a good cost to effectiveness ratio is important in order to ensure search and rescue robots are actually used. Two factors that impact cost the most are the size and amount of expensive parts used in the robot. Expensive parts include things such as motors and other components which require a lot of money to make. Low costing robots, therefore, should be small in order to avoid using too much material and should be simplistic so that moving parts are not used too much. Because of this, robots that are large and have many joints, such as rocker arm robots, segmented snake robots, and jointed snake robots will often cost more than the other designs. Tank and wheeled robots are the cheapest options, as their simplicity allows for a very cheap manufacturing process. An example of an ideal cheap robot is the Hector Darmstadt 2019 robot from the Robot Rescue League. This robot features a very small size with minimal parts, allowing the cost to build to be reduced to just 4,000 dollars (Table 1). This cost was much less than almost every other robot in the competition by a wide margin. Additionally, the team decided to build a larger rocker arm robot in 2022, and the cost of this robot was 30,000 dollars, far more than their tank robot from 2019 (Table 2). This shows how size and simplicity are major factors in the cost of the robot, allowing small tank and wheeled robots to be the cheapest option.

Complexity is important for search and rescue robots because it determines how well a design can be perfected and is closely related to cost. A complex robot will often lead to a more difficult manufacturing process and higher costs. However, these robots can also have additional features that make them better search and rescue robots. Finding a balance with complexity is difficult, as both simple and complex robots can succeed in different ways. This can be seen when comparing simple and complex robot designs such as the tank and snake designs. The tank robot design can be very effective as seen with the Howe & Howe Thermite S1 and QinetiQ Dragon Runner 10. Both of these designs are fairly simple but are good enough to be used in the field. Snake robot designs are good examples of complex designs, as they are still in the early stages and will take more time and experimentation to become fully developed. Due to their complexity, only prototypes of this design have been made, and none are ready for use on the field. However, their unique and complex design gives it many unique advantages and use cases, such as navigating tight and winding caves. Since less complex robot designs are easier to perfect, they are generally preferred over complex robots which are more difficult to perfect and are not as useful in the present. Because of this, designs including tank and wheeled robots are the best in terms of complexity as they have minimal subsystems and other moving parts involved in their designs.

Analyzing these categories gives important takeaways about each robot design and their strengths. However, these categories have made it clear that no one design can excel in all of these categories. This makes it difficult to choose one design that is better than the rest and should be used for all search and rescue purposes. Instead, a better approach would be to narrow down the number of designs and assign certain designs to different situations in which they perform best. Firstly, wheeled robots can be eliminated since they do not excel in many categories. The only strengths of this robot design are its complexity, cost, and flat ground speed. The simplicity of the wheeled robot design means it can be perfected and made very easily, but it also means there is little room for improvement, which will lead the wheeled robot design to become outpaced by other designs in the future. The flat ground speed of a wheeled robot is not useful in most search and rescue situations, as the ground is rarely flat, and search and rescue scenarios with simple flat terrain would find more use with drones which can scan a larger area faster than a ground robot. Because of this, the wheeled robot design does not have many areas in which it excels. The snake robot design is a must-use design because of the unique situations in which it is useful. No other robot design is small and flexible enough to wiggle through tight passages and caves the way a snake robot can. Additionally, the snake robot is only in its earlies stages, and will be improved a lot over the next few years as designs such as the vinebot become more robust. These unique abilities mean the snake robot design must be used in any underground situation involving small passageways or caves. Finally, for above-ground rescue missions involving rough terrain and hazardous conditions, the rocker arm robots and tank robots stand out as the best options. Both utilize the same drivetrain which allows for a lot of grip due to the tread, which also protects the drivetrain from the environment. The rocker arm provides additional mobility while sacrificing durability. The tank robot has a more durable system, but is less mobile due to the lack of an extra joint. Between the two, the rocker arm robot is the better option for a few reasons. Firstly, the durability issue that comes with rocker arm robots is something that can be improved. As research continues, new ways to manufacture the rocker arm joint will allow the robot to become much more durable. However, the maneuverability issue that comes with the tank robot design can't be fixed without modifying the design as a whole. Because of this, the rocker arm robot design is the better options since its flaws are fixable. In conclusion, the rocker arm and snake robots are the best options, and they should be used in search and rescue situations together, as their strengths provide different advantages depending on the situation.

Works Cited

1. Bryant, C. “How Search and Rescue Works.” *Mapquest,* https://www.mapquest.com/travel/pay-for-search-and-rescue.htm
2. Etzel, G. “Post-9/11 first responder deaths near total who died during attacks.” *Washington Examiner,* 2023, https://www.washingtonexaminer.com/news/2450928/post-9-11-first-responder-deaths-near-total-who-died-during-attacks/#google_vignette
3. Smith, E., Larkin, B., Holmes, L, et al. “20 years on, 9/11 responders are still sick and dying.” *Washington Examiner,* 2021, https://theconversation.com/20-years-on-9-11-responders-are-still-sick-and-dying-166033
4. Kleiner, K. “Robots to the Rescue?” *Slate*, 2023, https://slate.com/technology/2023/08/robot-search-and-rescue.html
5. Tae Ho Kim, Sang Ho Bae, Han, C.H., Hahn, B, et al. “The Design of a Low-Cost Sensing and Control Architecture for a Search and Rescue Assistant Robot”, *Proquest,* https://www.proquest.com/docview/2791667998?sourcetype=Scholarly%20Journals
6. Messina, E., Jacoff, A. “Performance Standards for Urban Search and Rescue Robots.” *National Institute of Standards and Technology*, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=822695
7. Huettenberger, G., Barth, K., Becker, K., et al. “RoboCup Rescue 2019 Team Description Paper Hector Darmstadt.” *RoboCup Rescue League TDP Collection*, https://tdp.robocup.org/wp-content/uploads/tdp/robocup/2019/robocuprescue-robot/hector-darmstadt-166/robocup-2019-robocuprescue-robot-hector-darmstadtL3gRETfkaU.pdf
8. Lin, X., Cardenas, I., Kanyok, N, et al. “RoboCup Rescue 2019 Team Description Paper ATR Team.” *RoboCup Rescue League TDP Collection*, https https://tdp.robocup.org/wp-content/uploads/tdp/robocup/2019/robocuprescue-robot/atr-kent-172/robocup-2019-robocuprescue-robot-atr-kent9qvmOUpaIM.pdf
9. Waterboro, M. “Howe & Howe Keeps Firefighters Safe Through Use of Thermite® Firefighting Robots.” *Howe & Howe,* https://www.howeandhowe.com/news-flash/articles/featured-news/howe-ho.we-keeps-firefighters-safe-through-use-thermite
10. Quick, D. “Dragon Runner 10 joins QinetiQ’s micro unmanned robot family.” *NewsAtlas,* https://newatlas.com/qinetiq-dragon-runner-10/19568/
11. Masato, M., Yuto, F., Shohei, I., et al. “RoboCup Rescue 2019 Team Description Paper SHINOBI.” *RoboCup Rescue League TDP Collection*, https://tdp.robocup.org/wp-content/uploads/tdp/robocup/2019/robocuprescue-robot/shinobi-163/robocup-2019-robocuprescue-robot-shinobihLyCQ2n8m0.pdf
12. Phunopas, A., Pudcheun, N., Blattler, A. et al. “RoboCup Rescue 2019 Team Description Paper iRAP SECHZIG.” *RoboCup Rescue League TDP Collection*, https://tdp.robocup.org/wp-content/uploads/tdp/robocup/2019/robocuprescue-robot/irap-sechzig-171/robocup-2019-robocuprescue-robot-irap-sechzigHCRQtdTZYU.pdf

13. iRobot 110 FirstLook Robot, *Army Technology,* 2017, https://www.army-technology.com/projects/irobot-110-firstlook-robot/?cf-view
14. 710 Kobra Multi-Mission Robot, *Army Technology,* 2015, https://www.army-technology.com/projects/irobot-710-kobra-multi-mission-robot/?cf-view
15. Zhu, S., Zhang, H., Lu, H., et al. "RoboCup Rescue 2019 Team Description Paper NuBot." *RoboCup Rescue League TDP Collection*, https://tdp.robocup.org/wp-content/uploads/tdp/robocup/2019/robocuprescue-robot/nubot-159/robocup-2019-robocuprescue-robot-nubot9SmrdWsEMT.pdf
16. Najafi, F., Bagheri, H., Hashemi, N.B., et al. "RoboCup Rescue 2019 Team Description Paper MRL.", *RoboCup Rescue League TDP Collection*, https://tdp.robocup.org/wp-content/uploads/tdp/robocup/2019/robocuprescue-robot/mrl-168/robocup-2019-robocuprescue-robot-mrlvTr5j0eCqb.pdf
17. Xu, Q., Shan, Z., Li, R., et al. "RoboCup Rescue 2019 Team Description Paper Mars-Rescue." *RoboCup Rescue League TDP Collection*, https://tdp.robocup.org/wp-content/uploads/tdp/robocup/2019/robocuprescue-robot/mars-rescue-165/robocup-2019-robocuprescue-robot-mars-rescueGxrHAVIupJ.pdf
18. Edlinger, R. "RoboCup Rescue 2022 Team Description Paper Team DYNAMICS." *RoboCup Rescue League TDP Collection*, https://tdp.robocup.org/wp-content/uploads/tdp/robocup/2022/robocuprescue-robot/team-dynamics-357/robocup-2022-robocuprescue-robot-team-dynamics74aUyQfnKO.pdf
19. Daun, K., Oehler, M., Schnaubelt, et al. "RoboCup Rescue 2022 Team Description Paper Hector Darmstadt." *RoboCup Rescue League TDP Collection*, https://tdp.robocup.org/wp-content/uploads/tdp/robocup/2022/robocuprescue-robot/hector-darmstadt-353/robocup-2022-robocuprescue-robot-hector-darmstadtC1JiORszUy.pdf
20. Reyes, M., Altamirano, F., Santiago, L., et al. "RoboCup Rescue 2019 Team Description Paper XFinder team." *RoboCup Rescue League TDP Collection*, https://tdp.robocup.org/wp-content/uploads/tdp/robocup/2019/robocuprescue-robot/xfinder-169/robocup-2019-robocuprescue-robot-xfindersSPeBnXyDi.pdf
21. Ackerman, E. "Robot Takes on Landmine Detection While Humans Stay Very Very Far Away." *Spectrum,* https://spectrum.ieee.org/husky-robot-takes-on-landmine-detection-while-humans-stay-very-very-far-away
22. Borenstein, J., Hansen, M., Borrell, A. et al. "The OmniTread OT-4 Serpentine Robot—Design and Performance", *University of Michigan,* https://deepblue.lib.umich.edu/bitstream/handle/2027.42/56171/20196_ftp.pdf
23. Spice, B. "Carnegie Mellon Snake Robot Winds Its Way Through Pipes, Vessels of Nuclear Power Plant." *Carnegie Mellon University,* https://www.cs.cmu.edu/news/2013/carnegie-mellon-snake-robot-winds-its-way-through-

pipes-vessels-uclear-power-plant#:~:text=The%20modular%20snake%20robot%20is,half%2Djoints%20on%20adjoining%20modules

24. Kubota, T. “Stanford researchers develop a new type of soft, growing robot.” *Stanford University,* https://news.stanford.edu/stories/2017/07/stanford-researchers-develop-new-type-soft-growing-robot

The Chemical Reaction that Fed the World - A Literature Review of the Haber Process and the History of Ammonia Production By Max Lee

Abstract

Nitrogen is essential for crop growth and thus has an indisputable role in agriculture. It is responsible for forming all the amino acids in plants, making proteins that aid the growth of crops. However, the most abundant form of natural nitrogen is diatomic gaseous nitrogen in the atmosphere, an unusable form for most plants and organisms. With the ever-increasing human population, especially in the late 19th to early 20th centuries, an inevitable lack of food was lurking. This literature review discusses Fritz Haber's attempts to artificially synthesize ammonia (NH_3), an extremely effective fertilizer, from hydrogen and nitrogen gas, and Haber's assistant's, Robert Le Rossignol, contraption that maintained a cyclic and continuous ammonia output. The review explains the methane steam reforming and dry air distillation processes from which hydrogen and nitrogen gas are obtained and the myriad of catalysts discovered and tried to make the reaction industrially viable. The review also looks at how fertilizers interact with soil and how they are taken by plants and assimilated into nitrogen-containing compounds. The Haber-Bosch Process provides seemingly endless food supplies to nurture the human population, allowing for the development and progression of technology into modern times.

Keywords: Haber Process, Ammonia Synthesis, Catalysts, Nitrogen Assimilation, Methane Steam Reforming, Air Distillation, WGSR, History of Haber Process, Fertilizers, Potash, Guano, Frank-Caro Process

Introduction

At the beginning of the 20th century, humanity was on the brink of mass starvation. The chemist William Crookes said that as the population multiplies, the food supply dwindles, and our agricultural production cannot keep up such strains for too long [1]. Since the 1800s, the population has increased from one billion to one and a half billion in just a century. This number increased by fivefold in the following 100 years [2]. The human population was growing exponentially, yet it was evident that the agricultural industry could not keep up with the same pace if provided with limited fertile soil. Thus, fertilizer needs have proportionally followed the rise of agricultural activity – especially the use of ammonia.

Ammonia is needed due to its nitrogen content, essential for efficient plant growth and increasing crop yield [3]. In the 21st century, fertilizers with nitrogen content production tallies to over 100 million metric tons, accounting for more than half of all fertilizer production [4]. Previous sources of ammonia or nitrogen included guano, potash, dry distillation of brown coal, or the Frank-Caro process, which all contained technological and natural resource limitations that hindered the response to the multiplying population [5]. Thus, an alternative, more efficient method of ammonia synthesis needed to be developed.

A clear candidate was ammonia synthesis from its elemental components since, according to German chemist Fritz Haber, plants cannot take in the abundant nitrogen gas in the air but need it to come in a form such as nitrate or ammonia, which it can absorb [5]. He suggested the formation of ammonia by combining its elemental components: hydrogen and nitrogen gas.

The synthesis of ammonia from gaseous nitrogen and hydrogen is a chemical reaction constituting an equilibrium. The forward reaction occurs when nitrogen and hydrogen gas (with a ratio of 1:3) form ammonia. This forward reaction has a favorable enthalpy, meaning energy is released from the system as the reaction proceeds [6]. However, the process is not as straightforward.

Producing an industrial-level yield of ammonia from the reaction between hydrogen and nitrogen gas is impossible with the technologies of the time. The reaction is compatible only at an extensive temperature and pressure to combat the difficulties regarding thermodynamic favorability and equilibrium [6]. A suitable catalyst is also needed to reduce the energy required for the reaction and the environmental requirements. The process, theoretically, required a system at 1000°C and 200 atmospheres [7]. That is equivalent to the heat of lava from a volcanic eruption and the pressure 2,000 meters into the ocean, enough to crush a human into a lump of flesh [8]. Generations of scientists have attempted to solve the issue but passed on the synthesis of ammonia from its elements to their successors.

Born in 1868, Fritz Haber was a German chemist who studied at Friedrich Wilhelm University. After experience working with chemical engineering, he approached the quest to synthesize ammonia from its gasses. His success in ammonia production revolutionized the world, and he was granted the Nobel Prize in 1918. As a patriotic German, his scientific contributions later extended into the battlefield, where he received the nickname “father of chemical warfare.” A decade after his death, his innovations would be used in the production of Zyklon B, the poison gas that killed millions of Jews [9].

Haber undeniably reshaped the world in many ways: what later became known as the Haber-Bosch process fed millions of people and was directly responsible for the billions of people alive today [10]. Ammonia is still one of the most dominant fertilizers in the world, being produced at over 150 million tons per year [11]. This literature review will explain the chemical concepts behind ammonia synthesis, analyze the techniques used during the discovery of Haber’s process, and give an overview of the consequences of the Haber-Bosch process.

1. The Synthesis of Ammonia from its Gasses

The Haber process is one of the most well-studied chemical equilibria [12]. A chemical equilibrium is a chemical reaction that is capable of being performed both “forward” (from left to right in a balanced equation) and “reversed” (from right to left). Le Chatelier's principle states that if a dynamic equilibrium is disturbed by changing the conditions, the position of equilibrium shifts to counteract the change to reestablish equilibrium. If a chemical reaction is at equilibrium and experiences a change in pressure, temperature, or concentration of products or reactants, the equilibrium shifts in the opposite direction to offset the change [13].

The Haber process (Eq. 1) releases energy, making the reaction exothermic [6]. If heat is added to the system without any other environmental changes, the reverse, or energy-demanding reaction, will be favored and occur at a greater rate.

$$N_2(g) + 3H_2(g) \leftrightharpoons 2NH_3(g) \qquad \text{(Eq.1)}$$

Other reactions, such as photosynthesis (Eq. 2), require energy, making it endothermic [14]. If additional heat energy is added to the system, the forward reaction will be favored because it fuels the reaction.

$$6CO_2(g) + 6H_2O(l) \rightarrow C_6H_{12}O_6(aq) + 6O_2(g) \qquad \text{(Eq.2)}$$

Furthermore, all non-spontaneous reactions have activation energy, which is the initial burst of energy input needed to break the intermolecular forces or bonds that hold the reactants together, forming unstable intermediates that have the potential to create the product. In organisms, enzymes dramatically increase the rate of biological reactions. For example, nitrogenase, responsible for breaking apart the tightly bonded diatomic nitrogen molecule, can cause a reaction that requires tremendous energy (such as a lightning strike) to occur [15].

Hence, enzymes are catalysts, substances that increase the rate of a chemical reaction without undergoing any permanent chemical change themselves. Many catalysts provide a surface for the reactants to bind to and assist in making and breaking bonds. This region is known as the active site (binding site) and is where the chemical reaction occurs [16]. Catalysts in the Haber process are adsorbent, meaning gas molecules attach to their surface and proceed with the reaction.

The ideal version of the Haber process would maximize ammonia output, be expandable to an industrial level, and have a suitable catalyst to optimize the process. The first step of the puzzle would be to gather the materials needed to start the reaction.

1.1 Obtaining the Reactants

The Haber Process involves the synthesis of ammonia from gaseous nitrogen and hydrogen. Both these reactants are abundant in nature. Our atmosphere comprises 78% nitrogen [17], and hydrogen is nearly omnipresent in plants, animals, and humans in the form of H_2O. However, when this reaction was first approached in the late 18th to early 19th century, the law of mass action and chemical equilibrium were not yet understood [18]. When available nitrogen and hydrogen gas were combined, ammonia formed only in minimal concentrations at low temperatures and decomposed at high temperatures [18], so many scientists believed the synthesis of ammonia from its elements was an insurmountable obstacle. A myriad of attempts to overcome this challenge followed, with efforts from eminent chemists and engineers at that time, including Fritz Haber, Carl Bosch, Alwin Mittasch, and later on Gerhard Ertl [18].

1.1.1 Nitrogen

Due to its stability, nitrogen is almost always found in its diatomic (N_2) form. Both atoms achieve a complete electron shell by forming a triple bond between the nitrogen atoms. The unreactive nature of the N_2 molecule is used in food processing, fire prevention, pressure testing, chemical blanketing, and many other industrial processes [19].

Nitrogen gas is mainly collected through liquid distillation of air. Air consists of gasses such as nitrogen (boiling point (bp): -196°C), oxygen (bp: -183°C), and argon (bp: -186°C), each with unique boiling points [21]. Liquid distillation first compresses all gasses into liquid states, then gradually heats the solution. According to their boiling points, the gasses will boil and evaporate separately (conveniently, nitrogen has the lowest boiling point and will boil first). The gas is then collected at the top of the distillation column [22].

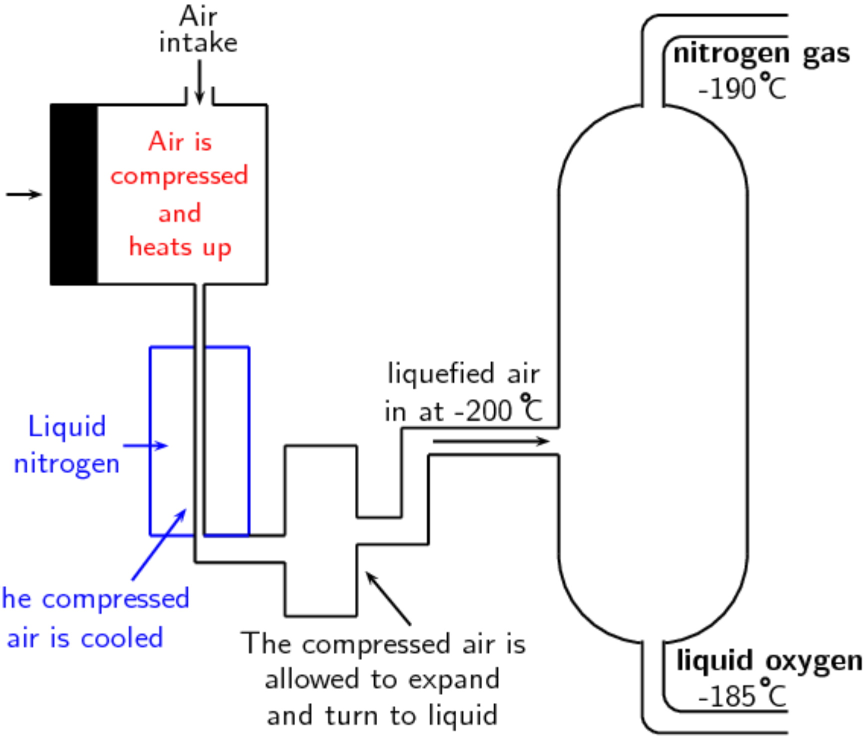


Fig. 1. Apparatus for Dry Air Distillation. Air enters the system as a gas. The first chamber involves a valve compressing the air, and the path below it cools it. The air swiftly turns into liquid in the middle chamber (where water and other liquids are often filtered) and moves into the third chamber, where nitrogen and oxygen is fractionated. From [20].

However, nitrogen is used naturally in all these cases and not as split nitrogen atoms because of the extraordinary energy input required to break the triple bond. This soon became a significant obstacle Haber faced when experimenting with the synthesis of ammonia, as the 941kJ needed to break apart one mole of nitrogen, which is usually only overcome by lightning bolts carrying gigajoules of energy, was much too extreme for simple laboratory setups to overcome [23]. The solution came by catalyzing the reaction at still extensive temperatures.

1.1.2 Hydrogen

What about hydrogen gas? Ammonia production alone accounts for over 50% of the total volume the hydrogen industry consumes [24]. Then, where do we obtain this irreplaceable fuel?

Water electrolysis (Eq. 3) is a chemical process involving reduction-oxidation reactions by moving an electrical current, which decomposes water into hydrogen and oxygen. Hydrogen ions in water are reduced, forming hydrogen gas, and the oxygen is oxidized to form oxygen gas. The amount of hydrogen to oxygen is the same as the original formula for water, where the molar ratio is 1:2, respectively, assuming a complete yield. Electrolysis plants operate at 70-90°C and consume 4-5 KWh/m^3 of hydrogen, obtained at a purity of 99.8% or more [25]. Because of its high energy consumption and substantial investment, water electrolysis is used for only 4% of world hydrogen production [25].

$$2H_2O(l) \rightarrow 2H_2(g) + O_2(g) \qquad \text{(Eq.3)}$$

Another alternative needed to be used: steam reforming of methane (Eq. 4). Methane steam reforming is the most common method of producing commercial bulk hydrogen, with the United States alone producing 9 million tons of hydrogen per year [26]. Natural hydrocarbons produced in agriculture, combustion of fossil fuels, and decomposition of landfill waste can now be used as a source of hydrogen.

$$CH_4(g) + H_2O(g) \rightarrow CO(g) + 3\,H_2(g) \qquad \text{(Eq.4)}$$

The reaction is industrially operated at a high temperature of around 800°C, over nickel-alumina-based catalysts to obtain a reasonable conversion of methane [27]. Methane is reacted with steam to produce carbon monoxide and hydrogen gas in an endothermic equilibrium reaction [27].

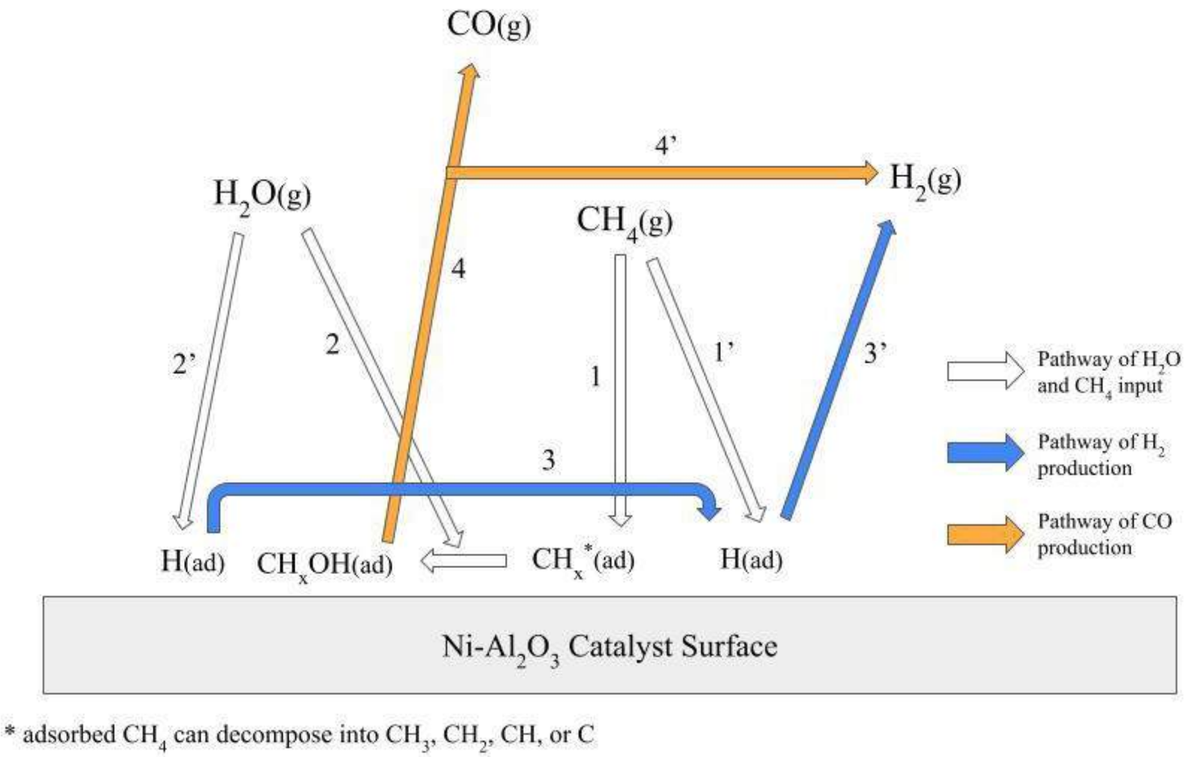

CO(g)
4'
$H_2(g)$
$H_2O(g)$
$CH_4(g)$
4
2'
2
1
1'
3'
3
H(ad)
$CH_xOH(ad)$
$CH_x^*(ad)$
H(ad)
$Ni-Al_2O_3$ Catalyst Surface
Pathway of H_2O and CH_4 input
Pathway of H_2 production
Pathway of CO production
* adsorbed CH_4 can decompose into CH_3, CH_2, CH, or C

Fig. 2. Mechanism of Methane Steam Reforming over nickel-alumina catalyst. 1, 1'. Methane is adsorbed onto the catalyst surface, decomposing into various CH_x species and lone hydrogens. 2, 2'. Water is adsorbed to form an alcohol with the CH_x species and lone hydrogens. 3, 3'. The lone hydrogens react to form H_2 gas. 4, 4'. The alcohol decomposes and forms CO and H_2 gas. Info. from [28]

To reduce the yield of the toxic carbon monoxide and further increase hydrogen production, the water gas shift reaction is coupled with steam reforming to maximize desired outputs.

The water gas shift reaction (WGSR) is the intermediate step for hydrogen production and CO reduction in the synthesis gas (Eq. 5). The use of WGSR exponentially increased with the advent of the Haber process and the development of relevant catalysts by Bosch in 1913 [24]. The catalyst consists of iron and chromium and could catalyze the reaction from 400°C to 500°C, reducing the exit carbon monoxide content to around 2% [29].

$$CO(g) + H_2O(g) \rightarrow CO_2(g) + H_2(g) \quad \text{(Eq.5)}$$

At lower temperatures, the iron-chromium catalysts lose their activity, and another alternative had to be developed for processes that demanded diminished temperatures. Copper-based catalysts operate at 2000°C and could achieve exit carbon monoxide concentrations of 0.1 to 0.3%. The iron-chromium and copper-based catalysts in the WGSR and methane steam reforming combine to produce the 20 million tons of hydrogen needed yearly for ammonia synthesis [30]. Nonetheless, methane reforming creates excessive carbon monoxide and attempts to reduce CO waste through WGSR, only to produce more CO_2, releasing greenhouse gasses into the atmosphere.

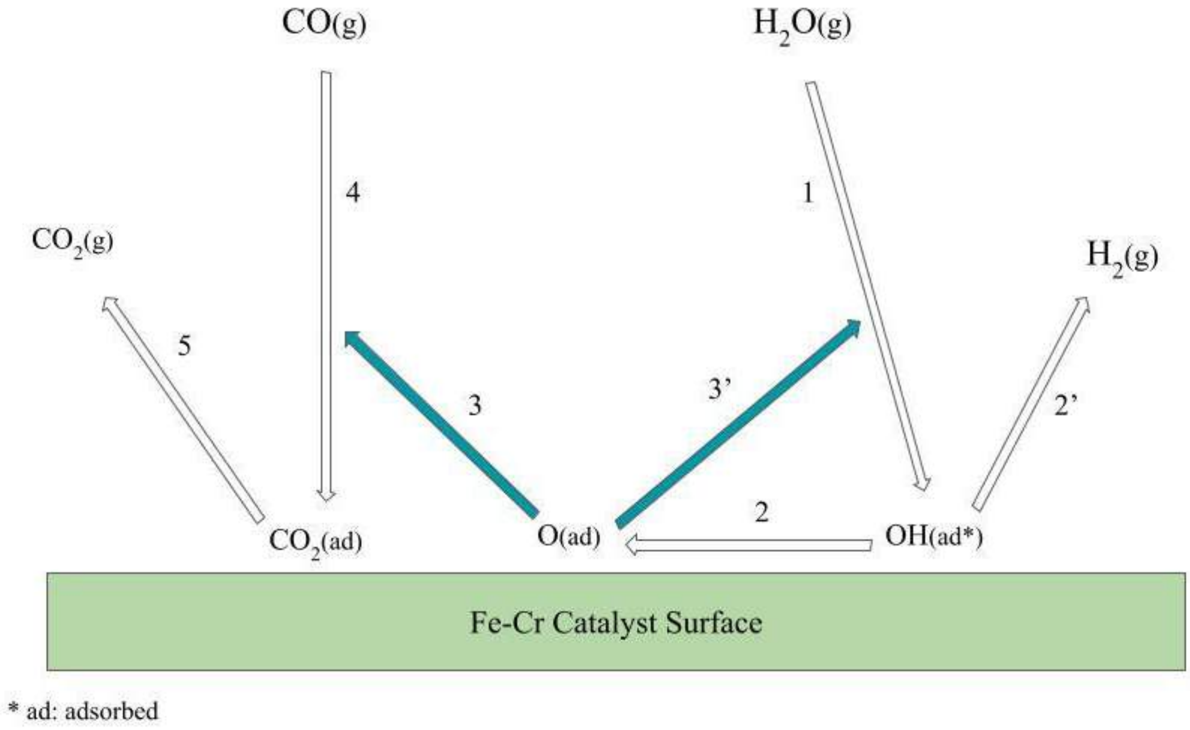


Fig. 3. Mechanism of Water Gas Shift Reaction over Iron-Chromium Catalyst. 1. Water is adsorbed onto the catalyst surface and catalyzed into hydroxides. 2, 2'. The hydroxides decompose into hydrogen at 2' and oxygen atoms at 2. 3, 3'. The oxygen is then used for the same reaction that produced it, and it also participates in the production of CO_2 at 4. 4. CO is adsorbed onto the catalyst surface and combines with the oxygen atom to form adsorbed CO_2. 5. CO_2 leaves the catalyst surface as a gas.

1.2 Finding a Catalyst

Now obtaining the needed reactants for ammonia production, experts sought a catalyst that made the Haber Process commercially viable. Similar to the catalysts in the steam reforming of methane and WGSR, gas molecules adsorb onto the surface of the catalyst, which helps "direct" and produce desired products.

The Haber process has three major steps: the dissociation of N_2, the dissociation of H_2, and the formation of N-H bonds. Catalysts in ammonia synthesis target the first and most tedious step. During Haber's initial successful trial in a laboratory setting, he discovered that an osmium or uranium-uranium carbide catalyst displayed excellent performance in ammonia synthesis [18]. Due to their costs, these choices soon proved unreasonable to reproduce at an industrial scale.

1.2.1 Magnetite-Alumina

Bosch and his assistant Mittasch were tasked with finding a more efficient and suitable catalyst for BASF (Baden Aniline and Soda Company), the leading proponent of industrializing the Haber process. Mittasch recognized that some metals showed little catalytic effect until combined with an additive. And, accidentally, their team stumbled upon an old iron sample infused with alumina, a small amount of calcium oxide, and potassium alkali, which proved to be the most effective catalyst so far [18]. They had found that “iron with a few percent of alumina and a pinch of potassium yielded a catalyst with acceptable reproducibility, performance and lifetime” [6]. Following the discovery in 1913, the iron was replaced with magnetite (Fe_3O_4), which was so effective that most manufacturers still use this catalyst today [18].

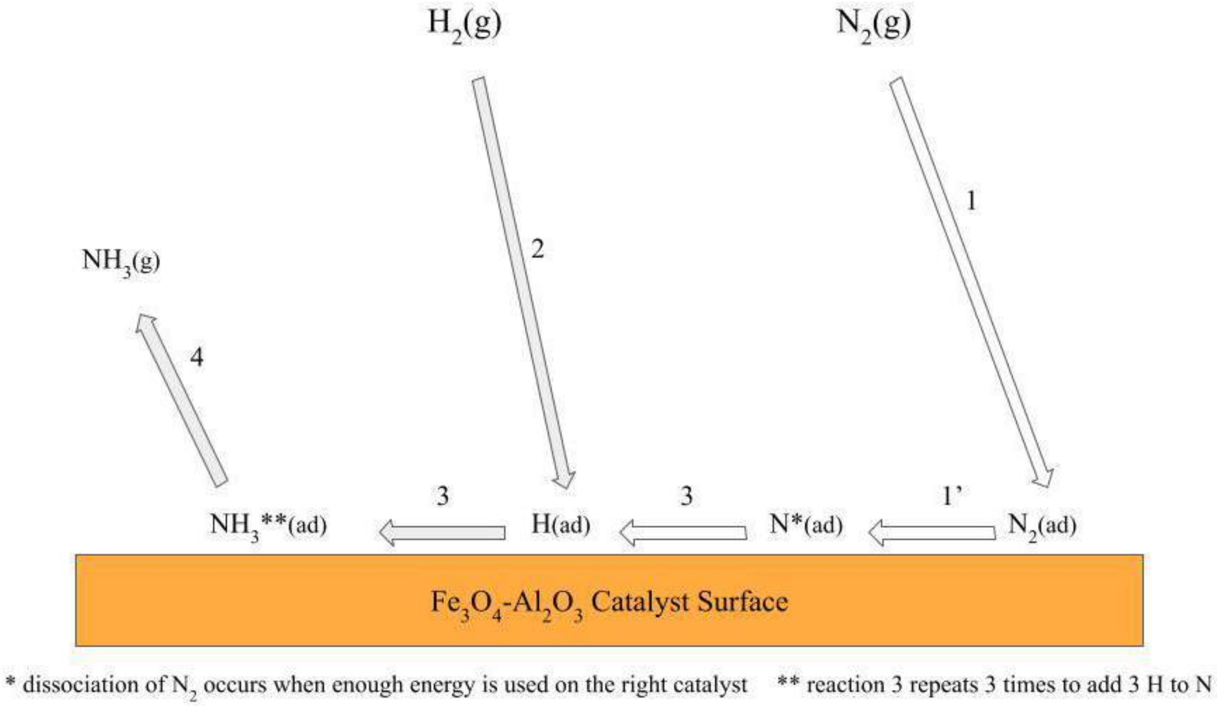

$H_2(g)$
$N_2(g)$
1
2
$NH_3(g)$
4
3
3
1’
NH_3**(ad)
H(ad)
N*(ad)
N_2(ad)
Fe_3O_4-Al_2O_3 Catalyst Surface
* dissociation of N_2 occurs when enough energy is used on the right catalyst
** reaction 3 repeats 3 times to add 3 H to N

Fig. 4. Mechanism of Ammonia Synthesis over Magnetite-Alumina Catalyst. 1, 1’. Nitrogen is adsorbed onto the catalyst surface and splits into the much more reactive nitrogen atom. 2. Hydrogen is adsorbed onto the catalyst surface. 3, 3’. Hydrogen reacts with the nitrogen atoms to form NH_x compounds, adding hydrogen atoms until ammonia is produced. 4. Ammonia leaves the catalyst surface as a gas.

1.2.2 Other Potential Catalysts - $Fe_{1-x}O$, Ru/C, and Co-Mo-N

Certain ameliorations and alternatives have been proposed and adopted, including the $Fe_{1-x}O$ catalyst. Invented in 1986, the $Fe_{1-x}O$ catalyst (which includes iron-based compounds such as FeO and Fe_2O_3), paired with aluminum, calcium, and potassium promoters, demonstrated improved performance compared to its alumina-based predecessor [18]. This marked the first improvement made in the iron-based catalyst system for the Haber process in nearly a century. Another is the ruthenium catalysts initially described in Mittasch's trials but revisited in 1972 and applied in 1992 by various countries and scientists. Using ruthenium as a base, potassium as a promoter, and carbon as a catalyst carrier, the Ru/C catalyst exhibited high activity for ammonia synthesis [18]. In two decades of refining, the catalyst reached industrial standards but lacked commercial appeal due to its high cost, facing the same fate as the original osmium catalyst. Not only that, but the Ru/C catalyst had a narrower optimal temperature range and shorter lifespan, leaving only 16 ammonia plants to use it in 2010 [18]. Finally, Ertl et al. theorized a cobalt-molybdenum nitride catalyst, indicating better performance than iron-based and Ru/C catalysts [18]. Molybdenum has a high affinity for dissociating nitrogen gas by forming nitrides. Thus, molybdenum can be used as an ammonia catalyst to form stable nitrides [31]. The adsorbed nitrogen and hydrogen atoms can directly react using these catalysts, producing ammonia more efficiently.

1.3 Completing the Reaction - Ammonia Synthesis

With the necessary reactants and catalysts, the only remaining obstacle is ammonia yield. To be scaled industrially, the Haber process needed to produce a large amount of ammonia that outweighed the energy expenses. Unfortunately, the original trial (before the osmium catalyst) ended with a negligible 0.005-0.012% ammonia yield, which proved far from viable in ammonia plants [32].

The invention of an innovative and exact "conical valve" by Haber's assistant Robert Le Rossignol and the unexpected efficiency of the uranium and osmium catalysts provided a much higher ammonia yield of 500-600g in an hour, helping the two obtain a contract with the BASF chemical engineering company [33]. The Haber process apparatus sought to fulfill two goals: to make the machine for ammonia synthesis continuous and to make the most ammonia as rapidly as possible.

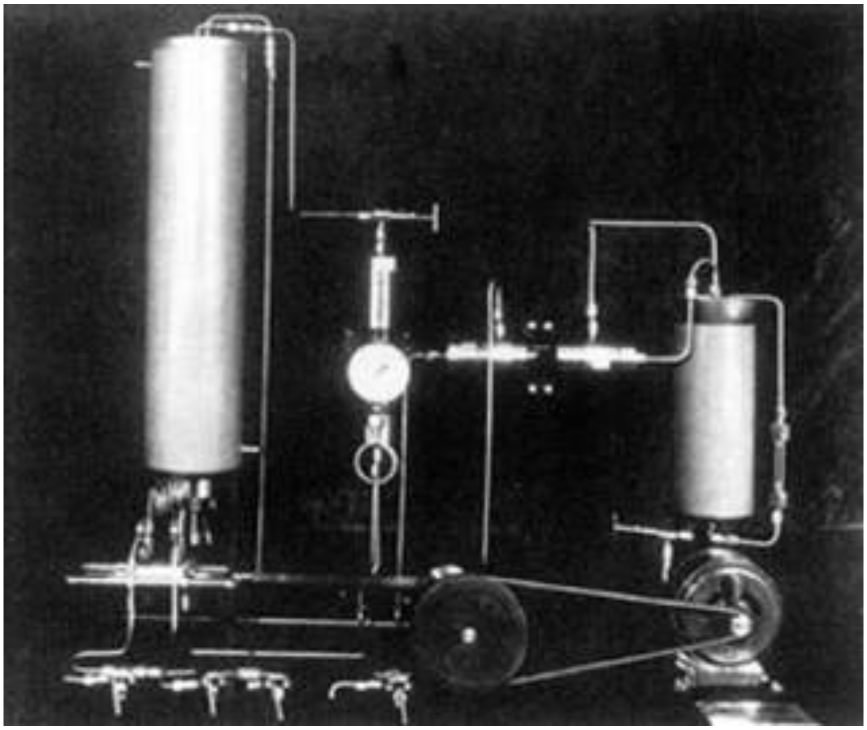


Fig. 5. The left cylinder contained the catalyst. The gas is funneled through the pipelines between the two cylinders (from the reaction chamber to the condenser and reversed). The right cylinder is the condenser, supposedly liquifying the ammonia produced. From [34].

To maximize the yield, the reaction occurred at 180-200 atm in Le Rossignol's experiments, shifting the equilibrium reaction towards ammonia as more moles of gas are on the reactant side. Additionally, temperatures of 500-600°C increased the rate of the reaction considerably [33]. However, because the forward reaction is an exothermic process, the added heat often hindered the forward reaction from being continuous. Thus, Le Rossignol thought of cooling and "removing the ammonia as a liquid, returning the 'remainder' gasses – replenished with a fresh charge of the mixture – to the catalyst" [33]. This ensures that the equilibrium is always shifted towards ammonia production (as its concentration will always be too low).

Knowing he has just experienced a historical milestone in chemical engineering history, Le Rossignol comically kept "two 6-inch sealed glass capillary tubes about two-thirds full of the first synthetic ammonia" produced that day for almost 44 years [33].

2. Approaches to Produce Ammonia Differently

The Haber process is efficient at producing ammonia. It dominates the fertilizer industry and has fed billions of people worldwide [10]. For these reasons, it has been an indispensable component of population growth in the last century, but it is far from the best alternative. Like the carbon dioxide produced in the WGSR (Section 1.1.2), the Haber process produces wasteful products that damage the environment. Hence, numerous other ammonia or nitrogen production methods have been studied and developed as fertilizer demand grows.

2.1 Alternative Methods for Ammonia Synthesis

Ammonia and other fertilizers have always been used as nutrient boosts to crop plants, increasing agricultural output. In the past, most fertilizers came from natural nitrogen sources and other nutrients like potassium. As natural resources are used up, a lack of fertilizer often occurs. Thus, more modern approaches to fertilizers sought to make mechanical processes more efficient (to feed the larger population) and more replenishable.

2.1.1 South American Natural Deposits

Seabird guano was an important fertilizer from the past, found in trace amounts dating back to the Late Intermediate Period (A.D. 1000-1476 in Peru) and other regions in Northern South America along the Western coast [35]. By tracking radioactive nitrogen isotopes, researchers found that seabirds along the Pacific coast consumed large amounts of nitrogen-rich fish, leaving behind guano of high nitrogen concentration [35]. This nitrogen was transferred onto human manures and finally into human remains. Guano also contained important nutrients like potassium and phosphate, also found in crops and humans of that time [35]. On a similar note, natural Chilean sodium nitrate rock deposits also serve a similar purpose in agriculture, leaving remains of nitrogen-rich crops in central Chile [36]. As European colonizers extended their reach onto the Western coast of South America, they also grew interested in these natural resources, leading to an exponential increase in their extraction and export back to Europe [37]. This inevitably fueled the ever-increasing yet unstable population, and as the supply of sodium nitrate and guano deposits diminished, the world was left in a difficult position.

2.1.2 Frank-Caro Process

In the late 19th century, Adolph Frank and Nikodem Caro engineered a mechanism that used calcium carbide to fix nitrogen, separating the diatomic molecule into individual atoms as cyanamide (Eq. 6) [38].

$$CaC_2(s) + N_2(g) \rightarrow CaCN_2(s) + C(s) \qquad \text{(Eq.6)}$$

In 1905, the Frank-Caro process was further extrapolated by turning the produced cyanamide into ammonia through steam reforming (Eq. 7). Nonetheless, the process was extremely energy intensive, reaching temperatures over 1000°C, making it much too expensive for mass production in plants [39]. Hence, the Haber process swiftly overshadowed it in the following years.

$$CaCN_2(s) + 3H_2O(g) \rightarrow CaCO_3(s) + 2NH_3(g) \qquad \text{(Eq.7)}$$

3. Nitrogen Assimilation in Plants

Plants have a “fundamental dependence” on inorganic nitrogen, as seen in the 85–90 million metric tonnes of nitrogenous fertilizers in 2004, which only increased to 196 million metric tonnes in 2019 [40, 41, 42]. Most fertilizers are ammonia, produced through the previously discussed Haber process that feeds the world.

Nitrogen, in the form of amino acids, is the building block for proteins, molecules responsible for catalyzing chemical reactions and structural support, and chlorophyll, the main pigment responsible for photosynthesis [43, 44]. Additionally, nitrogen has a crucial role in the development of plants, promoting leaf, stem, and root growth [43].

3.1 Nitrogenase and Nitrogen Fixing Bacteria

The bacterial enzyme nitrogenase can reduce atmospheric N_2 into ammonia. It consists of the catalytic molybdenum-iron protein and its specific reductase, the iron protein [45].

Nitrogen-fixing bacteria were first discovered in 1886 when certain legumes were found to use nitrogen gas from the atmosphere directly, which was later found to be due to the *Rhizobium leguminosarum*, a strain of bacteria that uses nitrogenase enzymes to convert nitrogen gas into ammonia (Eq. 8) [46]. Since then, attempts to apply nitrogen-fixing bacteria to crops have demonstrated “[stimulated] growth of crops [...] and improved yield of vegetables” [3]. However, the bacteria’s nitrogen-fixing efficiency is insufficient to meet the ammonia demand. Nevertheless, researchers believe that genetic engineering and modification, higher potency bacteria, or the direct modification of nitrogenase production in plant genes could solve the issue [3].

$$N_2(g) + 8H^+ + 8e^- \rightarrow 2NH_3(g) + H_2(g) \qquad \text{(Eq.8)}$$

3.2 Nitrogen Uptake in Plants

However, a question arises: Why do plants need ammonia? Although abundant diatomic nitrogen (N_2) is too tightly bonded for non-specialized plants to absorb, why not use another nitrogen-containing compound? This is because ammonia (ammonium, in most soils) is the most stable and suitable compound [43]. As the ammonia from the Haber process is implemented into the soil, it readily reacts with water to form ammonium (NH_4^+), the best nitrogen carrier for plant growth stimulation [47].

Ammonium is absorbed into the plant through the roots, collecting it from the nearby soil. It then directly enters the glutamine (the most abundant amino acid in plants and animals) synthetase (GS) cycle, where inorganic nitrogen is converted into organic nitrogen as it reacts with glutamate [48]. Nitrates, the other form of nitrogen absorbed by plants, are reduced into nitrite by nitrate reductase enzymes in the cytoplasm of the shoot, then into ammonium through plastidic nitrite reductase enzymes, finally joining the GS cycle [48].

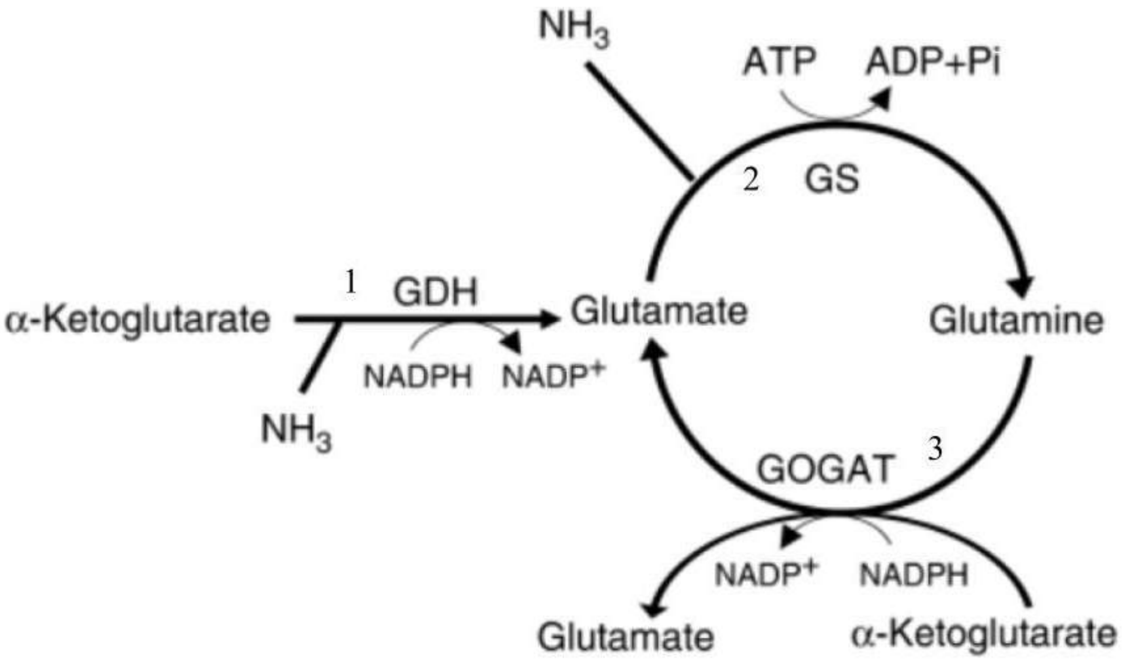


Fig. 6. Glutamine Synthetase Cycle and Ammonia Assimilation in Plant Systems. Ammonia can enter from two places in the cycle: 1. Conversion of α-ketoglutarate to glutamate by the glutamate dehydrogenase (GDH) enzyme; 2. Conversion of glutamate to glutamine by the glutamine synthetase (GS) enzyme. 3. Glutamine can be converted back into glutamate by the Glutamine oxoglutarate aminotransferase (GOGAT) enzyme, where the cycle continues. From [49]

Glutamate can then be used for other biosynthetic processes, donating nitrogen to form other amino acids and nucleotides [50]. Its nitrogen is responsible for part of the nitrogen in purines and pyrimidines and all the nitrogen in the other amino acids [50]. Hence, glutamate is vital for nearly all organisms and is found at much higher concentrations than any other amino acid [50]. For example, *Escherichia coli*'s concentration reaches two orders of magnitude greater than the second most abundant amino acid, leucine [50]. Thus, the nitrogen carried in ammonia is required for the development of plants and aids in the increased growth of crops in agriculture, feeding everyone on the planet.

Conclusion

A need for fertilizers grew as the population increased, demanding an industrial and efficient method of synthesizing the basic nutrients essential for plant growth. The obsolete natural fertilizers such as seabird guano and Chilean nitrate rocks were drained, and the energy-intensive Frank-Caro process did not solve the problem either. Ammonium and nitrates had to be artificially synthesized to provide food for the people, so the journey to synthesize ammonia from hydrogen and nitrogen gas began. Collecting and distilling the diatomic nitrogen from the air provided N_2. Natural gas steam reforming and the water gas shift reaction supplied H_2. With the carefully selected alumina-infused magnetite catalyst, the firm triple bond from N_2 broke, and ammonia can successfully be created in a lab. With Le Rossignol's original sketch of the machinery, Bosch extended it into factories and plants, eventually pumping out millions of tons of ammonia yearly [11, 33].

The Haber-Bosch process is a complex and ingenious chemical engineering feat invented by the brilliant Fritz Haber, designed by his assistant Robert Le Rossignol, and scaled up and industrialized by Carl Bosch [18]. Then, countless scientists who followed their predecessors built upon and improved it. Whether the contraption proceeded with a magnetite catalyst or a molybdenum catalyst, whether it obtained hydrogen gas from electrolysis, water gas shift reaction, or methane steam reforming, the Haber-Bosch process undoubtedly changed the world.

Works Cited

[1] Crookes, William, and C. Wood Davis. *The Wheat Problem.* 1899, 20-49.

[2] Jielin, Dong, et al. "How Does Technology and Population Progress Relate? An Empirical Study of the Last 10,000 Years." *Technological Forecasting and Social Change* 103 (2016): 57-70. https://www.sciencedirect.com/science/article/abs/pii/S0040162515003455.

[3] Zhang, Wenyao, et al. "Molecular Mechanism and Agricultural Application of the NifA–NifL System for Nitrogen Fixation." *International Journal of Molecular Sciences* 24 (2023): 907, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9866876.

[4] "Global Fertilizer Production by Nutrient." Statista, July 23, 2023. www.statista.com/statistics/1290786/global-fertilizer-production-by-nutrient/. Accessed 16 June 2024.

[5] Haber, Fritz. "The synthesis of ammonia from its elements." *Nobel lecture* 2 (1920): 1, https://www.ias.ac.in/public/Volumes/reso/007/09/0086-0094.pdf.

[6] Modak, Jayant M. "Haber Process for Ammonia Synthesis." *Resonance* 16 (2011): 1159–1167, https://doi.org/10.1007/s12045-011-0130-0.

[7] National Ocean Service. "NOAA's National Ocean Service: Ocean Facts." Noaa.gov, 2019, oceanservice.noaa.gov/facts/. Accessed 12 June 2024.

[8] Britannica. "Lava | Definition, Types, Composition, & Facts | Britannica." *Encyclopædia Britannica*, 2020, www.britannica.com/science/lava-volcanic-ejecta.

[9] Gal, Joseph. "Remembering Fritz Haber in the year 2015." *L'Actualité chimique* 397-398 (2015): 114-121, https://new.societechimiquedefrance.fr/wp-content/uploads/2019/12/2015-397-398-juin-juillet-p109-gal_hd.pdf.

[10] Harford, Tim. "How Fertiliser Helped Feed the World." BBC News, 2 Jan. 2017, www.bbc.com/news/business-38305504. Accessed 15 June 2024.

[11] Pattabathula, Venkat, and Jim Richardson. "Introduction to ammonia production." *Chem. Eng. Prog* 112 (2016): 69-75, https://www.aiche.org/sites/default/files/cep/20160969.pdf.

[12] Chen, Shiming, et al. "Chapter 2 - Electrochemical Dinitrogen Activation: To Find a Sustainable Way to Produce Ammonia." *Studies in Surface Science and Catalysis* 178 (2019): 31-46, www.sciencedirect.com/science/article/abs/pii/B9780444641274000021.

[13] Denbigh, Kenneth. *The Principles of Chemical Equilibrium with Applications in Chemistry and Chemical Engineering.* Cambridge, Cambridge University Press, 2002, 82-100.

[14] David Oakley Hall, and J M O Scurlock. *Photosynthesis and Production in a Changing Environment: A Field and Laboratory Manual*. London, Chapman & Hall, 1995, 59-64.

[15] Kim, Jongsun, and Douglas C. Rees. "Nitrogenase and Biological Nitrogen Fixation." *Biochemistry* 33 (1994): 389–397, https://doi.org/10.1021/bi00168a001.

[16] Whitaker, J. R. "ENZYMES | Functions and Characteristics." *Encyclopedia of Food Sciences and Nutrition*, 2003, www.sciencedirect.com/science/article/abs/pii/B012227055X004193.

[17] NASA. “10 Interesting Things about Air.” Climate Change: Vital Signs of the Planet, climate.nasa.gov/news/2491/10-interesting-things-about-air. Accessed 23 July 2024.

[18] Liu, Huazhang. “Ammonia Synthesis Catalyst 100 Years: Practice, Enlightenment and Challenge.” *Chinese Journal of Catalysis* 35 (2014): 1619–1640, https://doi.org/10.1016/s1872-2067(14)60118-2.

[19] “Omega Air.” Nitrogen Gas Applications | Omega Air | Air and Gas Treatment, www.omega-air.si/news/news/nitrogen-gas-applications. Accessed 15 June 2024.

[20] Nigerian Scholars. “Obtaining nitrogen: Fractional distillation of liquefied air.” https://nigerianscholars.com/tutorials/chemistry-real-world/obtaining-nitrogen/. Accessed 23 July 2024.

[21] Linde. “Gas School.” https://www.linde-gas.no/en/products_ren/gas_school/index.html. Accessed 23 July 2024.

[22] Britannica. “Distillation | Chemical Process.” *Encyclopædia Britannica*, 2019, www.britannica.com/science/distillation.

[23] “Periodic Table | Bond Energies.” Webelements, https://winter.group.shef.ac.uk/webelements/nitrogen/compound_properties.html. Accessed 20 July 2024.

[24] Baraj, Erlisa, et al. “The Water Gas Shift Reaction: Catalysts and Reaction Mechanism.” *Fuel* 288 (2020): 119817, https://doi.org/10.1016/j.fuel.2020.119817.

[25] Zoulias, Emmanuel, et al. "A review on water electrolysis." *Turkish Journal of Science and Technology* 4 (2004): 41-71. https://www.academia.edu/download/32222925/A_REVIEW_ON_WATER_ELECTROLYSIS.pdf.

[26] Gilani, Haris R., and Sanchez, Daniel L. “Introduction to the Hydrogen Market in California.” University of California-Berkeley, Sep. 2020, https://bof.fire.ca.gov/media/10190/introduction-to-the-hydrogen-market-in-california-draft-for-comment_ada.pdf. Accessed 21 July 2024.

[27] Maier, L., et al. “Steam Reforming of Methane over Nickel: Development of a Multi-Step Surface Reaction Mechanism.” *Topics in Catalysis* 54 (2011): 845–858, https://doi.org/10.1007/s11244-011-9702-1.

[28] Cai, Lei, et al. “Study on the Reaction Pathways of Steam Methane Reforming for H2 Production.” *Energy* 207 (2020): 118296, https://doi.org/10.1016/j.energy.2020.118296.

[29] Smith R J, Byron, Loganathan, Muruganandam and Shantha, Murthy Shekhar. "A Review of the Water Gas Shift Reaction Kinetics." *International Journal of Chemical Reactor Engineering* 8 (2010): 1-32. https://doi.org/10.2202/1542-6580.2238.

[30] Nallapaneni, Anuraag, and Sood, Shaifali. “Emission Reduction Potential of Green Hydrogen in Ammonia Synthesis.” Wri India, 4 Oct. 2022, https://wri-india.org/blog/emission-reduction-potential-green-hydrogen-ammonia-synthesis-fertilizer-industry. Accessed 17 July 2024.

[31] Kojima, Ryoichi, and Ken-ichi Aika. “Cobalt Molybdenum Bimetallic Nitride Catalysts for Ammonia Synthesis.” *Applied Catalysis A: General* 215 (2001): 149–160, https://doi.org/10.1016/s0926-860x(01)00529-4.

[32] Sheppard, Deri. *Robert Le Rossignol: Engineer of the Haber Process.* Springer Nature, 2020, 49-66.

[33] Sheppard, Deri. “Robert Le Rossignol, 1884–1976: Engineer of the Haber Process.” *Notes and Records: The Royal Society Journal of the History of Science* 71 (2017): 263–296, https://doi.org/10.1098/rsnr.2016.0019.

[34] Henning, Eckart, and Kazemi, Marion. *Dahlem, Domain of Science*. 2009, 68-83.

[35] Santana-Sagredo, Francisca, et al. “‘White gold’ Guano fertilizer drove agricultural intensification in the Atacama Desert from AD 1000." *Nature Plants* 7 (2021): 152-158. https://ora.ox.ac.uk/objects/uuid:091c9984-4b4b-41f2-b038-3f385b895824/files/s1j92g799z

[36] Ortega-Blu, Rodrigo, et al. “Nitrate Concentration in Leafy Vegetables from the Central Zone of Chile: Sources and Environmental Factors.” *Journal of Soil Science and Plant Nutrition* 20 (2020): 964–972, https://doi.org/10.1007/s42729-020-00183-4.

[37] Dunikowska, Magda, and Turko,Ludwik. “Fritz Haber: The Damned Scientist.” *Chem. Int. Ed.* 50 (2011): 10050 - 10062. https://doi.org/10.48550/arXiv.1112.0949

[38] MIT. “The Nitrogen Cycle.” https://web.mit.edu/5.03/www/readings/nitrogen/nitrogen.pdf

[39] Johnson, Benjamin. “The State of Ammonia Synthesis at the Turn of the Twentieth Century: The Arena for Discovery.” *Making Ammonia* (2012): 77–89, https://doi.org/10.1007/978-3-030-85532-1_9.

[40] Masclaux-Daubresse, Céline, et al. “Nitrogen Uptake, Assimilation and Remobilization in Plants: Challenges for Sustainable and Productive Agriculture.” *Annals of Botany* 105 (2010): 1141–1157, https://doi.org/10.1093/aob/mcq028.

[41] Good, Allen G., et al. “Can Less Yield More? Is reducing nutrient input into the environment compatible with maintaining crop production.” *Trends in Plant Science* 9 (2004): 597-605. https://doi.org/10.1016/j.tplants.2004.10.008.

[42] Ilinova, Alina, Diana Dmitrieva, and Andrzej Kraslawski. "Influence of COVID-19 pandemic on fertilizer companies: The role of competitive advantages." *Resources Policy* 71 (2021): 102019. https://www.researchgate.net/publication/349442980_Influence_of_COVID-19_pandemic_on_fertilizer_companies_The_role_of_competitive_advantages.

[43] Leghari, Shah Jahan, et al. "Role of nitrogen for plant growth and development: a review." *Advances in Environmental Biology* 10 (2016): 209-216. link.gale.com/apps/doc/A472372583/AONE?u=anon~d510777b&sid=googleScholar&xid=4820fb6a.

[44] Rasheed, Faiza, et al. “Modeling to Understand Plant Protein Structure-Function Relationships—Implications for Seed Storage Proteins.” *Molecules* 25 (2020): 873, https://doi.org/10.3390/molecules25040873.

[45] Owens, Cedric P., and Faik A. Tezcan. “Chapter Twelve - Conformationally Gated Electron Transfer in Nitrogenase. Isolation, Purification, and Characterization of Nitrogenase from Gluconacetobacter Diazotrophicus.” *Methods in Enzymology* 599 (2018): 355-386, www.sciencedirect.com/science/article/abs/pii/S0076687917303373.

[46] Soumare, Abdoulaye, et al. “Exploiting Biological Nitrogen Fixation: A Route towards a Sustainable Agriculture.” *Plants* 9 (2020): 1011, https://doi.org/10.3390/plants9081011.

[47] Iowa State University. “Understanding Anhydrous Ammonia Application in Soil | Integrated Crop Management.” 15 March 2019, Crops.extension.iastate.edu/cropnews/2019/03/understanding-anhydrous-ammonia-application-soil. Accessed 20 June 2024.

[48] Liu, Xiujie, et al. “Nitrogen Assimilation in Plants: Current Status and Future Prospects.” *Journal of Genetics and Genomics* 49 (2021): 394–404, https://doi.org/10.1016/j.jgg.2021.12.006.

[49] Yuan, Jie, et al. "Metabolomics‐driven quantitative analysis of ammonia assimilation in *E. coli.*" *Molecular Systems Biology* 5 (2009): 302. https://www.researchgate.net/publication/26751335_Metabolomics-driven_quantitative_analysis_of_ammonia_assimilation_in_E_coli

[50] Walker, Mark C., and Wilfred A. van der Donk. “The Many Roles of Glutamate in Metabolism.” *Journal of Industrial Microbiology & Biotechnology* 43 (2015): 419–430, https://doi.org/10.1007/s10295-015-1665-y.

Testing Extended Contact: A Prejudice Reduction Experiment Among Indians in India and the US By Krishnni Khanna

Abstract

Inter-group prejudice remains a pervasive social issue transcending boundaries of culture and ethnicity. Research on social psychology posits various modes of intergroup contact as potentially working to reduce prejudice across groups. In this paper, we test how imagined contact affects prejudice against Muslims among a sample of both Indian expatriates in the United States and individuals residing in India. Employing a rigorous experimental design, we replicate previous methodologies to assess the efficacy of imagined contact in prejudice reduction. Our sample is 42 Indian respondents, and we administer an online survey with a randomized experimental component to them. We find that imagined contact works to reduce prejudice only in the India sample, but not the US sample. This suggests that imagined contact may only lead to a reduction in prejudice levels among individuals who perceive intergroup contact as a plausible scenario, given their local contexts. This study contributes to a deeper understanding of the mechanisms underlying prejudice reduction and underscores the importance of contextual factors in shaping the efficacy of intergroup contact strategies.

Keywords

Behavioral and social sciences; Sociology and social psychology; prejudice reduction; intergroup contact; survey experiment

Introduction

In this study, we pose the question: How does imagined intergroup contact influence attitudes towards Muslims among Hindu Indians residing in both the US and India? Our goal is to understand how imagining interactions with Muslims influences attitudes of Indian Hindus towards Muslims, both in the United States and India.

To achieve this, the paper will first explore the concept of prejudice, its definition and significance in modern society. Then, the paper will focus on methodologies for measuring prejudice, followed by examining the contact hypothesis and its theories. However, we also assess the challenges associated with this hypothesis. Subsequently, we also detail the design and methodology of our own survey, presenting the top-line results and implications of our findings. Through the results of our own survey, we find that imagined intergroup contact is effective at reducing prejudice in situations where there are opportunities for such imagined contact to actually take place.

Literature Review

In this section, we provide a brief overview of research on prejudice and prejudice reduction. First, we define prejudice and discuss why it is useful to study. We then review literature theorizing and evaluating the contact hypothesis

Prejudice

Prejudice refers to a negative attitude towards individuals solely based on their membership in a particular community or group.[1] It is a prevalent social issue that has significant impacts on intergroup relations, societal harmony, and individual well-being. Understanding prejudice is crucial because it heavily influences the behavior and attitudes of those harboring prejudiced beliefs.[2]

The study of prejudice is motivated by its detrimental effects on individuals and society. Prejudice can lead to discrimination, social exclusion, and even violence against marginalized groups.[3] Additionally, prejudice hinders efforts towards social justice and harmony. Therefore, researchers aim to uncover the underlying mechanisms of prejudice formation and maintenance to develop effective strategies for reducing the prevalence and impact of prejudice.[4]

Various methods are implemented to measure prejudice, ranging from self-report questionnaires to implicit measures such as the Implicit Association Test (IAT).[1] Self-report questionnaires involve participants reporting their attitudes and beliefs directly, while implicit measures assess unconscious or automatic associations between concepts in one's mind. The choice of measurement method depends on the research objectives and the aspects of prejudice being investigated. For example, explicit measures may capture conscious beliefs, while implicit measures provide insights into underlying biases that individuals may be unaware of or unwilling to admit.[5]

Intergroup contact

The contact hypothesis is one of the most widely studied methods of prejudice.[2] The contact hypothesis suggests that intergroup contact can reduce prejudice under certain conditions.[1] According to this theory, contact between members of different social groups can lead to improved intergroup relations. Improvements include equal status between groups, common goals, cooperation, and support from authorities. The idea is that direct interactions between individuals from different groups can challenge stereotypes, increase empathy, and foster positive attitudes towards the outgroup.

Empirical evidence supporting the contact hypothesis comes from various sources, such as surveys and experiments. Studies have found that individuals who report more frequent contact with members of different social groups tend to have more positive attitudes towards those groups.[2,6] Experimental research has also demonstrated the effectiveness of intergroup contact in reducing prejudice. A recent study on inter-caste prejudice among Indian men showed that being randomly assigned to play in a cricket team with members of other castes reduced out-group prejudice and in caste-favoritism.[7] Another study finds that interparty contact among Democrats and Republicans can reduce out-party hostility.[8]

Research on mechanisms of inter-group contact theory finds that contact reduces prejudice through three main pathways: increasing knowledge about the outgroup, reducing feelings of anxiety from the outgroup, and increasing empathy.[9] However, interventions involving real intergroup contact are usually costly and resource-heavy, resulting in research on alternative methods.

Beyond real contact, researchers have examined extended or imagined contact. The extended contact hypothesis builds upon the contact hypothesis by suggesting that knowledge of ingroup members having positive relationships with outgroup members can also reduce prejudice.[10] In other words, even indirect or extended forms of contact, such as knowing someone who has a friend from a different group, can promote positive intergroup attitudes. This hypothesis is useful because it provides an additional method for prejudice reduction, especially in situations where direct intergroup contact may be limited or impractical.

Building upon the contact hypothesis, Turner and Crisp conducted an experiment to investigate the extended contact hypothesis, particularly focusing on intergroup contact with Muslims.[5] In their study, Crisp and Turner explored the impact of extended contact, which refers to knowledge of ingroup members having positive relationships with outgroup members, on intergroup attitudes towards Muslims, as well as imagined contact. Another meta-analytic study showed that imagined contact was indeed effective at reducing intergroup prejudice, but the effects were stronger for children than for adults.[11] However, while imagined contact has high potential for reducing prejudice, research finds that an understanding of context and norms within the groups is necessary for interventions to success.[12]

However, not all empirical evidence consistently supports the contact hypothesis. A review study discovered that while contact can reduce prejudice, the effects of intergroup contact are weaker in context to racial or ethnic prejudice.[13] Enos conducted a randomized controlled trial to
test the causal effects of intergroup contact on exclusionary attitudes towards immigrants.[3]

Contrary to expectations, the study found that repeated intergroup contact actually led to a shift towards more exclusionary attitudes among participants. The effectiveness of intergroup contact in reducing prejudice may depend on contextual factors, such as the nature of the contact and the sociopolitical context in which it occurs.

As this review of the literature shows, empirical research can be inconclusive around the effects of intergroup contact on prejudice. This ambiguity is especially significant when examining ethnic or racial prejudice in adult populations.[3] To build on this gap, our research addresses adult ethnic prejudices among a sample of Indian respondents.

Methods

This paper aims to replicate the Turner and Crisp study within a different context, focusing on Indians in the US and in India.[5]

In 2010, Turner and Crisp ran two studies to estimate the effect of imagined intergroup contact on prejudice in a sample of British undergraduates.[5] Their first study examined prejudice against the elderly and their second study examined prejudice against Muslims. Turner and Crisp's second study - which we now replicate in a different sample - found that imagining intergroup contact with Muslims led to more positive perspectives on Muslim people relative to not imagining intergroup contact on them. Our study replicated the experimental procedure of the Crisp and Turner study on a group of Hindu Indians in the US and India.

Why replicate?

India has a diverse and pluralistic society where Hindus and Muslims constitute two major religious communities - Hindus constitute the majority while Muslims make up less than twenty percent. Over the centuries, interactions between these communities have been shaped by a complex historical narrative that includes periods of coexistence, but also instances of tension and conflict. Muslims are often subjected to prejudice in economic, social, and political setting from the Hindu community.[14] By replicating the Turner and Crisp design on reducing prejudice against Muslims in the Indian context, the research aims to shed light on how these tactics for prejudice reduction would work in the Indian context.

Study Methodology

Similar to Turner and Crisp's study, we ran a randomized survey experiment among a convenience sample of Indian adults. We used Qualtrics to administer the survey electronically. The survey embedded an experiment almost identical to the one used by Turner and Crisp.[5]

Sample

We used a convenience sample to reach Indians in the US and India. This is because of resource and logistical constraints on our end. However, across both US and India, we used similar tactics of snowball sampling through organizations that the author was a part of - as a result, we expect that samples to be similar across both countries. We acknowledge the limited external validity of such a sample but believe our results provide an important starting point to answer the question. To ensure our sample only included individuals identifying as Indian, we included a screening question upfront. The screener would result in the survey being administered only if the respondent checked a box identifying as Indian. If they did not identify as Indian, the survey would end and give them a message thanking them for their time.

Survey Design

Respondents were randomly allocated into two different groups: treatment and control. Both groups were initially asked to provide demographic information, including their age, location, and sex, to help categorize and analyze responses within specific demographic groups.

Both groups were asked to spend the next 2 minutes imagining themselves meeting someone who is a Muslim for the first time - but the specifics of this task were different across the two groups. In an exact replication of Crisp and Turner, the experimental group was asked to imagine a relaxed, positive, and comfortable interaction and to write down a couple lines about this imagined interaction.[5] The control group was simply asked to imagine meeting a Muslim individual without specifying the nature of the interaction or asking them to reflect on the interaction.

Finally, both groups were then asked to rate Muslim individuals on a scale of 1 to 5 across several attributes, namely warmth, trustworthiness, positivity, friendliness, respect-worthiness, and admirability. This standardized set of attributes allowed for a comparative analysis of perceptions across the different populations under study.

Expectation

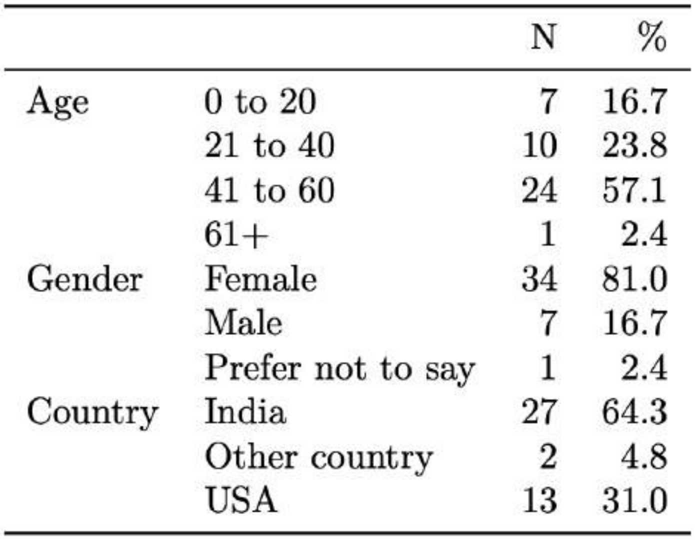
Based on the extended contact hypothesis, we expect that when people imagine interacting with a Muslims individual, it will make them see Muslims in a more positive light. By having non-Muslim participants imagine contact with a Muslim stranger, we aim to explore the positive effects of imagined intergroup contact on reducing prejudice. The mechanism behind why we would expect imagined contact to reduce prejudice is laid out by Crisp and Turner as follows: “imagining intergroup contact changes explicit out-group attitudes by activating conscious processes that parallel the processes involved in actual intergroup contact, for example thinking about what they would learn from the encounter and how that encounter would make them feel.”[5] Further, the reason behind using imaginary contact is because it is an alternative to direct contact, especially in situations where direct contact is not possible.

Results and Discussion

Top line results

We had 42 respondents in total to this question. Demographic characteristics of the respondents are shown below in Table 1. Most of our sample was in the 41 - 60 age range and the majority identified as female. The majority resided in India, with about a third in the US.

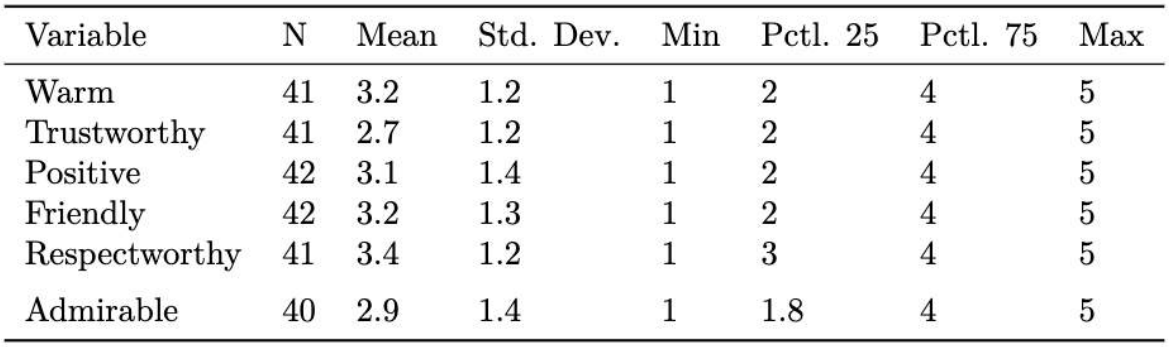
Table 1



		N	%
Age	0 to 20	7	16.7
	21 to 40	10	23.8
	41 to 60	24	57.1
	61+	1	2.4
Gender	Female	34	81.0
	Male	7	16.7
	Prefer not to say	1	2.4
Country	India	27	64.3
	Other country	2	4.8
	USA	13	31.0

The top line results for our dependent variables are shown below in Table 2.

Table 2



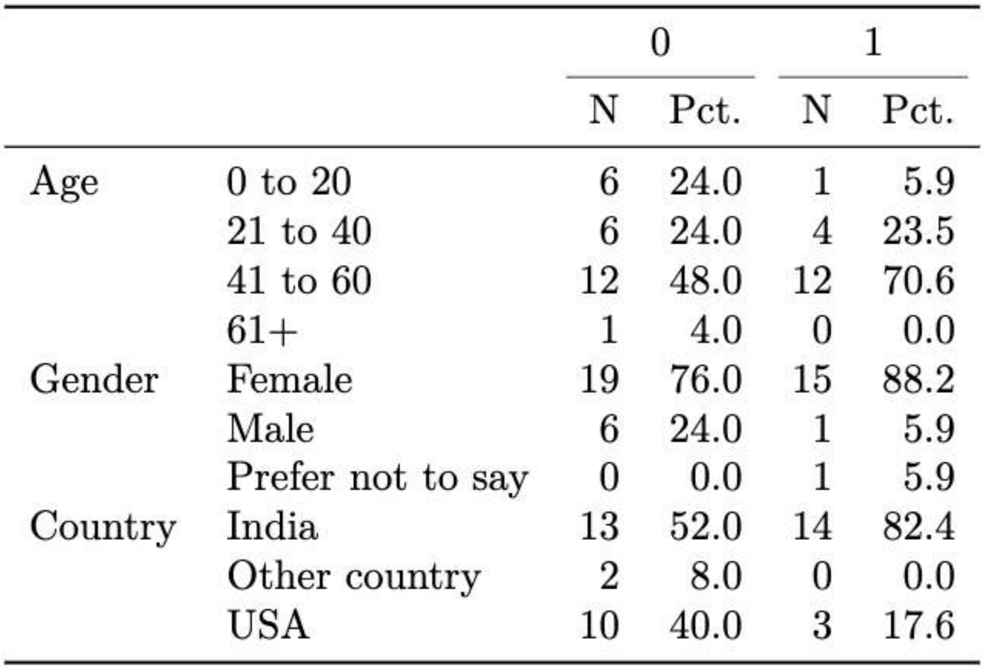
Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Warm	41	3.2	1.2	1	2	4	5
Trustworthy	41	2.7	1.2	1	2	4	5
Positive	42	3.1	1.4	1	2	4	5
Friendly	42	3.2	1.3	1	2	4	5
Respectworthy	41	3.4	1.2	1	3	4	5
Admirable	40	2.9	1.4	1	1.8	4	5

Participants were instructed to rank Muslim people on a scale from 1 to 5 in different categories. 1 symbolizes strongly disagreeing, while 5 represents strongly agreeing. Based on the mean scores, most participants rated Muslim people around 3 in all categories.

Randomization

Our software randomly allocated respondents to either the treatment or the control group, with a probability of 0.5. Within our sample of N=42, 25 respondents were treated (40 percent of respondents). Table 3 below shows the characteristics of treatment vs. control respondents.

Table 3



		0		1	
		N	Pct.	N	Pct.
Age	0 to 20	6	24.0	1	5.9
	21 to 40	6	24.0	4	23.5
	41 to 60	12	48.0	12	70.6
	61+	1	4.0	0	0.0
Gender	Female	19	76.0	15	88.2
	Male	6	24.0	1	5.9
	Prefer not to say	0	0.0	1	5.9
Country	India	13	52.0	14	82.4
	Other country	2	8.0	0	0.0
	USA	10	40.0	3	17.6

The table outlines general demographics for both the treated and control groups in the survey. Column header 0 represents the control group with a vague prompt about imagining contact with a Muslim individual. Column header 1 is the treated group, which includes a detailed prompt about imagining and reflecting upon contact with a Muslim individual.

Sub-column N displays the number of individuals for each version, and Column Pct. shows the percentage representation.

The data indicates an imbalance in randomization, with over half of the participants receiving the control condition. The distribution of age, gender, and location also indicate a lack of balance. For instance, around 70 percent of respondents in the treated condition were between 41 and 60, compared to less than 50 percent of respondents in the control condition. This is likely due to the small sample size of the survey. Given the imbalance in demographic characteristics, we will control for age, gender, and location in our treatment effect estimation.

Estimating treatment effect

To isolate the average treatment effect of our imagined contact intervention, we used an OLS regression model comparing the control and treatment group. The model specification is below.

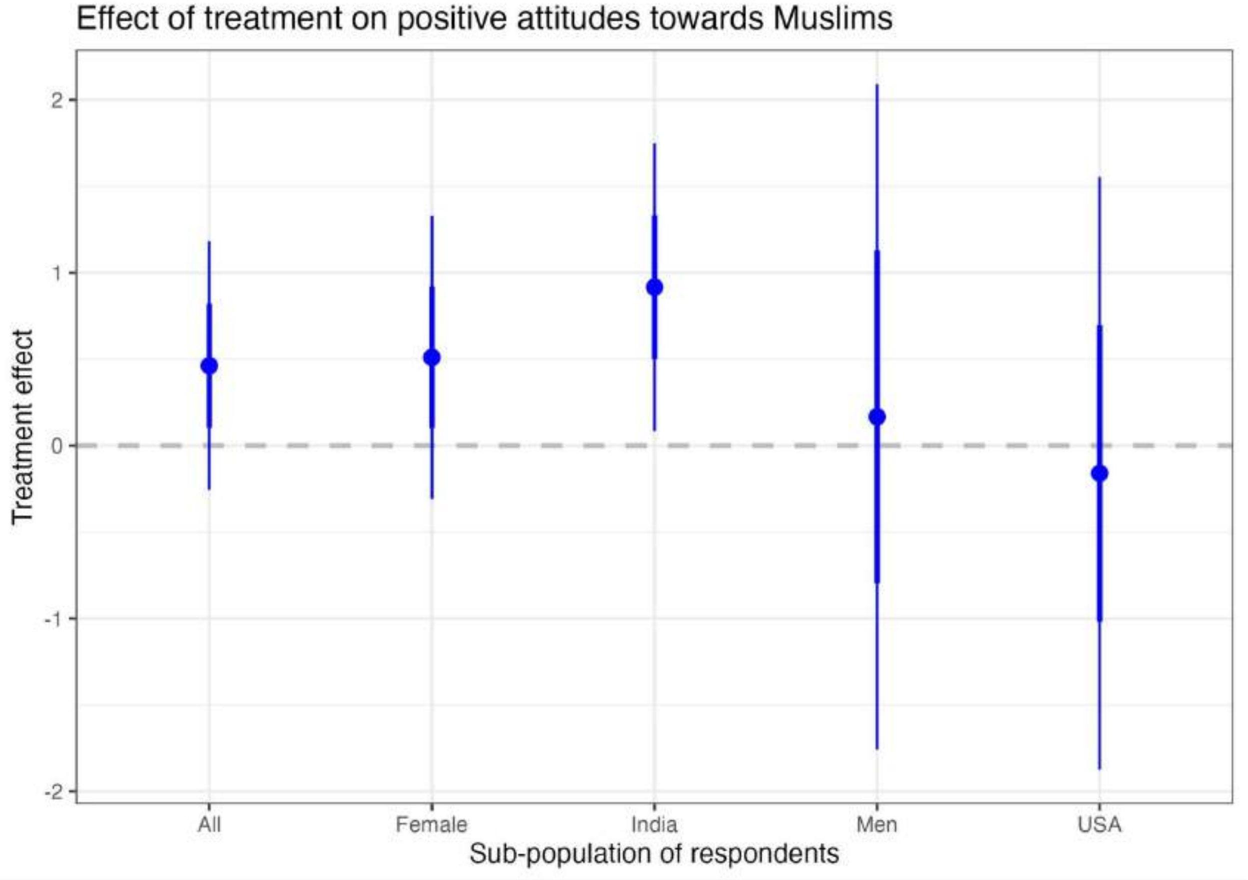
Y_(i)= β_0 +β_1·Treatment +β_(2)·AgeCohort_i +β_(3)·Gender_i + β_(4)·Country_(i)+ε _i

Where:

- Y_(i) is the score that the respondent has rated Muslims – averaged across all 6 categories
- β_0 is the intercept term
- Treatment is a binary variable indicating whether the respondent received the treatment (if experimental group, 1; if control group, 0)
- AgeCohort is a set of dummy variables indicating the age cohort of the respondent
- Gender is a set of dummy variables indicating the self-reported gender of the respondent (base group is female)
- Country is a set of dummy variables indicating the country that the respondent is located in (base group is India)
- β_1, β_2, β_3, β_4are the sets of coefficients to be estimated
- ε _i is the error term

We first used this model specification to perform a regression on the entire data set of our respondents. We then ran a similar regression on subsets of the data set to look for heterogeneous effects based on gender or location. The results of our regressions are in the appendix in Table 4. The coefficient on *Treatment* is shown below in Figure 1.

Figure 1



As we can see in the table, our treatment had a positive effect on attitudes towards Muslims across the entire group as well as most subgroups. Across all respondents, on average the effect of receiving the treatment is associated with a nearly 0.5 point increase in average score. Given our score is from 0 to 5, this represents a 10 percent increase in favorability of attitudes. However, likely due to a low sample size, many of these effects are not statistically significant.

We see some variation within geographic subgroups. Within the sub-group of our respondents who are in India, we see the highest effect, with a coefficient of 0.9, representing a 20 percent increase in favorability of attitudes. This result is statistically significant. For the USA sub-group, while the effect is technically negative, the low sample size gives us less confidence in this estimate. We ran similar regression models using the scores of each of the specific attributes as a dependent variable. However, neither was markedly different than the analysis using the composite dependent variable.

Discussion

Overall, the results from the replicated Turner and Crisp study showed a positive impact, but the statistical significance was not reached for the entire group due to a small sample size.

Interestingly, when looking specifically at the Indian subpopulation in India, the treatment effects became statistically significant. This may be for a few reasons. For one, participants in India may have encountered more real-world experiences of positive intergroup interactions with Muslims, influencing their willingness to engage in the imagined contact exercise with a more open mindset. This could be because there is a greater population of Indians in India compared to the United States, or even a more diverse population of Indians.

Conclusion

To conclude, our research aimed to replicate Crisp and Turner's study on imagined intergroup contact to explore its potential in reducing prejudice against Muslims among Hindu Indians in the US and India.[5] Despite limitations in sample size, the research results suggested that imagined contact had a positive effect on attitudes towards Muslims from Hindus. These findings highlight the potential of imagined intergroup contact as an alternative to direct contact, especially in situations where face-to-face contact may be challenging. Potential questions for future research include investigating the long-term sustainability of the observed effects using imagined contact, exploring the impact of imagined contact across different demographic cohorts, and assessing the generalizability of the results to other intergroup relations beyond Hindus and Muslims.

One of the most significant questions raised by this paper is centered around the reliability of the extended contact hypothesis in locations where respondents have limited or no actual contact with the members of a specific community. The reliability of the contact hypothesis may face challenges in environments where segregation or homogeneity are normalized. This is because opportunities for meaningful contact may be scarce, potentially hindering the ability to truly imagine positive contact.

Acknowledgements

Writing this paper and executing the research behind it would not have been possible without the exceptional support of my mentor Priyanka Sethy (PhD candidate at Harvard University). Her knowledge, guidance and laser-sharp attention to detail have been inspirational and vital to this research. I am also equally grateful to all the survey participants as without their participation and responses the research and its outcomes would not have been feasible.

Ethical Considerations

This study ensured to meet ethical consideration through the means of employing aspects like an Informed consent form which was collected verbally from the adult participants before they participated in the study. The data collected was kept anonymous and not shared with anyone except the researcher and the research mentor.

Works Cited

(1) Allport, G. W. The Nature of Prejudice, Unabridged, 25th anniversary ed.; Addison-Wesley Pub. Co: Reading, Mass, 1979.

(2) Pettigrew, T. F.; Tropp, L. R. A Meta-Analytic Test of Intergroup Contact Theory. Journal of Personality and Social Psychology 2006, 90 (5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751.

(3) Enos, R. D. Causal Effect of Intergroup Contact on Exclusionary Attitudes. Proc. Natl. Acad. Sci. U.S.A. 2014, 111 (10), 3699–3704. https://doi.org/10.1073/pnas.1317670111.

(4) Cameron, L.; Rutland, A.; Brown, R.; Douch, R. Changing Children's Intergroup Attitudes Toward Refugees: Testing Different Models of Extended Contact. Child Development 2006, 77 (5), 1208–1219. https://doi.org/10.1111/j.1467-8624.2006.00929.x.

(5) Turner, R. N.; Crisp, R. J. Imagining Intergroup Contact Reduces Implicit Prejudice. British J Social Psychol 2010, 49 (1), 129–142. https://doi.org/10.1348/014466609X419901.

(6) Tropp, L. R.; White, F.; Rucinski, C. L.; Tredoux, C. Intergroup Contact and Prejudice Reduction: Prospects and Challenges in Changing Youth Attitudes. Review of General Psychology 2022, 26 (3), 342–360. https://doi.org/10.1177/10892680211046517.

(7) Lowe, M. Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration. American Economic Review 2021, 111 (6), 1807–1844. https://doi.org/10.1257/aer.20191780.

(8) Wojcieszak, M.; Warner, B. R. Can Interparty Contact Reduce Affective Polarization? A Systematic Test of Different Forms of Intergroup Contact. Political Communication 2020, 37 (6), 789–811. https://doi.org/10.1080/10584609.2020.1760406.

(9) Pettigrew, T. F.; Tropp, L. R. How Does Intergroup Contact Reduce Prejudice? Meta‐analytic Tests of Three Mediators. Euro J Social Psych 2008, 38 (6), 922–934. https://doi.org/10.1002/ejsp.504.

(10) Wright, S. C.; Aron, A.; McLaughlin-Volpe, T.; Ropp, S. A. The Extended Contact Effect: Knowledge of Cross-Group Friendships and Prejudice. Journal of Personality and Social Psychology 1997, 73 (1), 73–90. https://doi.org/10.1037/0022-3514.73.1.73.

(11) Miles, E.; Crisp, R. J. A Meta-Analytic Test of the Imagined Contact Hypothesis. Group Processes & Intergroup Relations 2014, 17 (1), 3–26. https://doi.org/10.1177/1368430213510573.

(12) White, F. A.; Borinca, I.; Vezzali, L.; Reynolds, K. J.; Blomster Lyshol, J. K.; Verrelli, S.; Falomir‐Pichastor, J. M. Beyond Direct Contact: The Theoretical and Societal Relevance of Indirect Contact for Improving Intergroup Relations. Journal of Social Issues 2021, 77 (1), 132–153. https://doi.org/10.1111/josi.12400.

(13) Paluck, E. L.; Green, S. A.; Green, D. P. The Contact Hypothesis Re-Evaluated. Behav. Public Policy 2019, 3 (02), 129–158. https://doi.org/10.1017/bpp.2018.25.

Understanding drivers of wildfires in California By Jaden Randhawa

Abstract

This study investigates the intricate relationship between El Niño-Southern Oscillation (ENSO), precipitation patterns, and wildfires in California over the past two decades. California, known for its susceptibility to wildfires, has experienced significant variations in precipitation influenced by the ENSO phenomenon. The objective of this research is to examine the impact of ENSO events on California's precipitation patterns and subsequently assess their association with wildfire occurrences. Predicting wildfire and precipitation wildfires are important to California's significant agriculture production and better preparing for possible droughts or wildfires. To achieve this objective, a comprehensive analysis of climate data spanning the last 20 years was conducted. Historical records of ENSO episodes, obtained from reputable sources, were measured by correlation with precipitation data acquired from meteorological databases. Additionally, wildfire occurrence and severity data were collected from relevant agencies and organizations to be measured for correlation. As it turns out, while there was some positive correlation with ENSO and precipitation, there was little direct correlation between ENSO and wildfires. This means that there are many more complex factors to be assessed when figuring out causes of wildfires in addition to ENSO. The findings of this research will contribute to a better understanding of causes of wildfires in California. The results can inform policymakers, land managers, and disaster response agencies in developing more effective strategies to mitigate the risks associated with wildfires. Additionally, the study will provide valuable insights into the influence of climate variability on fire-prone ecosystems, aiding in the prediction and management of future wildfires.

Introduction

California, renowned for its scenic landscapes and diverse ecosystems, has long been plagued by the devastating impact of wildfires. Its unique combination of dry vegetation, frequent droughts, and strong winds make it highly susceptible to these natural disasters. Over the past few decades, the state has experienced a dramatic increase in the frequency, intensity, and scale of wildfires, causing widespread destruction of forests, wildlife habitats, and human settlements(NASA, 2021). Adding on, California, through its many dry seasons and "wet" seasons, has seen odd precipitation patterns, ranging from entirely dry seasons, to an extremely rainy mass flood-causing wet season(Leonard and Moriarty, 2023). Among the various factors influencing California's wildfire patterns and unpredictable precipitation, the El Niño-Southern Oscillation (ENSO) phenomenon has gained considerable attention. Normally, trade winds in the pacific ocean push warm waters towards the western pacific ocean, towards Asia and Australia. On the other side of the ocean, near the Americas, as the warm water is pushed away from that area, it's replaced by cold water from deep in the ocean. This balance of ocean temperature causes more in the ocean near Australia and Asia, causing more rainfall in that area. Conversely, near the Americas, a lack of evaporation near them usually leads to less precipitation. When those trade winds that cause this balance weaken, that's when a weather pattern known as El Nino occurs. This causes the warmth in the ocean to move eastward as there's less wind to push it towards the west. With newfound warmth in the eastern pacific, there is now more evaporation in that area and therefore, usually more precipitation than average(USGS, 2023). Studying the relationship between El Niño, precipitation, and wildfires in California is of paramount significance. Precipitation plays a crucial role in mitigating the risk of wildfires by replenishing soil moisture and reducing the flammability of vegetation(NASA, 2021). The influence of ENSO on precipitation patterns in California has been well-documented, with El Niño events typically associated with above-average rainfall in the state. Through understanding the interplay between El Nino and precipitation, researchers can potentially gain ways to track California's precipitation. In recent years, the field of predictive modeling has emerged as a powerful tool for understanding and managing natural disasters, including wildfires. By utilizing historical data, climate records, and advanced modeling techniques, researchers and policymakers can develop predictive models that estimate the likelihood, intensity, and spread of wildfires under varying climatic conditions. These models can provide critical information for proactive planning, resource allocation, and early warning systems, enabling more effective adaptation and response strategies to mitigate the impact of wildfires. The primary objective of this study is to delve into the complex relationship between El Niño, precipitation, and wildfires in California using predictive modeling techniques. By analyzing historical climate data, wildfire records, and ENSO indices, we aim to develop a robust predictive model that can accurately forecast the influence of El Niño on precipitation patterns and subsequently assess its impact on wildfire occurrence and behavior. This research endeavor will contribute to a deeper understanding of the intricate connections between climate phenomena, precipitation, and wildfires, ultimately leading to more effective wildfire management strategies and improved resilience in California.

Methodology

2.1 Data collection:

This research looked at four different variables (ENSO Index, Total Annual Rainfall, Number of Hectares Burned, and Number of Wildfires) across three databases. All data was averaged for the corresponding year between the time range of 2000 to 2023.

The ENSO Index is derived from satellite sea level observations published by NASA MEaSUREs/PO.DAAC and can found publicly at (https://ggweather.com/enso/oni.htm).

The annual precipitation data used in this study was obtained from NOAA (https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/county/mapping/4/pcp/201902/1/value")

The data for the number of hectares burned and the number of wildfires was collected from the Historical Wildfire Occurrence records maintained by the Department of Forestry and Fire Protection CAL FIRE statistics. This can be accessed publicly online at (https://www.fire.ca.gov/our-impact/statistics).

Overall, a summary of this data in its raw form can be seen in the annex.

2.2 Data Analysis

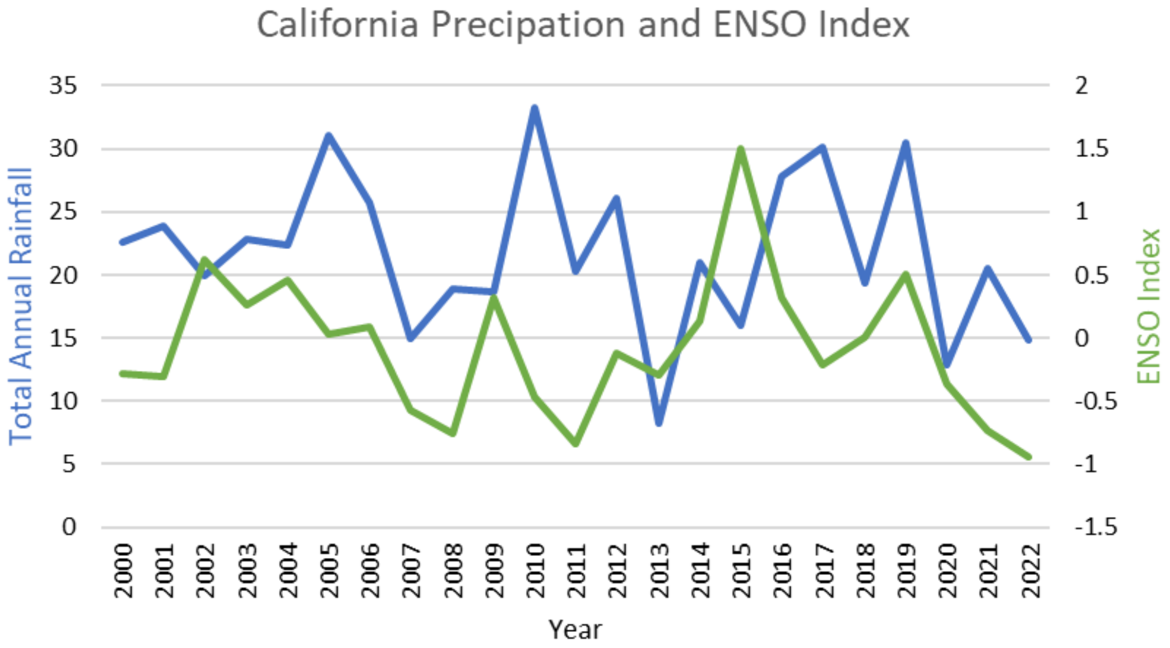
First, the trends in all the variables were assessed over time to gain an initial understanding of their temporal patterns and potential connections. This involved a comprehensive examination of how each variable evolved throughout the study period, allowing us to identify any notable shifts or patterns that might warrant further investigation.

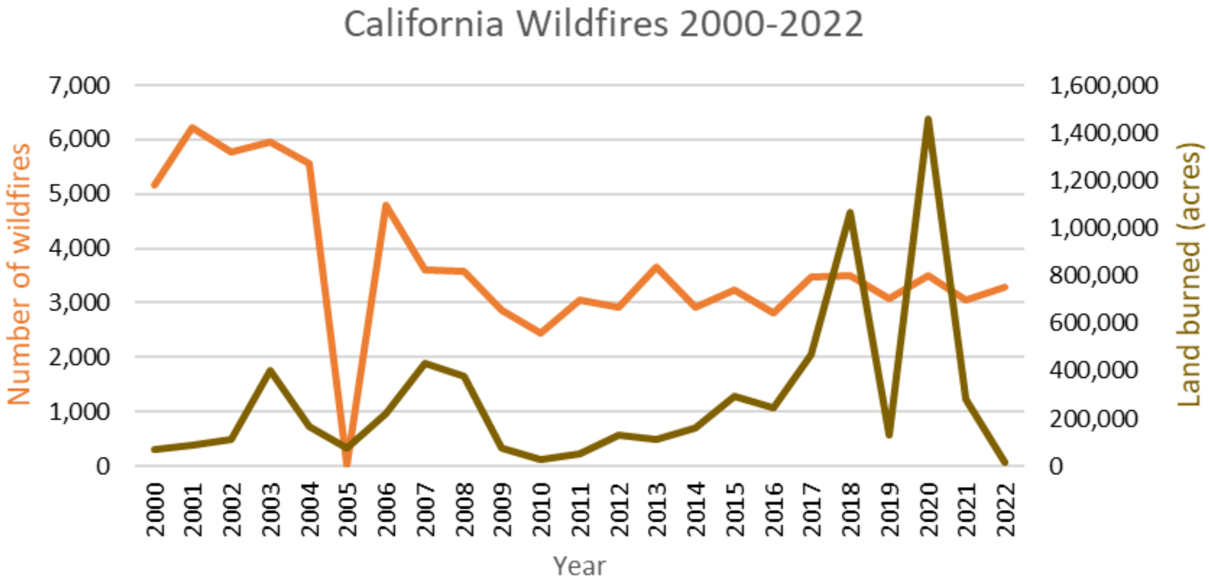
Next a Pearson's correlation matrix was conducted. This is a statistical tool used to measure the linear relationship between variables in a dataset. It provides a way to quantify how strongly two variables are related and in what direction (positive or negative) that relationship exists. The correlation coefficient, often denoted as "r," ranges from -1 to +1, where -1 indicates a perfect negative correlation, +1 indicates a perfect positive correlation, and 0 indicates no linear correlation. The purpose of constructing a Pearson's correlation matrix is to gain insights into the relationships between variables in a dataset so that linear relationships can be identified and the most relevant variables can be identified and selected for further analysis.

Lastly, a multiple linear regression model was used. A multiple linear regression model fits multiple variables and determines how “linear”, or in other words, correlated, they are. Multiple linear regression is a statistical technique used to establish a relationship between a dependent variable and two or more independent variables. In this method, the goal is to find the best-fitting linear equation that explains how the dependent variable changes concerning the independent variables. By analyzing the data, the model estimates the coefficients for each independent variable, representing their individual contributions to the dependent variable's variation. The resulting equation can then be used to make predictions or understand the impact of changes in the independent variables on the dependent variable(Wagavkar, 2023). Correlation matrices allow for more variables to be added in and measured at the same time, which can provide a better understanding of the dependent variable and the factors influencing it (wildfires). In this case, we have multiple independent variables (total annual precipitation, ENSO Index Average) that show correlations that are at least somewhat visible with the number of wildfires. By including these variables in a correlation matrix, we can examine their combined effects on the prediction of wildfires.

Results

3.1 Trends





3.2 Correlation Matrix

Measured Correlation between all Variables

	Total annual rainfall	Number of Wildfires	Land Burned (Acres)	ENSO Index Average
Rainfall	1	-0.21	-0.31	0.13

Wildfires	-0.21	1	0.03	0.13
Land Burned	-0.31	0.03	1	-0.04
ENSO Index	0.13	0.13	-0.04	1

The Pearson's correlation matrix was analyzed to determine the relationships between the variables: "Total Annual Rainfall," "Number of Wildfires," "Land Burned (Acres)," and "ENSO Index Average." The correlation coefficients reveal important insights into the associations between these variables.

Firstly, there was a weak negative correlation observed between "Total Annual Rainfall" and the "Number of Wildfires" (-0.21). This suggests that as the total annual rainfall increases, there tends to be a slight decrease in the number of wildfires. While the correlation is weak, rainfall still has some effect on decreasing wildfires, leaving in many other possible factors that decrease the amount of wildfires.

Additionally, a slightly stronger negative correlation was found between "Total Annual Rainfall" and "Land Burned (Acres)" (-0.31). This indicates that as the total annual rainfall increases, the total land burnt by wildfires annually tends to drop. This correlation makes logical sense as when rain slightly decreases the number of wildfires, the amount of land burned correlatively goes down.

Furthermore, a weak positive correlation was observed between "Total Annual Rainfall" and the "ENSO Index Average" (0.13). This implies that as the total annual rainfall increases, there tends to be a slight positive association with the ENSO (El Niño-Southern Oscillation) Index Average. The ENSO Index measures the state of the El Niño or La Niña climate patterns, which can influence weather conditions globally. This correlation suggests that higher precipitation levels might be influenced by specific phases of the ENSO cycle, which displays the wetter side of the Pacific Ocean caused by El Nino’s weather disturbance.

Regarding the relationship between the "Number of Wildfires" and "Land Burned (Acres)," a very weak positive correlation was found (0.03). Although this correlation is not significant, it suggests a slight tendency for a higher number of wildfires to be associated with a slightly greater extent of land burned. However, the correlation coefficient indicates that this relationship is not particularly strong.

Furthermore, a weak positive correlation was observed between the "Number of Wildfires" and the "ENSO Index Average" (0.13). This implies that there is a weak positive association between the number of wildfires and the ENSO Index Average. It suggests that some parts of the ENSO cycle may be a factor increasing wildfires in California.

Lastly, a weak negative correlation was found between "Land Burned (Acres)" and the "ENSO Index Average" (-0.04). This indicates that there is a weak negative relationship between the extent of land burned by wildfires and the ENSO Index Average. Although this correlation is not particularly strong, it suggests that certain phases of the ENSO cycle might have a slight influence on the severity and extent of wildfires.

3.3 Multiple Linear Regression

The objective of the multiple linear regression analysis was to establish a predictive relationship between the extent of land burned (measured in acres) and a set of five independent variables. The analysis yielded insightful statistical metrics for evaluating the model's performance. The multiple correlation coefficient (Multiple R) indicated a moderate positive association between the predictors and the land burned variable. The coefficient of determination (R Square) revealed that approximately 22.14% of the variability in land burned could be accounted for by the chosen predictors. However, the adjusted R Square, showing a negative value, suggests the potential presence of overfitting or underscores the necessity for additional pertinent variables. The computed standard error underscored the variability between predicted and observed values.

The p-value associated with the F-statistic surpassed the conventional significance level of 0.05, implying that the model's statistical significance in explaining the variance in land burned might be limited. The examination of the individual predictor coefficients unveiled that none of the variables exhibited statistically significant relationships with land burned at the 0.05 significance threshold. This observation was further supported by the p-values corresponding to the t-statistics, which all exceeded 0.05. Consequently, the individual predictors did not yield substantial predictive insights into the extent of land burned.

4. Discussion

The Pearson's correlation matrix was analyzed to determine the relationships between the variables: "Total Annual Rainfall," "Number of Wildfires," "Land Burned (Acres)," and "ENSO Index Average." The correlation coefficients reveal important insights into the associations between these variables.

There was a weak negative correlation observed between the rainfall" and the number of wildfires, a slightly stronger negative correlation between rainfall and land burned, a weak positive correlation between rainfall and El Nino.

Overall, these correlations provide valuable insights into the relationships between the variables studied. It is important to note that weaker correlations may have been resulted from large differences in numbers due to different ways in which the different variables were measured (i.e. acres of land, inches of rainfall, number of wildfires, etc.). It is also important to note that correlation does not imply causation, and other factors not considered in this analysis could also contribute to the observed patterns. However, these results may very well be the start of a better way to understand different drivers of wildfires in California. Further research and analysis are needed to fully understand the complex interactions between annual precipitation, wildfire occurrence, land burned, and the ENSO climate patterns.

5. Conclusion

The purpose of this experiment was to measure the correlation between the occurrence of El Nino with the amount of precipitation, number of wildfires, and land burned in each individual year of the 21st century in California so far. Through the correlation matrix and the multiple linear regression, there seems to be little to no correlation between the occurrence of El Nino with the amount of precipitation, wildfires, and land burned in California. This goes against the initial prediction that there would be a significant correlation. One way this experiment may be improved is by using first hand data in addition to secondary data as then there could be a bit more certainty on the findings of this experiment. Future research on this topic could allow for better and more accurate predictions on California's rainy seasons and dry seasons (Stevenson and Xingying, 2020).

Works Cited

Huang, Xingying, and Samantha Stevenson. ‘Contributions of Climate Change and ENSO Variability to Future Precipitation Extremes over California’. *Geophysical Research Letters*, vol. 50, no. 12, American Geophysical Union (AGU), June 2023, https://doi.org10.1029/2023gl103322.

Wagavkar, Sanskar. ‘Introduction to the Correlation Matrix’. *Built In*, 17 Mar. 2023, https://builtin.com/data-science/correlation-matrix.

Center for Climate Science. *The Future of Extreme Precipitation in California*. Jan. 2017, https://www.ioes.ucla.edu/project/future-extreme-precipitation-california/.

‘Less Predictable Precipitation’. *UCI News*, 16 Jan. 2018, https://news.uci.edu/2018/01/16/less-predictable-precipitation/.The Science behind California’s Extremely Wet Winter, in Maps. *The Washington Post*, WP Company, 10 Apr. 2023, www.washingtonpost.com/weather/2023/04/07/california-extreme-winter-storms-snow-climate/.

US Department of Commerce, and National Oceanic. *What Are El Nino and La Nina?* Mar. 2009, https://oceanservice.noaa.gov/facts/ninonina.html.

What Is ‘El Niño’ and What Are Its Effects? https://www.usgs.gov/faqs/what-el-nino-and-what-are-its-effects. Accessed 3 Sep. 2023.

What’s behind California’s Surge of Large Fires? NASA Earth Observatory, Oct. 2021, https://earthobservatory.nasa.gov/images/148908/whats-behind-californias-surge-of-large-fires.

Determinants of Tomato Demand in Delhi NCR: Analyzing Price and Income Elasticities During 2023 Price Spikes By Siya Mehra

Abstract

Tomatoes, an important ingredient in Indian cuisine, have highly volatile prices due to various factors. While previous studies have investigated demand determinants for vegetables and fruits, there is limited research focusing specifically on tomatoes in the context of Delhi NCR. This study aims to ascertain the factors influencing consumer demand for tomatoes in Delhi NCR, considering the price spike in the latter half of 2023 and its impact on traditionally inelastic demand. Using a quantitative approach, data was collected via a Google Forms survey from 98 households across different income groups in Delhi NCR. The study examines price elasticity, income elasticity, and the role of substitutes like ready-made tomato puree on the demand for tomatoes. Findings reveal that tomato demand is price inelastic across all income groups, with a negative income elasticity suggesting that tomatoes are viewed as an inferior good by higher-income households. Convenience drives the demand for ready-made puree, with a strong preference for homemade options. These insights are critical for policymakers to design effective price stabilization policies, marketers to target consumer preferences, and agricultural producers to enhance supply chain efficiency.

Keywords: Tomato, Consumer Demand, Price Elasticity, Income Elasticity, Delhi NCR

1. Introduction

India, renowned for its rich cultural heritage, is celebrated worldwide for its delectable and boldly spiced cuisine. Tomatoes, especially, hold a pivotal role in Indian cooking, featuring in various forms such as salads, purees, and gravies. Tomato puree, in particular, is a prevalent ingredient used in a myriad of dishes including vegetables, curries, salads, and chutneys across the country. India, is actually, one of the largest consumers of tomatoes worldwide, coming in second only to China. In 2023, India individually produced around 20.62 million metric tons of tomatoes. Internationally, India owns 12 per cent of the tomato market, with most of its imports going to Nepal, Bhutan and Bangladesh, according to Statista (2023)

Despite this demand, the price of tomatoes, like other agricultural products, is heavily influenced by factors such as supply quantity, seasonality, land fertility, and climate, which are typically uncontrollable. This leads to significant price fluctuations. Northern states like Uttarakhand, Himachal Pradesh and Jammu and Kashmir and Northeastern states like Assam, Sikkim and Arunachal Pradesh face higher prices due to being located farther from tomato producing states of the country like Haryana, Punjab, Maharashtra and most southern states (Zeebiz, 2023). In Delhi, the national capital, tomato prices typically range from ₹20-40 per kg, varying by region. However, prices can surge to ₹60-80 per kg during the winter season due to a shorter supply. (CEICdata.com, 2024)

In September 2023, the retail price of tomatoes in Delhi was ₹25/kg, marking a significant decrease from ₹43/kg in August 2023. This monthly price data has been recorded since August 2008, comprising 180 observations up to September 2023. Over this period, the average retail price of tomatoes stood at ₹30/kg. Notably, prices reached an all-time high of ₹173/kg in July 2023 and a record low of ₹10/kg in May 2011 in Delhi.

The fluctuations in prices of agricultural commodities like tomato in India is caused due to volatility in weather events like droughts, floods, and pest outbreaks. These significantly impact crop yields, directly affecting supply and pushing prices up. Additionally, inefficient infrastructure, spoilage, hoarding, and transportation bottlenecks can limit availability, leading to localised price spikes. These factors were major contributors to the price spikes observed from June to August 2023. A study conducted at the Indian Agricultural Research Institute (IARI) in 2020 aimed to forecast onion prices for the Varanasi market in Uttar Pradesh, India. Onions are similar to tomatoes in terms of their usage in the Indian market and are also a staple vegetable. The paper found that the demand as well as the price of the onions is primarily influenced by the time of harvest and local produce. When onions are locally grown and in season, the demand tends to be higher as consumers prefer fresh, locally sourced produce. On the demand side, fluctuations can occur due to seasonal variations in consumption and availability, leading to price adjustments. The availability and pricing of substitutes and complementary products can influence tomato demand and prices. A rise in income can also increase demand for tomatoes, putting upward pressure on prices. Additionally, tomatoes have a short shelf life, making them more susceptible to spoilage and price fluctuations compared to its substitute, tomato puree, which has a comparatively longer shelf life. It is also more convenient to purchase a readymade product instead of buying fresh tomatoes and making puree thereafter. Birthal et al. (2019) conducted a study that investigates causes of volatility in onion prices in India. As mentioned earlier, onions are an important staple in Indian diets and have fluctuating prices similar to tomatoes. The paper has a more production-side perspective in terms of their utility for the findings, which showed that the demand for onions was price inelastic. Price inelasticity is when even a large change in price would not affect demand heavily; for instance if the price increased by 15%, the demand would change by a lower percentage. Furthermore, even a slight output deficit appears to give supply chain players a chance to engage in anti-competitive behavior, such as stockpiling for a future increase in price, which aggravates the price consequences of lower supply.

“Socio-Economic Determinants of Fruits and Vegetable Consumption: Insights from a Survey in Delhi (Gupta, 2021)” is a research paper written on similar lines. It investigates the impact of non-price determinants of demand for fruits and vegetables in Delhi like health, convenience, gender, education, and income. One of the findings in the paper is that convenience, in terms of availability and preparation time, was a crucial factor in consumption. Moreover, the majority of middle-class and upper-class consumers who earn more than Rs. 25000 per month were seen to not prioritise price. Another important finding of the study is that, people in Delhi prioritise health and nutrition when choosing fruits and vegetables to eat. They also care about the taste and freshness of the produce. Due to this, most people in Delhi prefer to buy fruits and vegetables from local markets and street vendors rather than supermarkets or online. The justification presented for this outcome was that produce from local markets is perceived to be fresh and affordable. As for the impact of socio-economic factors, Factors like gender, education level, income, and occupation influence consumers' fruit and vegetable consumption patterns. They often don't have the affordability and availability to purchase their choice of vegetables, showing a gap in the market.

Another study by Lopez & Davis (2005) conducted a demand analysis for fresh tomatoes in the Dallas-Fort Worth area. This study explored the factors affecting consumer demand, including price, income levels, demographic characteristics, and the influence of alternative products available in the market

Although previous studies have been conducted regarding the demand of vegetables or fruits in India, there is a dearth of research in this domain that are trying to ascertain the determinants of demand of tomatoes in Delhi NCR. This study aims to investigate the factors influencing consumer demand for tomatoes in Delhi. Additionally, this study will be investigating the impact of a recent tomato price spike on the traditionally inelastic demand of tomatoes.

This study uses primary data-based research collected through a Google Forms survey. It is a quantitative study that incorporates economic theories of consumer demand to shed light on how factors like price, income, substitutes and socioeconomic status affect the demand for tomatoes.

2. Methodology

Our research aim is to ascertain the determinants influencing tomato demand in Delhi NCR, India, and to analyze the variations across distinct household income brackets. Delhi NCR has been chosen due to a variety in demographics, income levels, and infrastructure. It includes people from all ages, the majority being between 20 and 39 years of age. The population also varies in terms of socioeconomic wellbeing; the total population is divided into 3/4th urban population and 1/4th rural population. This study will be investigating a variety of factors affecting the quantity demanded for tomatoes (Q_D) such as price and non-price factors like substitutes (ready-made tomato puree) and income which will help us gain a better perspective of the demand of agricultural products, specifically, an essential agricultural product. This would be especially valuable in a country like India, where the agricultural sector is a large contributor to the economy. In order to collect data regarding the demand for tomatoes, the research design was a quantitative study through a survey. A survey was chosen as it allows for numerical analysis, providing a structured approach to understand factors influencing demand. The sampling technique used was convenience sampling where selection was done based on participants readily available. Responses were received from 98 households, belonging to the three income groups: low income group (Less than INR 5,00,000 annually), middle income group (INR 5,00,000 - 20,00,000 annually) and high income group (More than INR 20,00,000 annually). The responses were distributed well among the income groups with 30% from the high income group, 32% from the middle income group and 38% from the low income group. The survey was conducted using Google Forms for efficient data gathering. Ethical considerations were taken to ensure participants' data remains private. Participants were informed about the research purpose, their rights, and the voluntary nature of participation. By informing participants of where and how their data is being used, we prioritize participant well-being and confidentiality throughout the research process. Multiple choice questions were mainly used, structured to ask factors affecting tomato demand, consumption habits, preferences and lifestyle details.

3. Results

3.1 Consumer Behaviour

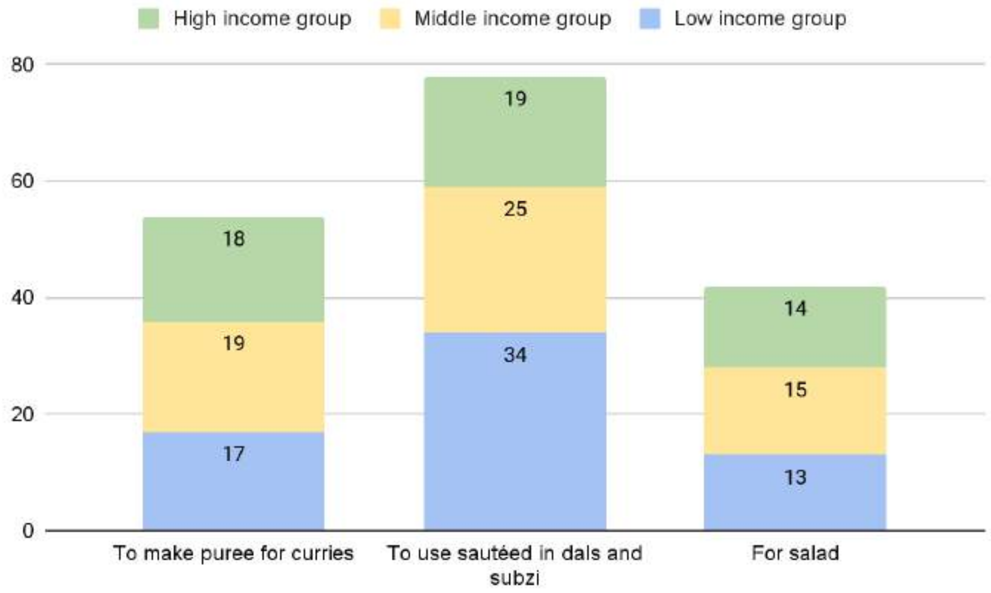


Figure 1: Reason for Buying Tomato Puree (N=88)

The data reveals distinct tomato usage patterns across income groups. Low-income households primarily use tomatoes sautéed in dals and subzi, followed by making puree for curries and in salads. Middle-income groups also favor sautéing but show a significant use for puree and salads. High-income households exhibit a balanced usage across all categories. This indicates that while sautéing in dals and subzi is common across all income levels, higher-income groups diversify their tomato usage more evenly, reflecting broader culinary habits and access to diverse food options.

3.2 Substitutes

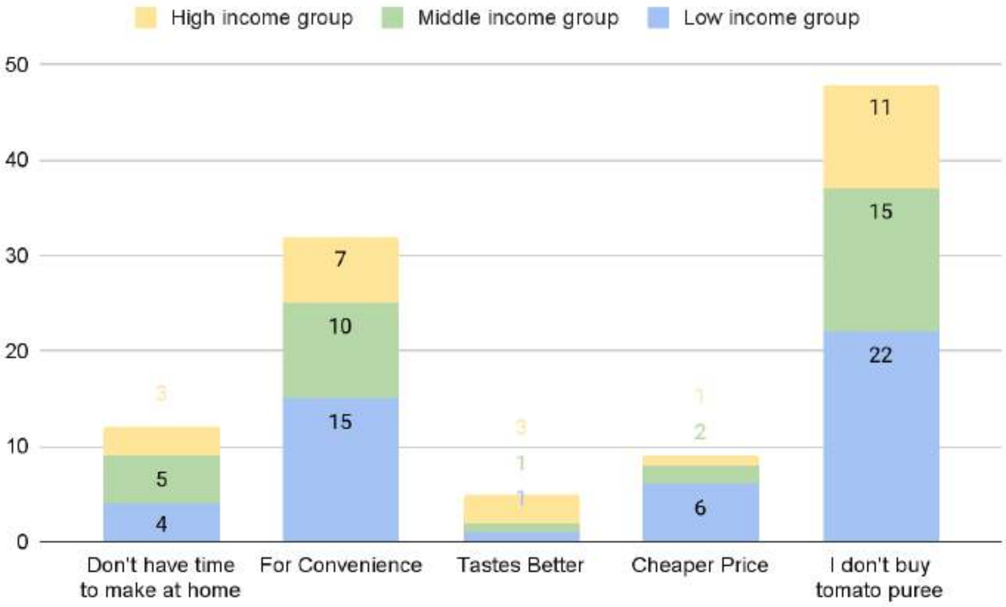


Figure 2: Reason for buying readymade tomato puree

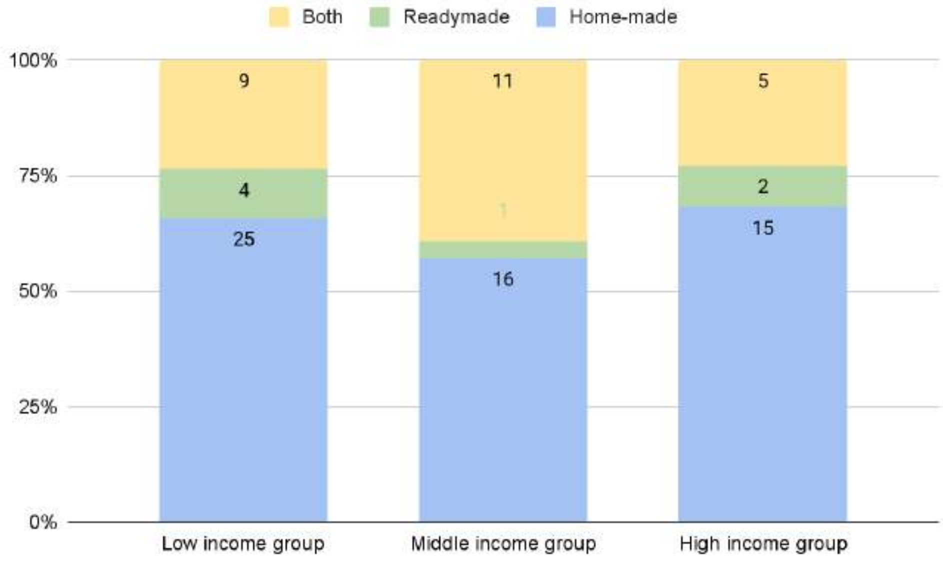


Figure 3: Q_D of Homemade and Readymade Tomato Puree

The data reveals key factors influencing tomato puree demand across the different income groups. A clear preference for homemade tomato puree is seen in all groups: almost 66% of the low-income group, 57% of the middle-income group, and 68% of the high-income group. Convenience is the primary reason for purchasing tomato puree, especially in the low and middle-income groups, with 15 and 10 respondents respectively citing it. Time constraints are a secondary factor, affecting 4 low-income, 5 middle-income, and 3 high-income respondents.

Taste and price are less influential. Notably, a significant portion of respondents—22 low-income, 15 middle-income, and 11 high-income—do not buy tomato puree at all, indicating either a preference for fresh tomatoes or non-consumption. Overall, the main drivers of demand for tomato puree are convenience and a preference for homemade puree, with time and cost playing lesser roles.

3.3 Price Elasticity

Price elasticity of demand (PED) measures how the quantity demanded of a good responds to changes in its price. The formula for calculating PED is:

$$Price\ elasticity\ of\ demand\ =\ \frac{\%\ Change\ in\ QD}{\%\ Change\ in\ Price} = \frac{\frac{Old\ QD - New\ QD}{Old\ QD} \times 100}{\frac{Old\ P - New\ P}{Old\ P} \times 100}$$

The new price was calculated by adding the lowest price during the price hike and the highest price, then dividing it to find the average price during this period.

$$New\ Price\ (after\ the\ price\ hike) \Rightarrow \frac{173+80}{2} = 126.5$$

Table 1: Price Elasticity of Demand for Tomatoes on the Basis of Household Income

Household Income	**Calculation of Price Elasticity (ε)**	**ε**
Low Income	$\frac{\frac{1.658-1.368}{1.658} \times 100}{\frac{44.737-126.5}{44.737} \times 100} = \frac{17.491\%}{-182.764\%}$	-0.096
Middle Income	$\frac{\frac{2.214-1.893}{2.214} \times 100}{\frac{46.429-126.5}{46.429} \times 100} = \frac{14.499\%}{-172.459\%}$	-0.084
High Income	$\frac{\frac{2.591-2.25}{2.591} \times 100}{\frac{49.091-126.5}{49.091} \times 100} \Rightarrow \frac{13.161\%}{-157.685\%}$	-0.083

Source: Author’s Calculation

Price elasticity of demand does not consider the negative sign as it is a given that there is a negative correlation between price and quantity demanded, which means an increase in the price will lead to a decrease in the quantity demanded of tomatoes. This negative correlation indicates that tomatoes are viewed as a normal good, but as seen in Table 1, the values are less than one in all cases, which means that the demand is inelastic. Additionally, the value is decreasing as the income groups progress, which means the scale of responsiveness is decreasing slightly. The price elasticity is 0.096 for the low income group, 0.084 for the middle income group, and 0.083 for the high income group. This suggests that the low income group is slightly more responsive to price changes compared to the middle and high income groups, but overall, all groups show relatively low responsiveness to price changes. Thus, while all income groups view tomatoes as a necessary good, the low income group is marginally more sensitive to price changes than the middle and high income groups.

3.4 Income Elasticity

Income elasticity of demand (YED) measures how the quantity demanded of a good responds to changes in consumer income. The formula for calculating YED is:

$$Income\ elasticity\ of\ demand \Rightarrow \frac{\frac{Change\ in\ QD}{Old\ QD}\times 100}{\%\ Change\ in\ income} \Rightarrow \frac{\%\ Change\ in\ QD}{\%\ Change\ in\ income}$$

Table 2: Income Elasticity of Demand for Tomatoes on the Basis of Household Income

Household Income	**Calculation of Income Elasticity (η)**	**η**
Low Income	$\frac{\frac{-0.026kgs}{1.658kgs}\times 100}{2.211\%} = \frac{-1.568\%}{2.211\%}$	-0.709
Middle Income	$\frac{\frac{-0.054kgs}{2.214kgs}\times 100}{3.179\%} \Rightarrow \frac{-2.439\%}{3.179\%}$	-0.767
High Income	$\frac{\frac{-0.068kgs}{2.591kgs}\times 100}{4.364\%} \Rightarrow \frac{-2.624\%}{3.179\%}$	-0.826

Source: Author's Calculation

According to Table 2, the income elasticity of demand for all income groups is negative, which means an increase in the income will lead to a decrease in the quantity demanded of tomatoes. This negative correlation could mean the sample views tomatoes as an inferior product, however, the values are less than one in all cases, which means the scale of change is low. Additionally, the value is decreasing as the income groups progress, which means the scale of responsiveness is increasing and the tomatoes are becoming more elastic in the higher income group compared to the middle income group as well as in the middle income group as compared to the low income group. Thus, the high income group views tomatoes as more of an inferior group than the middle and low income groups.

4. Discussion

The study indicates that the price elasticity of demand for tomatoes is inelastic across all income groups, with the low-income group being slightly more responsive to price changes than the middle and high-income groups. This is consistent with Birthal et al. (2019), who found that the demand for onions, a similar staple vegetable, is price inelastic in India. The inelastic demand suggests that tomatoes are a necessity, and consumers will continue to purchase them despite price fluctuations. Tomatoes are a staple in Indian cuisine, and are frequently used in dishes like curries, salads, and chutneys, making them an essential household item. The lower price elasticity in higher-income groups further supports Gupta's (2021) findings, which suggest that middle and upper-class consumers in Delhi prioritize health, convenience, and nutrition over price. This demographic is less sensitive to price changes because they can afford to maintain their consumption levels.

The income elasticity of demand findings are somewhat unconventional. The study reveals a negative income elasticity across all income groups, implying that an increase in income leads to a decrease in the quantity demanded for tomatoes. This suggests that tomatoes are viewed as an inferior good by the sample population, particularly among higher-income households. This outcome can be interpreted through the lens of dietary preferences and purchasing power. As income increases, consumers might opt for more diverse and potentially healthier food options, including organic or exotic vegetables, which are perceived to be superior. This aligns with Lopez & Davis's (2005) findings in the Dallas-Fort Worth area, where higher income levels lead to a preference for premium products. However, the perception of tomatoes as an inferior good contradicts Gupta (2021), which found that Delhi's middle and upper-class consumers prioritize health and nutrition, leading to a preference for fresh produce. This discrepancy may be attributed to the specific demographic characteristics of the sample population in this study, who may associate higher income with the ability to afford a more varied diet.

Moreover, this study considers ready-made tomato puree, which has a longer shelf life and is more convenient, as a substitute. This aligns with the findings of the IARI study (2020) on onions, where demand was influenced by local produce availability and convenience. Consumers in Delhi, especially those with higher incomes, prefer ready-made tomato puree due to its convenience, contributing to the negative income elasticity observed. The study indicates that convenience is the primary driver of demand for tomato puree across all income groups.

This aligns with Gupta's (2021) findings that convenience is crucial in determining fruit and vegetable consumption in Delhi. The data shows a significant portion of respondents prefer fresh tomatoes over ready-made puree, reflecting a cultural preference for fresh produce in Indian cuisine. Taste and price of the puree are less influential factors compared to convenience and time constraints, especially among higher-income consumers. This supports the notion that higher-income groups prioritize convenience over cost, as they can afford time-saving products. In summary, the demand for ready-made tomato puree is driven mainly by convenience and time constraints, with a strong preference for homemade puree. This highlights the importance of considering non-price factors, such as convenience and cultural preferences, in analyzing the demand for agricultural products in urban areas.

5. Conclusion

This study aimed to investigate the factors influencing consumer demand for tomatoes in Delhi NCR, with a particular focus on the impact of price, income, and substitutes. The primary findings indicate that the demand for tomatoes is price inelastic across all income groups, with the low-income group being slightly more responsive to price changes. The income elasticity results suggest that tomatoes are viewed as an inferior good, especially among higher-income households. Convenience is the main driver of demand for ready-made tomato puree, with a strong preference for homemade puree across all income groups.

The implications of these findings are significant for policymakers, marketers, and agricultural producers. For policymakers, understanding the inelastic nature of tomato demand can help in designing better subsidy and price control policies to stabilize market prices and ensure food security. Marketers and retailers can leverage the preference for convenience and homemade puree by offering products that cater to these needs, such as easy-to-use tomato puree packs or fresh tomato delivery services. Agricultural producers can focus on improving the supply chain efficiency to reduce spoilage and transportation bottlenecks, ensuring a steady supply of fresh tomatoes.

Despite the valuable insights, this study has limitations. The use of convenience sampling may not accurately represent the entire population of Delhi NCR. Additionally, the survey method depends on participants accurately recalling their historical consumption patterns, which may result in reporting inaccuracies due to memory limitations. Future research should consider a larger and more diverse sample and incorporate longitudinal data to capture changes in consumer behavior over time.

Overall, this study provides a comprehensive understanding of the factors affecting tomato demand in Delhi NCR, highlighting the importance of considering both price and non-price factors in agricultural and economic planning.

Works Cited

Birthal, P., Negi, A., & Joshi, P. (2019). Understanding causes of volatility in onion prices in India. Journal of Agribusiness in Developing and Emerging Economies, 9(3), 255–275. https://doi.org/10.1108/jadee-06-2018-0068

CEICdata.com. (2024, March 14). India Retail price: DCA: Month end: Tomato: North Zone: Delhi. https://www.ceicdata.com/en/india/retail-price-department-of-consumer-affairs-agriculture-commodities-month-end-by-cities-tomato/retail-price-dca-month-end-tomato-north-zone-delhi

Census of India. (2022). ECONOMIC SURVEY OF DELHI, 2022-23. In ECONOMIC SURVEY OF DELHI, 2022-23 [Report].
https://delhiplanning.delhi.gov.in/sites/default/files/Planning/ch._19_demographic_profile.pdf

Government of NCT of Delhi. (2022). Highlights of Economic Survey of Delhi 2022-23. https://delhiplanning.delhi.gov.in/sites/default/files/Planning/highlights_of_es_2022-23_english.pdf

Gupta, N., Bhattacharjee, M., & Roy Saha, A. (2022). Socio-economic determinants of fruits and vegetable consumption: insights from a survey in Delhi. Journal of Postharvest Technology, 10(4). https://www.researchgate.net/profile/Nisha-Gupta-22/publication/370003974_Socio-Economic_determinants_of_Fruits_and_Vegetable_Consumption_Insights_from_a_Survey_in_Delhi/links/64390f762eca706c8b5bfdf5/Socio-Economic-determinants-of-Fruits-and-Vegetable-Consumption-Insights-from-a-Survey-in-Delhi.pdf

Kumar, P., Badal, P. S., Paul, R. K., Jha, G. K., P, V., Kamalvanshi, K., P, A., M, B., & Patel, P. (2021). Forecasting onion price for Varanasi market of Uttar Pradesh, India. Indian Journal of Agricultural Sciences/Indian Journal of Agricultural Sciences, 91(2). https://doi.org/10.56093/ijas.v91i2.11160

Lopez, J. A., & Davis, C. (2015). A demand analysis for fresh tomatoes in the Dallas/Fort Worth grocery market. Texas Journal of Agriculture and Natural Resources, 30, 16–37. https://doi.org/10.22004/ag.econ.196847

Rajvanshi, A. (2023, July 28). In India, tomatoes are now more expensive than gas. Here's why. TIME. https://time.com/6298825/india-tomato-crisis/

Report Linker. (2022). India Tomato Industry Outlook 2022 - 2026. In ReportLinker. https://www.reportlinker.com/clp/country/484797/726396

Statista. (2023, August 24). Production volume of tomatoes in India FY 2015-2023. https://www.statista.com/statistics/1039712/india-production-volume-of-tomatoes/

ZeeBiz. (2023, July 12). The tomato crisis in India: Can we expect a price ease soon? Zee Business. https://www.zeebiz.com/economy-infra/news-tomato-rate-today-news-can-we-expect-a-price-ease-soon-cpi-inflation-data-to-be-released-243837#

Evaluation of Urban Renewal Methods for in Cheonan, South Korea By Junhyeon Bae

Abstract

Urbanization and the shift in city demographics over the past few decades has created an increasing demand for urban renewal, especially urban renewal through cultural diversity or cultural activities. Previous research has analyzed the effectiveness of urban renewal driven by cultural diversity in the cities across East Asia, but few have explored the viability of such an approach in South Korea, let alone its medium-sized cities like Cheonan City, which are in desperate need of renewal. The purpose of this paper is to explore Cheonan's city structure, changing demographic, and recent urban renewal attempts to determine if cultural activities can meaningfully contribute to the revitalization of Cheonan's old downtown area called Myeongdong. First, the paper examines case studies of Itaewon in Seoul, and Koreatown in Osaka, and Tompkins Square Park in New York to outline the advantages and disadvantages of using cultural diversity for urban renewal. What follows is an analysis of the viability of urban renewal in Cheonan using Korean traditional culture, following the successful case of Hanok village in Jeonju, Korea. Results indicate that foreign cultures and Korean traditional culture should be combined in Cheonan's urban renewal in order to mitigate the opposition of citizens and supplement the lack of cultural elements.

Introduction

Since 2007 over half the world has lived in urban over rural settings, the population in cities has continued to grow, and cities have had to change rapidly to accommodate the influx. While one may assume that growing cities mean they will take up more land, only 1 or 2 percent of "global land is defined as a built up area" (Ritchie et al.) Thus, instead of expanding outward, cities must constantly upgrade and renew existing areas in a process urban studies scholars call urban renewal. Merriam-Webster defines the term "urban renewal" as "a construction program to replace or restore substandard buildings in an urban area" ("Urban Renewal"). Urban renewal can be accomplished in several ways, such by introducing eco-friendly elements, introducing cultural elements, and introducing new technologies. Recent literature has emphasized the need for urban renewal by introducing cultural elements to the cities, especially cultural diversity and/or cultural activity. For example, Sasaki suggests that declining cities in developed countries should implement cultural diversity in order to develop their "creative economy" due to the decline of the manufacturing industries and hi-tech electronics that led to the economic crisis (Sasaki 36). In other words, culture-driven urban renewal can be a driving force for a city's economy.

However, it is debatable whether the concept of cultural diversity can be applied to urban renewal without exception. Cities across the world have different environments, demographics, histories, and cultures. Attempting to use cultural diversity for urban renewal without consideration of these elements can cause damage that outweighs potential benefits, leading to a failure in urban renewal and fracturing of the community. Cheonan, South Korea is currently undergoing urban renewal, but it has not been clearly established if urban renewal using cultural diversity would benefit the city. Therefore, this paper will consider the city's demographics, industries, city structure, and cultural elements to examine the viability of urban renewal using cultural diversity in Cheonan. The paper will also consider various case studies in other East Asian cities to evaluate the advantages and disadvantages of cultural urban renewal in the South Korean context.

Introduction to Cheonan City

Cheonan City is located in South-Chungcheong Province below the Seoul Capital Area that includes both Seoul City and Gyeonggi Province. The population of Cheonan surpassed 650 thousand in 2017, and the population continues to steadily increase because of the city's booming manufacturing and service industries, which comprise 95.2% of the city's total industry. However, despite the population increase, the old downtown near Cheonan Station, named "Myeongdong," did not receive appropriate development because recent investment has focused on the Northern and the Western side of the city ("2035 Cheonan Masterplan" 17, 22, 24, 29, 160). Namely, the northern bus terminal was removed, a new train station with a high-speed railroad was built in the west, and City Hall was also moved to the western side (Sung and Lee 3235). These changes resulted in the "doughnut phenomenon" where the city center was underdeveloped in comparison and eventually abandoned by residents.

One other notable development in Cheonan is the recent increase in the foreign population. The city's foreign population, the majority are Chinese and Vietnamese, has increased from 34,908 in 2021 to 38,456 in 2022, which is a 0.4% increase ("Status of Foreign-born and Multicultural Households"). According to its population statistics by June 2024, the foreign population in the old downtown area was 1099, comprising 7.9% of the total population in that district. This was 2.7% and 3.5% higher than the foreign population of the entire city and the foreign proportion of South Korea, respectively. The changing demographic has made Korean citizens perceive the area as unsafe due to the general lack of cultural diversity in Korea ("Population Status"; "2.26 Million Foreign-born Residents"). Therefore, it can be concluded that the old downtown district that requires urban renewal has a relatively higher foreign population and more elements of multicultural society than other areas of Korea. Furthermore, if not handled properly, it is possible that multiplying culturally diverse elements through urban renewal could increase tensions between the Korean and non-Korean population in Cheonan.

Current Urban Renewal Plans

Currently, the city is implementing different approaches to renew Myeongdong. There were attempts to reconstruct the Cheonan station with private capital in 2000 and to construct skyscrapers, highrise residences, and a new ward office in 2007; however, projects were canceled or had minimal effects on the urban renewal (Sung and Lee 3236-3237). More recently, the city opened an urban regeneration center and established a regeneration plan in the district, which included the renewal of an underground shopping mall, the construction of themed streets, the demolition of a brothel, and the construction of a new ward office (An 15). However, these plans were also ineffective, as indicated by the failure of cultural events in Cheonan. For example, Cheonan has held the annual Christmas festival in the Myeongdong area for 2 years (Lee 449). According to the survey that was distributed in person to the merchants in Myeongdong, the local merchants rated the effects of the festival on the increase in nearby store sales as 2.74, on the increase in restaurant sales as 2.72, and on the increase in cultural exchange as 2.70. All of them were outnumbered by the need for program diversification, which scored 4.00 (Lee 449,450). This leads to the conclusion that one of the current urban renewal plans in the district was not effective enough to bring economic benefits to the area.

Advantages and Disadvantages of Implementing Cultural Diversity

This begs the question: is using cultural diversity for urban renewal appropriate for Cheonan's context? First it is necessary to weigh the pros and cons of this urban renewal method. Implementing cultural diversity in urban renewal makes the district more unique, appealing, and safe, but there is also a possibility of promoting prejudice, especially in a culturally and ethnically homogenous country like Korea.

Itaewon, a district in Seoul previously occupied by the US Army in the 1950s, succeeded in urban renewal by introducing cultural diversity. Formerly, Itaewon was perceived as a contaminated area by Koreans because of the presence of sex workers and gay clubs. However, from the 1990s, Itaewon started to accept a more diverse multicultural population—such as English teachers from Western countries, Muslims who attend the nearby central mosque, and second-generation Koreans from abroad—and foreign businesses. As a result, the younger generation of Koreans started to visit Itaewon to experience authentic foreign cuisine and fashion, changing the perception of Itaewon as a dangerous and foreigner-only area into a booming commercial district (Kim 4). The fact that the district was renewed within a decade by diverse immigrants who established multicultural businesses is an indication that cultural identity can be an effective approach to urban renewal in the old downtown of Cheonan too. Furthermore, the fact that Cheonan was previously occupied by US soldiers coincides with the situation of the old downtown in Cheonan, which also has a relatively high foreign population and is considered dangerous.

Osaka, another East Asian city, mixed Korean culture with the traditional Japanese culture to accomplish urban renewal, demonstrating another example of urban renewal based on cultural diversity. According to Sasaki, "the Korea town neighborhood and community [In Osaka's old downtown] still possesses an air of the warm and casual interpersonal relations that have long been considered a defining characteristic of the old downtown" (Sasaki 44). The old downtown of Osaka went through the financial crisis and doughnut phenomenon in the 1990s, when Osaka engaged in waterfront development and construction of the World Trade Center that eventually failed and drove the city to almost bankruptcy as a result of the end of the bubble era (Sasaki 42). Meanwhile, Koreatown, with its Korean culture that cannot be found in other parts of the city, remained appealing by mixing unique Korean culture and existing nostalgia of the old downtown. The success of Koreatown in Osaka hints at the possibility of revival of an old downtown with a combination of traditional culture and foreign culture. Considering the higher foreign population proportion in the old downtown in Cheonan, introducing foreign cultures such as Chinese and Vietnamese cultures to Korean cultures in the district may also be effective.

However, there are drawbacks to this method of urban renewal. Fainstein claims that if "[different] lifestyles are too incompatible, it only heightens prejudice." She then offers an example of Tompkins Square Park in New York, which is located in an "extremely heterogeneous neighborhood encompassing new gentrifiers; older, white working-class Eastern Europeans; self-styled anarchists, artists and other Bohemians; students; and bikers" that ended up with "loud noise and raucous behavior." She concludes that "exposure does not always breed acceptance—it can, in fact, produce mutual loathing" (Fainstein 13). Her claim that exposure to drastically different cultures may lead to disapproval may also be valid in the case of Cheonan. Although the old downtown possesses significantly higher proportions of foreigners, the majority of the population is Korean. This means urban renewal can ultimately end up with a failure if they start to consider multicultural elements in the district as "incompatible," which is probable because of the low magnitude of cultural diversity in Korea.

Advantages and Disadvantages of Introducing Pre-Existing Cultures

Considering this glaring drawback, experts in Korea have experimented with introducing Korean traditional culture to urban renewal instead of cultural diversity. Using traditional cultural elements in urban renewal has been effective in the Korean cities of Jeonju and Changwon. Jeonju used Hanok, Korean traditional houses, in the old downtown to construct Jeonju Hanok village, which became so famous that it is described as "one of the representative Korean traditional Hanok villages" (Hwang 82). This case successfully used existing local culture to turn the old downtown into a lively attraction. However, this method also has its disadvantages. Hwang explains that such methods have limitations including "finding of cultural contents to industrialize them" and maximizing "ripple effects to their periphery areas based on each main district" (Hwang 82). This means that the city should already have some Korean cultural attractions, such as heritage sites or traditional villages, to renovate and advertise to renew the entire district. The cultural attractions should be impactful enough to expand the effects of that content to the entire area. In Jeonju's case, this was not a problem because it was the major administrative city of Joseon, the dynastic kingdom that existed on the peninsula before Korea became a modern nation. Thus, Jeonju had more than enough cultural content for this type of urban renewal. However, Cheonan is unlikely to have such impactful cultural attractions in the downtown area because of its relatively short history. Cheonan experienced rapid growth after South Korea was established, as it turned from a county to a city in 1963 ("History"). Therefore, it is unlikely that Cheonan would have enough cultural contents and heritage that can be industrialized for urban renewal.

Conclusion

Given that both introducing cultural diversity and making use of existing culture in urban renewal have advantages and disadvantages, it is best to use a hybrid approach for urban renewal in Cheonan. That is, future renewal projects for the old downtown area should include both pre-existing cultural elements and cultural diversity. By using a combination of Chinese and Vietnamese elements as well as Korean traditional culture to renew the streets, establish community centers, and grow the district into live attractions, the district can succeed in urban renewal and become a diverse community while still possessing the identity of Korea. This hybrid method can reduce the disadvantages of only using cultural diversity because Korean residents will be reassured by the familiar Korean cultural elements. The disadvantage of only using traditional culture can also be mitigated by decreasing the reliance on traditional cultural content that Cheonan lacks. In conclusion, both traditional cultures and foreign cultures have to be introduced in urban renewal in order to revive the area into an inclusive and attractive community.

Works Cited

"2024년6월말 인구현황" ["Population status at the end of June 2024"]. *Cheonan City*, July 2024, www.cheonan.go.kr/prog/stat/sen_pop/list.do?siteCode=stat&mno=sub02_01_01. Accessed 3 July 2024.

Fainstein, Susan S. "Cities and Diversity: Should We Want It? Can We Plan for It?" *Urban Affairs Review*, vol. 41, no. 1, Sept. 2005, pp. 3-19, https://doi.org/10.1177/1078087405278968. Accessed 3 July 2024.

Hwang, Kyu Hong. "Finding Urban Identity through Culture-led Urban Regeneration." *Journal of Urban Management*, vol. 3, nos. 1-2, 2014, pp. 67-85, https://doi.org/10.1016/s2226-5856(18)30084-0. Accessed 3 July 2024.

Kim, Ji Youn. "Cultural Entrepreneurs and Urban Regeneration in Itaewon, Seoul." *Cities*, vol. 56, July 2016, pp. 132-40, https://doi.org/10.1016/j.cities.2015.11.021. Accessed 3 July 2024.

Lee, Je-Yong. "Effects of Cheonan Station around the Festival Program on the Local Commercial Activation." 한국철도학회 학술발표대회논문집, 15 Apr. 2018, pp. 449-51, www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07455013. Accessed 3 July 2024.

Ritchie, Hannah, et al. "Urbanization." *Our World in Data*, September 2018, https://ourworldindata.org/urbanization. Accessed 1 July 2024.

Sasaki, Masayuki. *Urban Regeneration through Cultural Diversity and Social Inclusion*. Faculty of Fine and Applied Arts, Chulalongkorn University, 2011, https://doi.org/10.14456/JUCR.2011.2. Accessed 3 July 2024.

Sung, Min-Ho, and Heewon Lee. "Schematic Regeneration Strategy of Old Downtown, Myeongdong, in Cheonan." *Journal of the Korea Academia-Industrial Cooperation Society*, vol. 15, no. 5, 31 May 2014, pp. 3231-39, https://doi.org/10.5762/kais.2014.15.5.3231. Accessed 3 July 2024.

"2035 천안도시기본계획" ["2035 Cheonan Masterplan"]. *Cheonan City*, Aug. 2022, www.cheonan.go.kr/kor/sub06_09_01_01.do. Accessed 3 July 2024.

"국내 거주 외국인주민 수 226만 명, 총인구 대비 4.4%, 최대규모 기록" ["2.26 Million Foreign-born Residents in the Country, 4.4% of the Total Population, the Largest Ever"]. *Republic of Korea Policy Briefing*, 8 Nov. 2023, www.korea.kr/briefing/pressReleaseView.do?newsId=156598606. Accessed 3 July 2024.

"연혁" ["History"]. *Cheonan City*, 3 July 2024, www.cheonan.go.kr/kor/sub04_01_01.do. Accessed 3 July 2024.

"외국인주민 및 다문화가구원 현황" ["Status of Foreign-born and Multicultural Households"]. *Cheonan City*, www.cheonan.go.kr/kor/sub06_07_01_01.do. Accessed 3 July 2024.

Beyond Borders: America's Global Conquest By Olivia Weller

Americans expanded westward across the rugged continent throughout much of the 19th century. By 1890, however, the census declared that, with so many pioneering the West, there was no longer a new frontier.[14] But, this was not the end of American expansion. With European nations pursuing empires around the globe in the late 1800s, America shifted its focus to acquiring overseas territories. This period of imperialism was motivated by a desire to compete with Europe for global economic and military power. America also believed it had the right and duty to "civilize" the inhabitants of foreign lands. While global power was achieved, civilizing other nations was morally unjust. In both cases, it was an appeal to a new American masculinity embodied by the person most identified with America's imperialistic ambition, Theodore Roosevelt, that helped drive the country to embrace overseas expansion.

The key to greater economic and military power was securing naval outposts overseas. As articulated by naval officer Alfred T. Mahan in 1890, "no nation could prosper without a large fleet of ships engaged in international trade, protected by a powerful navy operating from overseas bases."[15] As Assistant Secretary of the Navy in 1897, Roosevelt emphasized this belief: "We need a first-class navy"[16] that can confront the European powers and "meet them on the seas, where the battle for supremacy must be fought."[17] The fight came the following year when America declared war on, and quickly defeated, Spain and acquired the territories of Puerto Rico, Guam, the Philippines, and parts of Cuba. That same year, America annexed the island of Hawaii, providing a strategic Navy base in the Pacific. In 1899, as Governor of New York, Roosevelt argued in favor of this overseas expansion, stating that if America did not seize the moment, European nations would "win for themselves the domination of the world."[18] With its new overseas territories and naval bases, America began to fulfill its objective of becoming a global military and economic power.

America's new global power was depicted in a political cartoon appearing on the cover of "Puck" magazine in 1901. The cartoon, "Columbia's Easter Bonnet", shows Columbia - a name that represents America - dressed as a naval officer in red, white, and blue, adjusting her fancy Easter bonnet as she looks calmly in the mirror. Her bonnet is a battleship named "World Power" with protruding cannons and billowing smoke that contains the word "Expansion." She uses a tiny military sword as a hat pin, and wears an ammunition belt. The image symbolizes American military strength through overseas expansion. By featuring a woman at home, rather than a soldier in battle, the cartoon projects society benefiting from cultural and economic opportunities acquired from expansion, suggesting that the nation should feel pride in embracing imperialism. Significantly, Columbia has a tiny purse with a dollar sign fastened to her waist, symbolizing that territorial expansion brings economic prosperity.

[14] Turner, page 1
[15] Foner, page 536
[16] Roosevelt, page 68
[17] Roosevelt, page 68
[18] Roosevelt, page 67

A second, less successful, justification for expanding overseas involved the belief in white American cultural and racial superiority, and the need to civilize foreign lands. President McKinley explained that it was America's duty to "uplift and civilize"[19] its "little brown [Filipino] brothers,"[20] and Roosevelt echoed this by stating that it is "the idlest of chatter to speak of savages as being fit for self-government."[21] If the Filipinos were permitted to govern themselves, Roosevelt feared the islands would be reduced to "savage anarchy"[22] causing a "hideous calamity to all mankind."[23] Thus, Roosevelt urged Americans to resist the bloody battle for Filipino independence in order to "advance the cause of civilization."[24]

The duty to civilize populations that Americans viewed as racially inferior was highlighted in the political cartoon called "The White Man's Burden" appearing in "Judge" magazine in 1899. The title comes from the 1899 Rudyard Kipling poem urging America to assume the responsibility of civilizing its overseas native populations. The cartoon depicts men representing Britain (John Bull) and the United States (Uncle Sam) carrying the "burden" of inhabitants from newly acquired territories on their backs. The two men climb a hill covered in boulders toward the peak featuring a glowing person with the promise of "Civilization" and offering "Education" and "Liberty." Each boulder displays a negative trait meant to apply to the native populations, including "barbarism," "ignorance," and "vice." This cartoon shows America rescuing the inhabitants of the overseas territories from their current way of life and delivering them to civilization. Uncle Sam has a medical red cross symbol on his arm, supporting that America was "saving" these inhabitants. This promise of civilization is covered in gold also symbolizing that America will lead the territories to economic prosperity. Notably, the cartoon features the populations of the Philippines, Cuba, Puerto Rico, and Hawaii as unruly savages with dark skin, emphasizing the perceived inferiority of these non-white races, and the "burden" of civilizing them.

Equally prominent, however, were political cartoons supporting the anti-Imperialist cause and highlighting the morally unjust notion of governing others without their consent and the irony of using acts of barbarism to eliminate barbarism. The cartoon, "The Harvest In The Philippines" that appeared in "Life" magazine in 1899 illustrates this perfectly. It shows Uncle Sam, fully armed, in the foreground with countless rows of dead Filipinos stretching into the horizon behind him. This cartoon highlighted the awful hypocrisy of destroying innocent lives in the name of saving them and reflected the misguided nature of this second justification for expansion.

[19] Foner, page 538
[20] Foner, page 540
[21] Roosevelt, page 66
[22] Roosevelt, page 67
[23] Roosevelt, page 67
[24] Roosevelt, page 67

The two motivations for imperialism arose in an American population that, at the turn of the 20th century, was suffering from a "masculinity crisis."[25] For most of the 1800s, wealth and power were distributed broadly and the vast majority of men were self-employed.[26] However, with the Industrial Revolution and rise of manufacturing, men were forced to become tenant farmers or wage laborers, and suffered through poor working conditions with little hope for advancement.[27] Consequently, men had reduced standing in their family because they could not provide as prior generations. They were also less able to "control the country" because "the power of manhood . . . encompassed the power to wield civic authority . . . and to shape the future of the nation."[28] Accordingly, men began to compensate for their loss of manliness by embracing the "strenuous" life where "virility", "cowboy novels", "hunting", and "he-man" pursuits like football, boxing, and bodybuilding were popular.[29] The ideal male image became "physical bulk and well-defined muscles,"[30] and, at this same time, negative terms such as "sissy" and "pussy-foot" were used to describe effeminate men.[31]

Roosevelt, a national hero for his bravery in the Spanish-American war, appealed to these manly qualities, and became the perfect promoter for the age of American imperialism. He encouraged expansion by appealing to the masculine desires of men, using words like "manly" and "virile" in contrast to words like "timid" and "weak." He stated that a nation that is "unwarlike and isolated . . . [will] go down before other nations which have not lost the manly and adventurous qualities."[32] He also argued that peace "breeds timidity" and if men did not choose to fight, the nation would lose its "fearless, virile qualities" leading to "shameful disaster."[33] Roosevelt criticized the "undersized man of letters . . . with his delicate, effeminate sensitiveness . . . [who] cannot play a man's part among men,"[34] and he dismissed arguments against imperialism as an "unwillingness to play the part of men,"[35] which he viewed as showing "ourselves weaklings."[36] Instead, he urged the country to join him in this "hard and dangerous endeavor" that would achieve national greatness."[37]

The crisis of masculinity that dominated American men in the late 1800s was susceptible to Roosevelt's unique appeal to manliness. This strongly affected the drive for overseas expansion. As a result, America was able to achieve global military and economic power. On the other hand, America's desire to civilize was morally misguided and rested on inappropriate notions of racial superiority.

[25] Bederman, page 152
[26] Bederman, page 153
[27] Bederman, pages 152 - 153
[28] Bederman, page 154
[29] Bederman, pages 152 + 154 + 155
[30] Bederman, page 154
[31] Bederman, page 155
[32] Roosevelt, page 66
[33] Roosevelt, page 66
[34] Roosevelt, page 65
[35] Roosevelt, page 67
[36] Roosevelt, page 67
[37] Roosevelt, page 67

Works Cited

Bederman, Gail. *Manliness and Civilization: A Cultural History of Gender and Race in the United States*.Chicago: UChicago Press, 1996.

Roosevelt, Theodore. "Remarks at the Lincoln Dinner of the Republican Club in New York City."13 February 1905. Republican Club, New York.

Foner, Eric. *Give Me Liberty! An American History Vol. 2*.New York: W. W. Norton & Company, 2023.

Turner, Frederick J. *The Frontier in American History*.New York: Dover Publications, 1996.

How and why did Nazi policies towards the Jews of Europe change between 1938 and 1942? By Olivia Weller

On November 13th, 1919, a thirty-year-old Adolf Hitler stood on a rickety wooden table in a Munich beer cellar before a crowd of 130 people. His voice rose over the clinking of beer steins. This was his first public speech and he used this opportunity to blame the Jews for Germany's defeat in World War I.[38] Two months earlier, he provided a written response to the "Jewish question,"[39] which was the concern held by many Germans that Jews were too influential in German society.[40] In particular, Hitler described two phases to his approach to answer the Jewish Question: He wrote that "anti-Semitism, as a political movement" required "systematic legal opposition and elimination of the special privileges which Jews hold . . . but its final objective must unswervingly be the removal of the Jews all together."[41] Echoing this letter, Hitler gathered the attention of the crowd in the beer cellar and told them that Germany "will carry on the struggle until the last Jew is removed from the German Reich."[42] Fourteen years later, in 1933, Hitler became the Chancellor of Germany and he began to construct his diabolical plan for the Final Solution to the Jewish Question. Initially, Hitler's Nazi party enacted laws that stripped Jews of their rights to fully participate in public life, government, culture, and in many professions.[43] In 1938, however, Nazi policies toward the Jews changed drastically. Hitler's push for Lebensraum (living space) for the German people, beginning with the annexation of Austria in 1938, and the conquest of neighboring territory that followed, brought millions of Jews under Nazi Germany's control. This accelerated the need for Hitler to implement the removal phase of his response to the Jewish Question and ultimately it led to the 1942 Wannsee Conference's Final Solution and the subsequent deportation of the Jews to the death camps.

During the early 20th century, anti-Semitism was prevalent throughout Europe. However, anti-Jewish attitudes in Germany were particularly extreme. Germany lost over 3 million soldiers and civilians in World War I and following the country's defeat, its economy was shattered. On top of this, the peace Treaty of Versailles forced Germany to pay burdensome reparations.[44] The Jews became the explanation for these hardships. Many Germans accused the Jews of not supporting the German war effort, and of only being interested in profiting from it.[45] Many Germans also spread the *Dolschstoss* myth, (translated to the "stab in the back" myth), the anti-Semitic theory that Jewish communists provoked labor unrest which caused a lack of aid

[38] Lucy S. Dawidowicz, *The War Against the Jews,* page 17.

[39] Dawidowicz, 16.

[40] Arno J. Mayer, *Why Did the Heavens Not Darken?* page 108.

[41] Quoted in Dawidowicz, 17.

[42] Quoted in Dawidowicz, 17.

[43] Dawidowicz, 58.

[44] "World War I: Aftermath," *Holocaust Encyclopedia*, United States Holocaust Memorial Museum.

[45] Dawidowicz, 45.

for the German army.[46] ** The Jews were convenient scapegoats for the Germans because of the historic anti-Semitism that was already prevalent in society. Hitler absorbed these anti-Semitic views and expanded on them, making the removal of the Jews from Europe the centerpiece of his solution to the Jewish Question. Hitler crystallized his anti-Jewish ideology in his 1924 autobiography, *Mein Kampf*.[47] In blaming the Jews for Germany's defeat in the war, Hitler wrote that Germany should have poisoned the Jews with gas, just "as hundreds of thousands of our very best . . . on the battlefield had to endure it."[48] Hitler then wrote that the elimination of the Jewish "traitors . . . might have saved many real Germans."[49] He also thought that Jews were dominating the press, cultural and artistic life, and they were producing "literary filth, artistic trash, and theatrical nonsense."[50] Additionally, Hitler said that Jews were carriers of the "virus"[51] of Marxism, which he described as "an intellectual pestilence worse than the Black Death."[52] While these ideas set out in *Mein Kampf* reflected a rabid anti-Semitism, it was Hitler's obsession with the concept of a "master race" that evolved into his desire for Lebensraum, and that ultimately led to the annihilation of the Jews. Hitler wrote that the master race was a "perfection of human existence, whereas the Jews were the embodiment of evil."[53]

Less than a decade later, in 1933, Hitler took the oath as Chancellor of Germany and began to enact laws to deprive the Jews of their rights. This was phase one of his ultimate plan to exterminate the Jewish population of Europe. Germany first passed a number of laws that excluded Jews from public life, government, culture, many professions[54] and from entertainment and the press.[55] The Law for the Restoration of the Professional Civil Service, passed on April 7, 1933, excluded Jews from government. Hitler extended this exclusion and prohibited Jews from teaching at universities, from being judges, and limited Jewish attendance at schools.[56] Additionally, Hitler created the Reich Chamber of Culture and National Press Law which excluded Jews from the entertainment industry and placed newspapers under state supervision and out from under Jewish influence.[57] These laws, eventually reaching four hundred, were the

[46] Dawidowicz, 14.

** The photographs are from the exhibit, "The Holocaust: What Hate Can Do" featured at The Museum of Jewish Heritage - A Living Memorial to the Holocaust, in New York City.

[47] Dan Stone, *The Historiography of the Holocaust*, page 176.

[48] Quoted in Mayer, 101.

[49] Quoted in Mayer, 101.

[50] Quoted in Mayer, 101.

[51] Quoted in Mayer, 101 - 102.

[52] Quoted in Dawidowicz, 12.

[53] Quoted in Dawidowicz, 19.

[54] Dawidowicz, 59.

[55] Dawidowicz, 58 - 59.

[56] Dawidowicz, 58 - 59.

[57] Dawidowicz, 58 - 59.

first anti-Jewish regulations in over 60 years.[58] Importantly, these laws displayed Hitler's promise, made in 1919, that the Jewish Question would first require a legal response.[59]

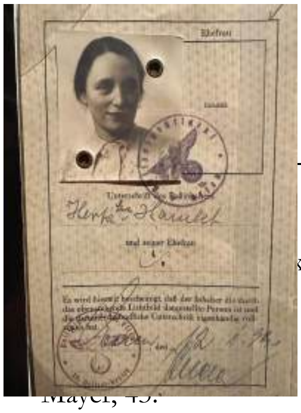
The laws against the Jews became more severe with the passing of the infamous Nuremberg Laws of 1935. These laws were designed to protect the racial purity of those with German blood.[60] The Nuremberg laws focused on inequalities based on racial differences, and deprived the Jews of many rights. For example, the Reich Citizenship Law permitted citizenship in the German Reich[61] only to those who were racially pure. This law specifically rejected the equality of men, and instead embraced the inequality of men based on racial differences.[62] The Nazi party was obsessed with racial purity and enacted another Nuremberg Law called the Law for the Protection of German Blood and German Honor. This law defined a Jew as anyone with at least three full Jewish grandparents, or two full Jewish grandparents and who belonged to the Jewish religious community.[63] This thorough effort to identify Jews highlighted the Nazis' obsession with categorizing the Jews in order to separate them from society and legally define them based on ancestry. Additionally, under this same law, marriage between Germans and Jews was prohibited.[64] Overall, these laws effectively removed Jews from daily life in Germany in a virtual sense. This bureaucratic, legislative anti-Semtism was the completion of phase one of Hitler's answer to the Jewish Question. With Jewish influence now weakened, the Jews were vulnerable and marginalized. This set the stage for the second phase, which was the actual removal of the Jews from German life.

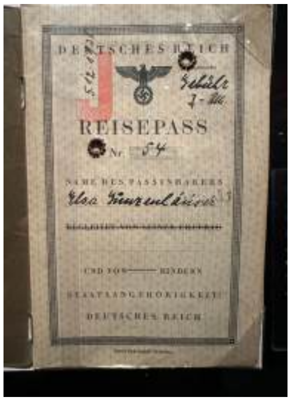
The Nazi policies towards the Jews changed in 1938 when Hitler began his push for Lebensraum in order to obtain new territory and resources for Germany. He understood that the German Reich would soon be expanding and millions of Jews would be under his control. Accordingly, the removal of the Jews from Europe, phase two of Hitler's solution to the Jewish Question, became a priority. In March of 1938, Hitler annexed Austria marking the start of Germany's territorial expansion. At the same time, the Nazis confiscated Jewish property and assets. The Nazis decreed that Jews must be removed from the economy, and all of their businesses and professions terminated by the end of the year.[65] Advancing this, every Jew had to list and assess all of their property followed by the liquidation of thousands of Jewish businesses.[66] Jewish children were prevented from attending school, Jews were forbidden to access most public places, and a curfew was implemented for them.[67] In November 1938, the Germans burned hundreds of synagogues, damaged thousands of Jewish stores, and tens of

[58] Mayer, 133 - 134.
[59] Dawidowicz, 17.
[60] Mayer, 149.
[61] The term "Reich" is the German word for "Empire" and Germany was commonly known during this time as the German/Third Reich. Britannica.com, Encyclopædia Britannica, 2023.
[62] Dawidowicz, 67.
[63] Dawidowicz, 68.
[64] Dawidowicz, 63, 68.
[65] Dawidowicz, 96 - 97.
[66] Dawidowicz, 95 - 96.
[67] Dawidowicz, 102 - 103.

thousands of Jews were arrested.[68] This planned attack was known as *Kristallnacht*, the night of broken glass. After this horrific event, a one billion Mark fine was imposed on the Jews ironically to pay for the damage.[69] By the end of 1938, Jews were excluded from the German economy,[70] and every aspect of Jewish life was disposed of "except the Jews themselves."[71] The Jews had their property, money, jobs, ability to go to school, and freedom of movement taken away from them. This was very different from the restrictions of rights in prior years. The Jews were now barely able to subsist. This step was carefully calculated by Hitler because, in order to remove the Jews from German society, he first had to transfer all Jewish businesses to German control and get rid of everything the Jews owned. This was all in preparation for Hitler's "removal of the Jews all together."[72]

In 1939, the vast increase of the Jewish population under German control accelerated the physical removal of the Jews from society and into ghettos. After the 1938 annexation of Austria, the population of Jews in the German Reich was approximately 750,000.[73] However, in 1939, Hitler took over part of Czechoslovakia and then invaded Poland (triggering World War II). This acquisition of territory increased the population of Jews under German control to almost four million.[74] With this increase, Hitler now had to address the Jewish Question with more urgency. Accordingly, the Nazis issued a directive in late 1939 called The Jewish Question in the Occupied Territory which moved all Jews to Poland and into ghettos.[75] The largest of these ghettos were Warsaw and Łódź, two cities in Poland where the Jews were locked in a labyrinth of squalid and crowded streets filled with malnutrition and disease.[76] In order to identify German Jews for transport to the ghettos, Jewish males were required to have "Israel" and Jewish females needed to have "Sarah" on their identification papers. Additionally, the Germans stamped Jewish passports with a red letter "J" to identify the passport holder as a Jew.[77] ** The purpose of the ghettos was to aid the Nazis in the steps leading to the Final Solution.[78] They were "temporary measure[s] to be implemented prior to the ultimate goal."[79] The installment of the ghettos allowed Hitler to concentrate the Jewish population and weaken them. This made it easier for the Germans to eventually export the Jews to the concentration camps.





xpulsion" by Frank Bajohr - quoted in Stone, 53.

Mayer, 43.

[75] Dawidowicz, 116.

[76] Mayer, 14.

[77] Dawidowicz, 97.

** The photographs are from the exhibit, "The Holocaust: What Hate Can Do" featured at The Museum of Jewish Heritage - A Living Memorial to the Holocaust, in New York City.

[78] "Ghettoization" by Tim Cole - quoted in Stone, 74.

[79] "Ghettoization" by Tim Cole - quoted in Stone, 74.

Hitler's final push for Lebensraum, after conquering much of Europe, was invading Russia to eliminate Jewish Bolshevism in 1941. The German army encountered millions of Russian Jewish civilians, and for the first time, the Nazis committed mass murder of Jews. Approximately four million Jews fell into the hands of the German army.[80] The Einsatzgruppen were a paramilitary force that followed the German soldiers in their invasions and slaughtered Jews along the way. These forces were specifically indoctrinated into the teachings of Hitler, were trained to "fight world Jewry as one has to fight a poisonous parasite,"[81] and were told that the Jews must be "wiped out in accordance with the [Hitler's] aims."[82] Hundreds of thousands of Jews were executed during the first few weeks of the Russian invasion. For example, in June of 1941, over the course of four days, 3,800 Jews were murdered in Kovno, Lithuania. Hundreds of synagogues and Jewish homes were set ablaze.[83] Several months later, the Einsatzgruppen slaughtered over 33,000 Jews in Kiev, Ukraine.[84] This was the first outbreak of large scale murder and violence towards the Jews in connection with Hitler's solution to the Jewish Question. The violence of these mass public executions would soon be replaced by the systematic horror of the genocide in the concentration camps.

While the Jews were being rounded up and murdered by the Einsatzgruppen in the East, the Nazis were preparing the concentration camps for the mass extermination of Europe's Jews which numbered approximately 8 million under German control.[85] In the summer of 1941, Hitler gave orders that "the occupied Eastern territories [were] to become free of Jews."[86] Shortly thereafter, the construction of the death camps started and the "killing with showers of carbon monoxide" was explicitly discussed.[87] The Nazis addressed the coordination and implementation of the fate of the Jews a few months later during the Wannsee Conference in Berlin.[88] At this conference, Reinhard Heydrich, a high ranking Nazi official, spoke of the approaching "Final Solution of the Jewish Question"[89] and said that to implement it, the Nazis would need to "comb Europe from west to east" to cover a total of over eleven million Jews.[90] Heydrich's superior, Heinrich Himmler, wrote immediately after the Wannsee Conference to "prepare the concentration camps," which would receive "great economic contracts and assignments" in the coming months.[91] In March of 1942, the first Jews arrived at Auschwitz, which led to Joseph Goebbels, a Nazi official, to write that the Jews "are now being evacuated eastward. The procedure is a pretty barbaric one and not to be described here more definitely. Not much will

[80] Mayer, 254 - 255.
[81] Dawidowicz, 114 - 115.
[82] Dawidowicz, 125.
[83] Dawidowicz, 259.
[84] Dawidowicz, 268.
[85] Mayer, 255.
[86] Quoted in Dawidowicz, 129.
[87] Dawidowicz, 131.
[88] Mayer, 290.
[89] Quoted in Dawidowicz, 136.
[90] Quoted in Mayer, 304.
[91] Quoted in Mayer, 311.

remain of the Jews."[92] After the Wannsee Conference in Berlin and the construction of the death camps, the Jews' fate was sealed. By the end of 1942, Hitler's second phase of his solution to the Jewish Question, the removal of the Jews, was in motion.

The anti-Semitism in Germany following World War I was reflected in the line of a popular children's book of the day: "Without solution of the Jewish question // No salvation of mankind."[93] This widespread anti-Semitism allowed Hitler's extreme and irrational views towards the Jews to flourish. However, Hitler's delusional hatred of the Jews did not, by itself, lead to the genocide of the Jews. Rather, the Final Solution would not have occurred without Hitler believing that the Germans were a master race entitled to more living space. It was this territorial expansion for Lebensraum, beginning in 1938, that brought millions of Jews to Hitler. This then led to the change in Nazi policies towards the Jews. For Hitler, legal deprivation of rights was no longer a sufficient solution to the Jewish Question. Instead, "the removal of the Jews all together"[94] was, for Hitler, the only practical option. In other words, but for Hitler's desire for Lebensraum, it is likely that the Holocaust, although horrific, would have been confined to Germany.

[92] Quoted in Dawidowicz, 139.
[93] Dawidowicz, 165.
[94] Dawidowicz, 17.

Works Cited

Dawidowicz, Lucy S. *The War against the Jews, 1933-1945*. New York: Holt,Rinehart and Winston, 1975.

Mayer, Arno J. *Why Did the Heavens Not Darken?: The 'Final Solution' in History*.New York: Pantheon Books, 1988.

Stone, Dan, ed. *The Historiography of the Holocaust*. Houndmills [England]:Palgrave Macmillan, 2004.

“Third Reich.” Accessed May 14, 2023 https://www.britannica.com/place/Third-Reich

“World War I: Aftermath.” Accessed April 24, 2023 https://encyclopedia.ushmm.org/content/en/article/world-war-i-aftermath

Childhood Obesity during global COVID Pandemic and Preventive strategies: Time for action By Shriya Katukuri 1, Neelima Katukuri , MD 2

Key Words: Obesity, childhood preventive disorders, diabetes

Abstract:

Worldwide public health challenges brought about by the COVID-19 pandemic contributed to childhood obesity. The objective of the study is to examine the effects of the pandemic on the prevalence of childhood obesity by examining the contributing factors and related risk factors. Important discoveries show an alarming pattern of rising childhood obesity during the COVID-19 pandemic, which is linked to decreased physical activity and limited access to healthy foods. Prolonged periods of lockdown lead to reduced physical activity and social interaction, causing children to consume calorie-dense foods and increasing the risk of obesity. Using insights from healthcare providers and quantitative data, the review will contain a holistic understanding on the factor that drove childhood obesity during COVID-19. The COVID-19 pandemic has increased the growth of childhood obesity, requiring immediate public health measures. To decrease long-term health effects in children, strategies should emphasize encouraging physical activity and enhancing nutrition accessibility.

Introduction:

During Covid, many children began to gain weight and disrupted different aspects of their life. Children and adolescents gain weight due to disrupted food patterns. Kids didn't know what to do during their time off, as they had much more free time than before the pandemic hit. Not only did children gain weight more rapidly, but studies also show there was an increase in stress, limited outdoor activity and more screen time. Families had less and less access to nutritious foods causing problems for the household, especially for children. COVID-19 highlights the importance of addressing childhood obesity. Childhood obesity is contributed to several factors including behavior, genetics, and community circumstances such as access to healthy food and safe places for physical activity. In fact, long lasting habits like healthy eating habits and practicing yoga might have prevented this obesity epidemic.

Prevalent Causes of Childhood Obesity during COVID-19 and Preventative Strategies:

According to the CDC, a study of 432,302 children ages 2 to 19 years found the rate of Body Mass Index (BMI) doubled during COVID-19 pandemic compared to the pre-pandemic period in children with overweight or obesity and younger school-aged children. Based on the initial BMI, obesity prevalence was 16.0% including 4.8% of persons who were severely obese, in August of 2020, the estimated percentage of obesity among people ages 2-19 was 19% (Simonnet, A. et al. 2020). This not only shows how there are flaws in public health but also the effects of leaving children at home alone during pandemic. During school before the pandemic, children were restricted to an extent and there was a certain schedule of when they couldn't and could eat. This highlights serious shortcomings in the pandemic's health promotion initiatives, which were made worse by disturbed habits and a rise in inactivity at home. Using methods based on evidence is crucial to finding creative solutions to these problems. Regular BMI screenings can assist identify children who are at-risk early on and enable prompt interventions in healthcare settings. Furthermore, it is essential to implement programs that are specifically designed to encourage physical activity and healthy eating habits in homes.

Obesity is related to unhealthy lifestyle, eating a lot of food containing excessive fat, Intake of sugary drinks, excessive consumption of simple carbohydrates like pasta, brown rice, grains, vegetables and raw fruits, too little sleep and lack of physical activity i.e., walking, regular exercises and yoga (Bjarnadottir, A. et al. 2017).

The prevention of this condition is a daunting task, but we cannot remain complacent and expect all overweight children to outgrow it or we might find ourselves facing even more alarming statistics. Preventive strategies include screening by health care providers for BMI, increased access to evidence-based pediatric weight management programs, identifying social determinants of health, and food assistance resources. School resources to facilitate healthy eating, physical activity, and chronic disease prevention (US Department of Agriculture. 2020).

In addition to helping you burn calories, yoga can improve the tone and mass of your muscles. Yoga may help with joint pain, allowing you to boost your everyday activities and exercise regimen (Anekwe, C. et al. 2021). Body weight is not enough to treat the body by special diets, technique, changes in lifestyle. It is also necessary to work on the level of consciousness, which is mental energy, or a scientific system for developing our body and for the expression of consciousness. (Nieman, D. C. et al.. 2019) Therefore, in yoga we do not exercise the body for the sake of burning extra calories, but to develop body awareness, to understand the language of our body, the way it works, and what suits it best the needs of our body and mind (Jones, A. W. et al. 2019).

Discussion:

Obesity in childhood persists into adulthood predominantly when there is a strong genetic component. According to the US Department of Agriculture, systemic changes in food economies combined with inflation have severely restricted the availability of healthy foods, leading to almost 33.8 million people facing food insecurity in 2021. The pandemic had a detrimental effect on nutrition quality and eating habits. Children's snacking between meals and emotional overeating increased in the spring of 2020 (Monica L. et al. 2024). The complications in adulthood are well-known with estimation of annual cost related to obesity around $100 billion. Childhood obesity carries its own morbidity related to type 2 diabetes, which is now the most common type of diabetes diagnosed in several pediatric diabetic centers. Treatment of obesity in children and the relative comorbidities is expensive or lengthy and generally only effective if the whole family is involved, even then it’s not curative. Hence a nationwide population-based approach to prevention of childhood obesity is essential.

Conclusion:

COVID -19 pandemic emphasizes the importance of acting on childhood obesity to protect and promote healthy lives for all children. During the pandemic, involvement in public health decreased. These findings also show how public health should increase access to efforts on promoting healthy behavior in the future to prevent situations like this from happening again. Government agencies, healthcare professionals, parents and caregivers, and others can take steps to reduce childhood obesity. By implementing specific actions of encouraging healthier eating habits, innovating to promote physical activity and prioritizing health equity, we can foster healthier futures for all children.

Works Cited

CDC National Center for Health Statistics (NCHS) data brief

Simonnet, A. et al. (2020). High prevalence of obesity in severe acute respiratory syndrome coronavirus‐2 (SARS‐CoV‐2) requiring invasive mechanical ventilation. *Obesity*.

US Department of Agriculture and U.S. Department of Health and Human Services. Dietary Guidelines for Americans, 2020-2025. 9th Edition. December 2020.

Nieman, D. C. et al.. (2019). The compelling link between physical activity and the body's defense system. Journal of sport and health science, 8(3), 201-217.

Jones, A. W. et al. (2019). Exercise, Immunity, and Illness. In Muscle and Exercise Physiology (pp. 317-344). Academic Press.

Monica L. et al. (2024). *Child health behaviors and obesity after COVID-19*. JAMA Pediatrics. https://jamanetwork.com/journals/jamapediatrics/fullarticle/2815511#:~:text=As%20COVID%2D19%20captured,to%2022.4%25%20in%20August%202020.

Anekwe, C. et al. (2021). *Yoga for weight loss: Benefits beyond burning Calories*. Harvard Health. https://www.health.harvard.edu/blog/yoga-for-weight-loss-benefits-beyond-burning-calories-202112062650

Bjarnadottir, A. et al. (2017). *Why Refined Carbs Are Bad For You*. Healthline. https://www.healthline.com/nutrition/why-refined-carbs-are-bad#:~:text=On%20the%20other%20hand%2C%20refined

Assessing the Diagnostic Effectiveness of CNN and MobileNetV2 in Detecting Brain Cancer from MRI Scans By Nopparut Rukkha-anankul

Abstract

Brain cancers have to be detected and removed early to improve patients' survival rates as it wouldn't have spread around as much. However, there are multiple places that have a shortage of doctors, especially in rural areas and developing countries. This leads to many patients going undiagnosed or misdiagnosed. This study aims to understand how accurate and reliable a CNN model can be when diagnosing brain cancer and also compare between two models to see which is the better choice for usage in future healthcare systems. This is so that these algorithms can act as radiologists for diagnosis in areas without specialised doctors. This study compared the two models, Convolutional Neural Network (CNN) and MobileNet, using accuracy as the deciding factor. The publicly available dataset included 7,023 MRI scans that has 4 types of scans: no tumour, pituitary tumour, glioma, and meningioma. 2000 were no tumours, 1,757 were pituitary tumours, 1,621 were gliomas, and 1,645 were meningiomas. The CNN model achieved a 98.82% accuracy, and the accuracy of MobileNet was 95.65%. This indicates that the CNN model might be a better approach, compared to the MobileNet, for brain tumour diagnosis. However, the CNN model should be used to assist radiologists for more accurate diagnosis. AI can be expected in healthcare systems in the future and will be more prevalent in the future. However, the models are not going to be replacing doctors any time soon.

Keywords: Convolutional Neural Networks; MobileNet; Brain tumour; Brain cancer

Introduction

Tumours can be either malignant (cancerous) or benign (Farias et al.). A brain tumour appears as a lump of abnormally growing tissues that presses against the spinal cord and disrupts the brain's functioning. If brain cancer is detected at an early stage where it hasn't developed and spread as much, the survival rate of the patient will be higher (Farias et al.). There are three main types of tumours including gliomas, meningiomas, and pituitary. Gliomas are usually malignant tumours that form from glial cells which are aggressive, and prone to recurrence, making them challenging to treat. Meningiomas are usually benign tumours forming from the brain's protective covering, they are less aggressive and have a better prognosis compared to malignant brain tumours (Ogasawara et al.). Pituitary tumours form in the pituitary gland and are also generally benign and can often be effectively addressed with medical or surgical treatments, with a generally good prognosis. However, there is a chance that benign tumours may become malignant over time.

Brain tumours are very rare compared to other types of tumours and only 35.9% of brain tumours are malignant (Voisin et al.), but they have high case fatality rates. Brain cancers rank as the third leading cause of cancer-related death in males aged fifteen to fifty-four and the fourth leading cause in women aged fifteen to thirty-four (Walker et al.). The treatment of malignant brain tumours can take various forms, such as surgery for tumour removal, radiotherapy or chemotherapy. The classic method of human radiologists diagnosing scans has some limitations, such as delays and errors in diagnosis (Farias et al. ; Singh et al.). These limitations can unfavourably affect patient outcomes, showing the need for a faster and more accurate method for diagnosing brain tumour scans. The integration of artificial intelligence (AI), especially convolutional neural networks (CNNs), into medical diagnosis can be a solution to overcoming these limitations (Abhishek). CNN is a form of deep learning model specialised for image classification, and MobileNetis a type of CNN designed for devices with smaller computing power. Both offer potential solutions to improve diagnostic accuracy.

Research suggests that AI is becoming more popular in healthcare (Shaheen). Therefore, will AI be used in the near future in healthcare systems? Both CNN and MobileNetV2 have been tested out in image classifications, with both achieving very high accuracies (Mohammad Monirujjaman Khan et al. ; Khan et al. ; Haq et al. ; Khasoggi et al. ; Samee et al. ; Arfan et al.). However, when looking at diagnosing brain tumours, CNN has a higher accuracy compared to MobileNet. Hence is CNN more effective than MobileNet in diagnosing brain cancer? With the development of these deep learning models and their high accuracy, there is a potential that they can be used along with radiologists. This leads to the question of whether CNN and MobileNet can replace radiologists in the diagnosis of brain cancer in MRI scans?

Literature Review: Traditionally used methods in diagnosis by radiologists

There are a total of seven different medical imaging techniques available: X-ray, computed tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), Single-photon emission computed tomography (SPECT), digital mammography, and ultrasound (Hussain et al.). With X-rays, the devices emit X-rays towards the patient, forming an image on an X-ray film. These images help to directly see medical conditions in the bones or joints. Similar to X-rays, CT scans produce 3-D scans using X-rays that are emitted from different angles, and computer processing. Physicians can evaluate the CT scans using computers. PET scans work by detecting radioactivity, which is discharged from a small dosage of radioactive tracer that was injected into the body. MRI scans use strong magnets to align ions within the body. The alignment is disrupted by radiowaves and the ions return to their initial position; the machine detects the change and an image is created. MRI is commonly used for tumours. Another scanner, known as SPECT emits, gamma rays to generate 3D images, which is mostly used for detecting very small changes in any part of your body. Mammography uses lower-frequency X-rays for the detection of breast cancer. Lastly, an ultrasound emits high-frequency waves to visualise the internal body. The machine will generate an image of the internal body, in which the doctor can see the internal body as the doctor moves the device along the body.

Some of the methods have limitations, starting with Computed tomography (CT). CT carries small risks when numerous scans are used. Younger people are particularly vulnerable to this method of scanning because CT scans use a lot of ionising radiation, which can increase the risk of cancer (Hussain et al. ; Pearce et al.). Leukaemia and brain tumours have been associated with CT scan radiation exposure, so CT scans are used for tumours or internal damage and not many other cases. Single-photon emission computed tomography (SPECT) is costly, requires careful radioactive material, and like PET, involves ionising radiation, posing radiation-related side effects. Ultrasound in the medical field has many applications but also possible limitations, such as hormone changes, and chromosome damage at low frequencies. Furthermore, research shows that due to the complexity of diagnosing diseases, doctors can make mistakes in diagnosis. Such mistakes can lead to delayed or incorrect treatment, which affects a patient's outcome (Singh et al.). To support the statement before, research was conducted and it showed that diagnostic error rates are approximately 10-15% (Graber). The percentage of diagnostic errors is emphasised by the fact that around 795,000 people lose their lives or become permanently disabled each year due to these mistakes (O'Mary).

MRI can be very costly and the machine takes a relatively long time to scan the patient compared to other methods (Hussain et al.). However, even though MRI is expensive, it is considered the safest method for diagnosing cancer. This is due to the fact that people are worried about how other methods being used that involve radiation, such as X-rays or CT scans, may bring risks of additional harm to the patient's body due to radiation exposure (Kauppinen and Peet). Because of these characteristics, MRI, which doesn't use radiation, is the best choice for the detection and visualisation of cancer.

An Introduction to Convolutional Neural Network

The Convolutional Neural Network (CNN) has revolutionised deep learning, particularly in computer vision, enabling new features like autonomous vehicles. (Alzubaidi et al.). It works by utilising convolutional layers with small filters that slide across the input image, detecting features like edges and textures. Each convolutional layer is followed by optional pooling layers that downsample feature maps and an activation function (Alzubaidi et al.). After several convolutional and pooling layers, the feature maps are flattened into a single vector and passed through fully connected layers, which are responsible for learning more detailed representations and making final decisions. The output layer has a loss function that calculates the error rate that shows the difference between the real input and the predicted output in the training dataset. During training, CNNs use backpropagation and the error is calculated to adjust weights that help the algorithm reach the decision.

To learn about these deep learning models' potential, a study researched the potential of AI in diagnosing medical scans. The researchers trained a CNN model on actual Traumatic Brain Injury patients (Wu et al.). Their evaluation is based on metrics such as the dice similarity coefficient. A group of 828 patients underwent the experiment which revealed the promising ability of the mode with its accuracy reaching 95.67% in medical scan diagnosis. Additionally, a study attempted to confirm CNN's ability to diagnose by using it to detect kidney rejections post-transplant (Abdeltawab et al.). This research had a total of 56 subjects and the dataset was MRI scans of the patients. The research used a computer-aided design system based on a CNN classifier and was able to achieve an impressive 91% accuracy, 90% sensitivity, and 92% specificity.

Despite all the high accuracies and potential, CNNs have some limitations (Alzubaidi et al.). First, CNNs require huge amounts of data for effective learning due to their complex architecture. Hence when data is scarce, transfer learning can be used to allow the model to train well with a smaller dataset. This method uses existing knowledge from the pre-trained model, that has done a similar task before, and adapts to the new training data that is being given to the algorithm. Furthermore, data augmentation is also another method, particularly for image classification tasks. Doing things like translating, mirroring, and rotating scans can help broaden datasets, improving model adaptability. Secondly, some biological datasets contain class imbalances, biassing one class heavily because underdeveloped countries don't have enough datasets to have much effect on the algorithm. Such bias can lead to models excelling in diagnosing the majority class but underperforming the minority class. In which strategies such as up-sampling the minority class can help lessen the effect of class bias. CNN models' complexity often leads to a "black box" effect, which challenges transparency as doctors won't be able to understand how the algorithm came to the decision. In critical areas like cancer diagnoses, knowing how the algorithm thinks is important. Therefore, techniques like back-propagation-based methods can help tell a thing or two about how the weight decides, which can lessen the "black box" effect. Lastly, overfitting is also a problem and it is the risk of performing well on training data but poorly on new data due to too much training on the training dataset and failing to generalise to new data. This can be solved by techniques such as weight decay and less training epochs.

CNN in the diagnosis of brain cancer

Brain tumours are a global health concern and it has 3 main types, including Meningioma, Glioma, and Pituitary (Haq et al.). Detecting brain cancer early is crucial for effective treatment, so deep learning (DL) models, particularly convolutional neural networks (CNNs), have gained popularity for brain cancer diagnosis due to its exceptional ability to do image classification. These models work by extracting main features of brain cancer in scans for image classification. Many models have been tested and all of it were able to consistently achieve impressive accuracy rates ranging from 90% to 98% with some even achieving 100%, showing the high potential of CNNs being integrated into clinical practice.

Firstly, a study was conducted to assess the diagnostic capabilities of CNNs in brain tumour diagnosis, testing two proposed models, VGG16 and a 23-layer CNN (Khan et al.). These CNN models are designed to categorise them as meningioma, glioma, or pituitary tumours. The proposed VGG16 model achieved a remarkable 100% accuracy, while the CNN model achieved an accuracy of 97.8%. Additionally, acquiring more diverse datasets will be crucial for enhancing both the accuracy and reliability of these models. Furthermore, due to the fact that cancer diagnosis by radiologists is highly subject to errors, another study decided to put computer-assisted tools (CAT) to the test, and ML models and DL models, such as CNN, are being tested to see its accuracy and its ability to speed up the process in diagnosing cancer when the image of the MRI scans are imperfect (Farias et al.). The study used a dataset containing 8099 3D MRI images and made use of the Gaussian blur in order to blur the images. The CNN model used has been pre-trained by 100,000 images and 200 different classes and further trained as a Resnet34 architecture, where it takes residuals from each layer and uses them in subsequent connected layers. This is further trained using 80% of the dataset and the 20% was used as testing for the CNN model. Performance results were measured in terms of accuracy, precision, F1 score, and recall, with these values averaged over 10 simulation folds. The results of normal MRI images indicated that the Resnet34 architecture achieved high performance with minimal variation across the folds, with less than 2% misclassification. However, as the blurring or the scans intensified, the accuracy and the F1 Score started to drop. This shows how neither AI nor radiologists will be able to diagnose cancer if the image is very degraded due to noise or blurring.

Even recent studies support the older studies that CNN has a high potential of being integrated into the medical diagnosis system. Gupta et al. 's study has results similar to previous studies that CNNs can diagnose brain tumours with remarkable accuracy rates that are much higher than those of radiologists (Khan et al.). Gupta's CNN model was able to achieve an astounding 100% accuracy with a 100% recall rate as well, while the VGG16 model was only able to achieve an accuracy of 96% which is still very high. Lastly, another study used a multi-classification method by utilising a CNN model (Srinivasan et al.). The model was tested on three different tasks that range from basic to advanced tasks. The three tasks are detecting the presence of a tumour, classifying which brain tumour type it is, and lastly grading the tumour from level 1 to 5, 5 being the most severe cases. The accuracy percentages were 99.53%, 93.81%, and 98.56%, respectively. This shows how accurately the model has performed across all experiments. The limitations to all of the studies above were that public datasets were used for training and testing, therefore before using it in real-world settings, the researchers will have to implement experiments testing its ability in clinics and hospitals to ensure the model's reliability.

An introduction to MobileNet and its use in the diagnosis of brain cancer

A neural network model called MobileNets, a type of transfer learning technique, a model created from CNN models, uses depth-wise separable convolutions (Khasoggi et al.). This method divides the computation into two parts to improve overall efficiency by lowering the amount of computing needed, particularly in the early layers of the neural network. Because MobileNets are efficiency-focused in design, they can be deployed on low-processor devices like mobile phones and embedded devices. In a study, MobileNet was executed on a Raspberry Pi 3 which has relatively small computing power compared to an everyday laptop. Despite its small computing power, it achieves an accuracy of 92.6% (Khasoggi et al.). Moreover, another research revealed that the model requires only 56% of the time needed to train compared to other models (Rybczak and Kozakiewicz). However, with such speed, the model will have to sacrifice its accuracy along with its ability to differentiate high-level features due to its smaller architecture (Maiti).

MobileNet has been tested in the diagnosis of brain tumours from MRI images (Samee et al.). It is trained using MRI scans of healthy brains and brains with tumours. MobileNet shows great performance by achieving an accuracy of 99.51%, which shows its potential integration in the healthcare industry. Furthermore, a researcher uses transfer learning with the MobileNetV2 to train a CNN for brain tumour classification. The dataset used consists of 2475 MRI images and is divided into training and testing sets (Arfan et al.). Evaluation metrics, such as precision, recall, and F1-Score, were used to measure how well the MobileNet did, and the model achieved a testing accuracy of 94% (Arfan et al.). Further investigation was conducted on the use of CNNs, specifically MobileNetV2 and VGG19, in the diagnosis of brain cancer using X-ray and CT scans (Mohammad Monirujjaman Khan et al.). This was investigated due to the traditional methods of manual categorising, being inaccurate and slow. The use of pre-trained CNN models in conjunction with transfer learning has shown an increase in the accuracy of tumour classification. The MobileNetV2 achieved an accuracy of 97%, and this shows how precise it can perform in medical scan classification. The study clearly shows that CNNs have the potential to improve the accuracy and speed of brain cancer diagnosis.

The present study

For the past few years, doctors have constantly been using traditional methods for the diagnosis of multiple diseases, including brain cancer. However, there are some limitations to such methods as the literature suggests. For example, multiple doctors will have to work together to diagnose cancer, as one doctor’s diagnosis is not reliable and this results in inefficiency since doctors have to work for the whole day, they are more likely to make mistakes in the diagnosis if they’re tired, therefore making it even less reliable (Singh et al.). Furthermore, a lack of specialised expertise in specific regions worsens this problem. As a result, some people may not receive a thorough diagnosis, which can lead to inaccurate diagnoses. These delayed or incorrect diagnoses can affect patient outcomes in a bad way, especially in cases of diseases like cancer, where delayed treatment will decrease survival rates. AI is a good tool to use to overcome these limitations as past research suggests (Davenport and Kalakota). This is due to its ability to be much more efficient, accurate and consistent compared to doctors (Alzubaidi et al.). CNN is a deep-learning model that is the best option for the diagnosis of diseases due to its ability to classify images (Mohammad Monirujjaman Khan et al. ; Khan et al.). This is because CNN can be accurate and efficient in the diagnosis of cancer compared to doctors, with its accuracy constantly above 90% as long as the model has been trained with enough data (Mohammad Monirujjaman Khan et al. ; Khan et al. ; Haq et al. ; Rong et al. ; Rauschecker et al. ; Wu et al. ; Abdeltawab et al. ; Akter et al.). This is very important as cancer patients will have a higher recovery chance if cancer is detected early. With that being said, CNN will not be effective if it has not been trained with sufficient data and the image scans are too degraded. Furthermore, since CNN has been achieving over 90% accuracy, which is already higher than most diagnosis specialists, CNN could potentially replace or assist radiologists (Haq et al. ; Rauschecker et al.). In addition to the traditional CNN model, MobileNet offers higher efficiency and requires less computational power for operation. Nevertheless, it exhibits limitations in effectively training with large datasets (Khasoggi et al.).

The main purpose of this paper is to compare CNN and MobileNet based on accuracy in diagnosing brain cancer and whether or not it will replace the doctors in diagnosis. Specifically, our hypotheses are H1: AI will be a huge helping tool for doctors to use in healthcare systems. H2: CNN will be a better model compared to MobileNet based on its higher accuracy (Mohammad Monirujjaman Khan et al.). H3: It is hypothesised that eventually, CNNs would replace the doctors, however, currently, CNNs will have to work along with the radiologists (Davenport and Kalakota).

Methods

The dataset was taken from Kaggle and it contains 7023 images of human brain MRI images which are classified into 4 classes: glioma (severe) - meningioma (mild) - no tumour and pituitary (not serious). This dataset is publicly shared by Masoud Nickparvar, which can be opened by:
https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset

Approximately, 22% of the images were used as test data for the model and the rest as training data. The CNN model, similar to the dataset, is also taken from Kagle and is publicly available. The code's purpose is to delve into the analysis, classification, or visualisation of brain tumour data, by using deep learning models written in Python.
https://www.kaggle.com/code/abdallahwagih/brain-tumor/notebook

The CNN model program's image preprocessing is performed using TensorFlow's Keras library. The images are uniformly resized to 224 by 244 pixels and converted to the RGB colour mode. Data shuffling is applied to the training and validation datasets to prevent any order-based biases, and a batch size of 16 is used to allow efficient data processing. These preprocessing steps standardise the data for deep learning, enhancing the model's accuracy and generalisation capabilities during training and testing. The model will randomly select 50% of the testing dataset for its recall test.

The code for the MobileNet model was also taken from Kaggle, and like the CNN model, it is used to analyse brain scans using deep learning. However, this model is less complex compared to the CNN.
https://www.kaggle.com/code/slavenjabuka/cnn-98-84-mobilenet-v2-99-35/notebook

The image preprocessing is performed using TensorFlow's Keras library. The colour images undergo uniform resizing to dimensions of 224 by 224 pixels and are also converted to RGB colour mode. Additionally, a normalisation step was applied to the pixel values of the resized colour images. This is done by dividing each pixel value by 255.0, the resulting normalised colour images have consistent intensity levels within the range between 0 and 1. This is done to ensure that all pixel values in the images are within a consistent and standardised range. Unlike the CNN model above, this model will use all of the testing data in the dataset.

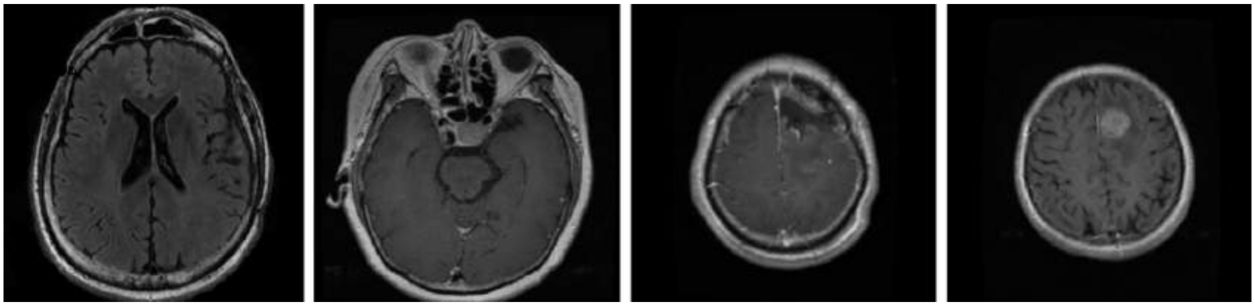


Fig 1: Examples from the dataset visualising what data is being used in the model. The figures show images of no tumour, pituitary, glioma and meningioma.

Results

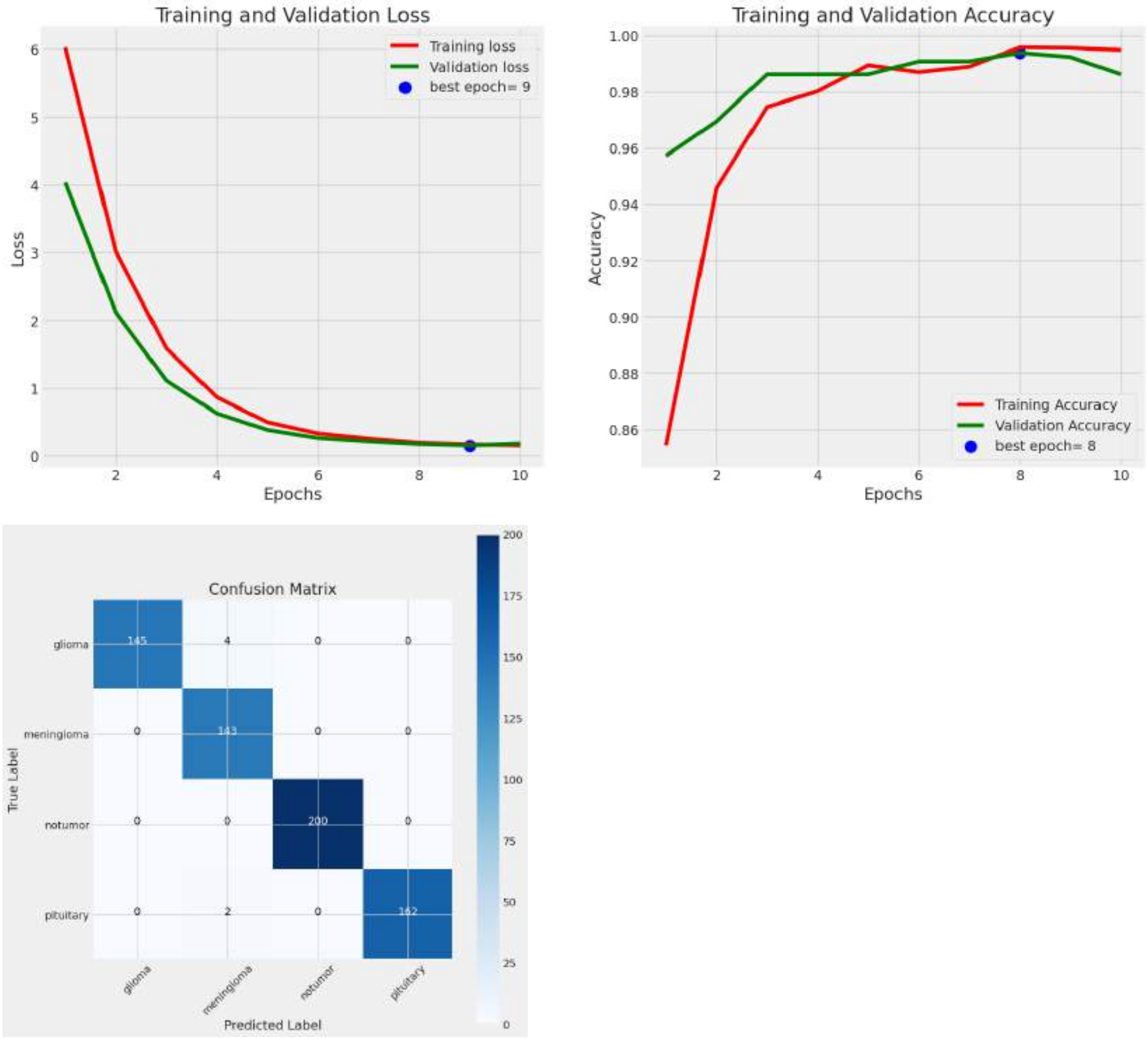


Fig 2: This figure shows two graphs, one on the left and one on the right, which represent the overview of the classification ability of the CNN model. The left graph shows training loss and validation loss over a number of epochs, while the right graph shows training accuracy and validation accuracy over a number of epochs. The confusion matrix also provides a visual

summary of the model's performance on a classification task. It shows the distribution of predictions across different categories.

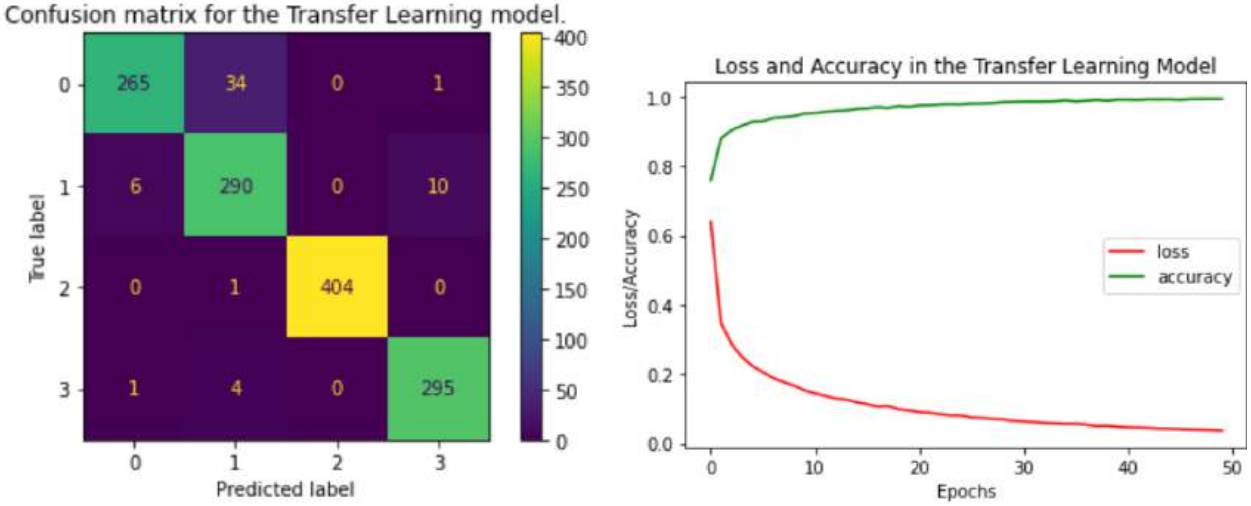


Fig 3:This figure shows the performance of MobileNet on a classification task, with 0,1,2,3 representing glioma, meningioma, pituitary and no tumour. The confusion matrix shows how well the model differentiates between different categories of brain tumours. Ideally, high values appearing along the diagonal will indicate accurate classifications. The loss and accuracy graph tracks the model's training process over epochs. As the model trains, the training loss should decrease and the training accuracy should increase. If the model is generalising well, the validation loss and accuracy curves should follow a similar trend.

Table 1: The performance of CNN and MobileNet for Brain Tumour Classification is shown in the table. The CNN model achieved an overall accuracy of 99%, which is higher than the MobileNet model's accuracy of 96%.
Precision: The proportion of positive predictions that were truly positive.
Recall: The proportion of actual positive cases that were correctly identified.
F1-Score: The harmonic mean of precision and recall.
Support: The number of data in each class.

CNN					
	Glioma	Meningioma	Pituitary	No tumour	Overall Metrics
Precision	1.00	0.96	1.00	1.00	0.99
Recall	0.97	1.00	0.99	1.00	0.99
F1-Score	0.99	0.98	0.99	1.00	0.99

Accuracy	-	-	-	-	0.99
Support	149	143	164	200	-
MobileNet					
Precision	0.97	0.88	1.00	1.00	0.96
Recall	0.89	0.95	0.99	0.98	0.95
F1-Score	0.93	0.91	0.99	0.99	0.95
Accuracy	-	-	-	-	0.96
Support	300	307	300	413	-

CNN Model Performance:

The results show that the CNN model, during the first few epochs, has relatively high training and validation loss, which shows that the model was not performing very well initially, which is probably due to not having enough training yet. As more epochs passed, the model's loss decreased tremendously. The accuracy of the model was inversely proportional to the loss, meaning that as the model's loss decreases, the accuracy increases. The accuracy seems to plateau off at around 99%. Furthermore, precision, recall, and F1-score metrics were consistently high across all tumour types, including glioma, meningioma, pituitary tumours, and cases with no tumour present. However, the CNN model made a few mistakes, including misclassifying four gliomas as meningiomas and one pituitary tumour as a meningioma. The error rate for glioma is 2.68% and 1.22% for pituitary.

MobileNet Model Performance:

The MobileNet model had a slightly lower overall accuracy of 96% in diagnosing brain tumours. While precision and recall metrics were generally high, the model had a higher rate of misclassifications than the CNN model. Specifically, 34 cases of Glioma were misclassified as Meningioma, 1 case of Glioma was misclassified as no tumour, 6 cases of Meningioma were misclassified as Glioma, 10 cases of Meningioma were misclassified as no tumour, 1 case of Pituitary tumour was misclassified as Meningioma, 1 case of no tumour was misclassified as Glioma, and 4 cases of no tumour were misclassified as Meningioma. Despite the efficient performance of MobileNet, which is evidenced by the rapid decrease in validation loss with minimal epochs, its simpler architecture limits its accuracy to a maximum of 96%.

Conclusions

The results show that the traditional CNN model achieves a higher accuracy at 98.82%, whilst the MobileNet achieved an accuracy of 95.65%. This happened despite the training epoch of the CNN being less than the Mobilenet's epoch. While overfitting is a potential concern, the consistently high accuracy throughout the final epochs, without any decline, suggests that the model may have adequately generalised to the test data and thus shows that the model is not overfitted. The reason that this could have happened would be that the CNN model was much more complex. This can be proven true as in the code for MobileNet, the model has 4 layers and the dense layer in this model has only 4 units, while the CNN model has 5 layers with the dense layer having 256 units. This made the time taken to train the CNN much longer than the MobileNet, as the CNN model with more units, will need to perform much more computations.

The results show that these models may surpass us humans in the diagnosis of cancer, which is accurate to almost every study about using CNN in the diagnosis of diseases as a radiologist's accuracy is averaged at around 86% (Mohammad Monirujjaman Khan et al.; Khan et al.; Haq et al.; Rong et al.; Rauschecker et al.; Wu et al.; Abdeltawab et al.). This means that it will certainly be used once it demonstrates consistent and reliable performance in real-world hospital settings. Currently, artificial intelligence might only be used in administrative tasks such as updating patient records and billing. However, AI will definitely be integrated into diagnosis and treatment prediction. For the diagnosis of diseases such as cancer, the doctors should opt for the traditional CNN as the accuracy of the model was higher than the MobileNet's because the accuracy is what is the most important when diagnosing diseases, not speed. Even though the accuracy of these models are incredibly high, it still can't fully replace radiologists or doctors in diagnosing diseases in the present. This is because AI still can't be fully relied on currently without further testing as there is the "black box" problem where the doctors can't see how the model is thinking and what decisions has it made to reach the final answer. Furthermore, doctors still have more experience and more expertise compared to the models. In scenarios such as diagnosing rare diseases or diseases that look very similar, the doctors will be the better choice for diagnosis. Therefore, the use of CNNs will need to be utilised with the collaboration of doctors in diagnosis for the time being.

The dataset used was too small to properly train a CNN model for real use in a hospital and it had only a few types of tumours. Hence, further research on a larger dataset is needed for more reliable results. Additionally, another limitation was that no information was provided from the scans on the malignancy of tumours. The models can be tested in real-life situations following these steps: 1-letting the model process the scan; 2-let the team of radiologists discuss and analyse the scan; 3-comparing the results between the 2 steps. After comparison, the information should be passed to the model for it to generalise and improve its accuracy. This process is important as it allows the model to repetitively learn new data from a new patient, thus allowing it to be adapted to real-world scenarios.

Works Cited

Abdeltawab, Hisham, et al. "A Novel CNN-Based CAD System for Early Assessment of # Transplanted Kidney Dysfunction." Scientific Reports, vol. 9, no. 1, Apr. 2019, https://doi.org/10.1038/s41598-019-42431-3

Abhishek, Kumar. "Introduction to Artificial Intelligence." Simple Talk, 10 May 2022, www.red-gate.com/simple-talk/development/data-science-development/introduction-to-artificial-intelligence

Akter, Atika, et al. "Robust Clinical Applicable CNN and U-Net Based Algorithm for MRI Classification and Segmentation for Brain Tumor." Expert Systems With Applications, vol. 238, Mar. 2024, p. 122347. https://doi.org/10.1016/j.eswa.2023.122347

Alzubaidi, Laith, et al. "Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions." Journal of Big Data, vol. 8, no. 1, Mar. 2021, https://doi.org/10.1186/s40537-021-00444-8

Arfan, T.H., Hayaty, M. and Hadinegoro, A, et al. "Classification of Brain Tumours Types Based on MRI Images Using Mobilenet." IEEE Conference Publication | IEEE Xplore, ieeexplore.ieee.org/abstract/document/9590183

Davenport, Thomas, and Ravi Kalakota. "The Potential for Artificial Intelligence in Healthcare." Future Healthcare Journal, vol. 6, no. 2, June 2019, pp. 94–98. https://doi.org/10.7861/futurehosp.6-2-94

Farias, Q., et al. "The Influence of Magnetic Resonance Imaging Artifacts on CNN-Based Brain Cancer Detection Algorithms." Computational Mathematics and Modeling, vol. 33, no. 2, Springer Science+Business Media, Apr. 2022, pp. 211–29, https://doi.org/10.1007/s10598-023-09567-4

Fusco, Roberta, et al. "Artificial Intelligence and COVID-19 Using Chest CT Scan and Chest X-ray Images: Machine Learning and Deep Learning Approaches for Diagnosis and Treatment." Journal of Personalized Medicine, vol. 11, no. 10, Sept. 2021, p. 993. https://doi.org/10.3390/jpm11100993

Graber, Mark L. "The Incidence of Diagnostic Error in Medicine." BMJ Quality & Safety, vol. 22, no. Suppl 2, June 2013, pp. ii21–27. https://doi.org/10.1136/bmjqs-2012-001615

Gupta, Manali, et al. "Classification of Brain Tumor Images Using CNN." Computational Intelligence and Neuroscience, vol. 2023, Oct. 2023, pp. 1–6. https://doi.org/10.1155/2023/2002855

Haq, Amin Ul, et al. "DACBT: Deep Learning Approach for Classification of Brain Tumors Using MRI Data in IoT Healthcare Environment." Scientific Reports, vol. 12, no. 1, Sept. 2022, https://doi.org/10.1038/s41598-022-19465-1

Hussain, Shah, et al. "Modern Diagnostic Imaging Technique Applications and Risk Factors in the Medical Field: A Review." BioMed Research International, vol. 2022, June 2022, pp. 1–19. https://doi.org/10.1155/2022/5164970

Kauppinen, Risto A., and Andrew C. Peet. "Using Magnetic Resonance Imaging and Spectroscopy in Cancer Diagnostics and Monitoring." Cancer Biology & Therapy, vol.

12, no. 8, Oct. 2011, pp. 665–79. https://doi.org/10.4161/cbt.12.8.18137

Khan, Md. Saikat Islam, et al. “Accurate Brain Tumor Detection Using Deep Convolutional Neural Network.” Computational and Structural Biotechnology Journal, vol. 20, Jan. 2022, pp. 4733–45. https://doi.org/10.1016/j.csbj.2022.08.039

Khasoggi, Barlian, et al. “Efficient Mobilenet Architecture as Image Recognition on Mobile and Embedded Devices.” Indonesian Journal of Electrical Engineering and Computer Science, vol. 16, no. 1, Oct. 2019, p. 389. https://doi.org/10.11591/ijeecs.v16.i1.pp389-394

Maiti, Agniva. “MobileNet V3 Model.” OpenGenus IQ: Learn Algorithms, DL, System Design, 16 Oct. 2023, iq.opengenus.org/mobilenet-v3-model

Mohammad Monirujjaman Khan, et al. “A Novel Approach to Predict Brain Cancerous Tumor Using Transfer Learning.” Computational and Mathematical Methods in Medicine, vol. 2022, no. 1, Hindawi Publishing Corporation, June 2022, pp. 1–9, https://doi.org/10.1155/2022/2702328

Ogasawara, Christian, et al. “Meningioma: A Review of Epidemiology, Pathology, Diagnosis, Treatment, and Future Directions.” Biomedicines, vol. 9, no. 3, Mar. 2021, p. 319, https://doi.org/10.3390/biomedicines9030319

O’Mary, Lisa. “Misdiagnosis Seriously Harms 795,000 People Annually: Study.” WebMD, 19 July 2023, www.webmd.com/a-to-z-guides/news/20230719/misdiagnosis-seriously-harms-people-annually-study

Pearce, Mark S., et al. “Radiation Exposure From CT Scans in Childhood and Subsequent Risk of Leukaemia and Brain Tumours: A Retrospective Cohort Study.” Lancet, vol. 380, no. 9840, Aug. 2012, pp. 499–505. https://doi.org/10.1016/s0140-6736(12)60815-0

Rauschecker, Andreas M., et al. “Artificial Intelligence System Approaching Neuroradiologist-level Differential Diagnosis Accuracy at Brain MRI.” Radiology, vol. 295, no. 3, June 2020, pp. 626–37. https://doi.org/10.1148/radiol.2020190283

Rong, Guoguang, et al. “Artificial Intelligence in Healthcare: Review and Prediction Case Studies.” Engineering, vol. 6, no. 3, Mar. 2020, pp. 291–301. https://doi.org/10.1016/j.eng.2019.08.015

Rybczak, Monika, and Krystian Kozakiewicz. “Deep Machine Learning of MobileNet, Efficient, and Inception Models.” ProQuest, vol. 17, no. 3, 2024, p. 96, https://doi.org/10.3390/a17030096

Samee, Nagwan Abdel, et al. “Classification Framework for Medical Diagnosis of Brain Tumor With an Effective Hybrid Transfer Learning Model.” Diagnostics, vol. 12, no. 10, Oct. 2022, p. 2541. https://doi.org/10.3390/diagnostics12102541

Shaheen, Mohammed Yousef. “Applications of Artificial Intelligence (AI) in Healthcare: A Review.” ScienceOpen Preprints, Sept. 2021, https://doi.org/10.14293/s2199-1006.1.sor-.ppvry8k.v1

Singh, Hardeep, et al. “The Global Burden of Diagnostic Errors in Primary Care.” BMJ Quality

& Safety, vol. 26, no. 6, Aug. 2016, pp. 484–94, https://doi.org/10.1136/bmjqs-2016-005401

Srinivasan, Saravanan, et al. “A Hybrid Deep CNN Model for Brain Tumor Image Multi-classification.” BMC Medical Imaging, vol. 24, no. 1, Jan. 2024, https://doi.org/10.1186/s12880-024-01195-7

Voisin, Mathew R., et al. “Incidence and Prevalence of Primary Malignant Brain Tumours in Canada From 1992 to 2017: An Epidemiologic Study.” CMAJ Open, vol. 9, no. 4, Oct. 2021, pp. E973–79. https://doi.org/10.9778/cmajo.20200295

Walker, David, et al. “Strategies to Accelerate Diagnosis of Primary Brain Tumors at the Primary–Secondary Care Interface in Children and Adults.” CNS Oncology, vol. 2, no. 5, Sept. 2013, pp. 447–62, https://doi.org/10.2217/cns.13.36

Wu, J. et al. “An Artificial Intelligence Multiprocessing Scheme for the Diagnosis of Osteosarcoma MRI Images.” IEEE Journals & Magazine | IEEE Xplore, ieeexplore.ieee.org/abstract/document/9802666

Elements in a Local Economy: A Tale of Two Towns By Eshaan Khera

Abstract

This article attempts to investigate the key aspects that drive the local economy of a town or city. Although it is easier to list the usual elements, this study aims to identify the key factor that influences other factors. To remove the noise of variables such as geographic profile, I chose two neighboring towns in the same county with significantly different economic profiles. By evaluating factors such as ethnic makeup, income, property prices, education, and crime, we can observe how some of these factors interact, resulting in an amplifying effect that leads to a shift to the two extremes of the economic spectrum. Based on these findings, it is reasonable to assume that, of all the potential variables, a town's ethnic makeup has the most direct or indirect influence on its economy. I also corroborated this finding by using data from neighboring counties in New Jersey to compare poverty rates by ethnicity. One of the proposals for reducing the economic disparity is to diversify these towns. Higher rents and property values are one of the most significant hurdles to doing so. For affluent towns like Westfield, a campaign to boost the number of affordable housing units, particularly if mandatory for all new developments, will help attract new people who cannot otherwise afford to live there. Similarly, municipalities that require external financial stimulus, such as affluent inhabitants and new enterprises, can enact legislation that encourages investment and raises the town's prestige. One caveat is that while doing so, an effort must be made to ensure that the negative aspects of gentrification are addressed. Gentrification, or the influx of money into historically low-income neighborhoods, is a complex phenomenon with both positive and negative consequences. While gentrification can help to enhance infrastructure, cut crime rates, and provide better amenities, it can also cause displacement of existing residents, loss of cultural identity, and strained infrastructure. Municipalities can utilize a range of strategies to secure the benefits of gentrification while reducing its negative effects. Initiatives such as rent control and developer tax credits can help prevent the uprooting of low-income residents while encouraging development. A well-articulated program at the county and state levels can be a catalyst for the implementation of these initiatives to reduce the economic gap among various localities in the state.

Introduction

One of the objectives of local governments and city planners is to determine the variables that contribute to or limit the vibrancy of a town or city. Financial characteristics are one of the factors that influence the appeal of a community. This is because communities with strong economies have traits that attract outsiders while extending the stay of current residents. So, it is critical to determine the key component that not only contributed directly to economic inequality but also indirectly influenced other factors. This study focuses on identifying and analyzing this factor. In addition, I examine other criteria that distinguish towns' desirability. I hypothesize that the ethnic composition of the residents is the most relevant factor. The ethnic makeup of a town directly affects its prosperity. My secondary consideration is to develop feasible strategies to close the economic gap between towns. Furthermore, I attempted to investigate the potential consequences, such as gentrification, of new development. Gentrification, or the influx of money into historically low-income neighborhoods, is a complex phenomenon with both positive and negative consequences. I investigated ways to resolve any reservations that may arise as a result of gentrification. Reducing compartmentalization where some areas are more desirable and prosperous while others are impoverished, contributes significantly to achieving economic justice for all.

Methods

For my comparison investigation, I chose the towns of Westfield and Plainfield. The selection of these two adjoining towns eliminates variables related to location. The similarities between Westfield and Plainfield are many. First, the towns are adjacent and belong to Union County in New Jersey. Second, both are served by similar modes of public transit. For example, the NJ Transit Raritan Valley Line passes through both communities. Furthermore, there is a comparable network of main roadways surrounding the towns. Third, the area in square miles is approximately the same for both towns.

To help with the analysis, I gathered census data on demographics, poverty levels, income taxes, education, crime, and unemployment for these two communities. To generalize my findings, I examined New Jersey's statewide ethnicity and poverty indices. To demonstrate the interplay of other factors, I used national census data to determine the link between education and income disparities. To get a comprehensive view of these two communities, I included both residential and commercial spaces. In particular, I examined the impact of a thriving business environment in the town in terms of tax collection, influence over municipal regulations, level of out-of-town foot traffic, and so on. A vibrant downtown in any city or town adds to the overall attractiveness of the area.

Finally, I examined Plainfield's gentrification issues and the mayor's response to the issue. Gentrification is always considered a factor in housing unaffordability since it might diminish the supply of lower-cost units, resulting in the relocation of low-income, minority people.

Table 1: Westfield and Plainfield statistics (2021).[1,3,4,16]

	Plainfield	Westfield
Population	54,513	30,539
Area	5.969 mi²	6.74 mi²
Ethnicity	41% Hispanic, 39% Black, 9% White	5% Hispanic, 3% Black, 78% White
Average Income / Person	$27,532	$72,844

Property Value	$287,300	$810,400
Average Property Tax Bill	$9,130	$17,427
Home Ownership	41.6%	80.1%
Poverty Rate	17.1%	2.02%
Median Age	34.1 years	40.9 years
High School Education (% with high school diploma, 2017-20210)	77.7%	98.3%
NJ High School Ranking	352 - 399	49
Unemployment Rates	4.7%	2.9%
Violent Crime Incidents	195	4

Results and Discussion

A brief inspection of the data in Table 1 reveals the following differences between the two municipalities. It reveals that Westfield is a more expensive town, with greater property values and household incomes. Westfield has about twice the rate of home ownership as Plainfield. Plainfield's poverty and unemployment rates are substantially higher.

When comparing the highest education level reached in the two towns and its impact on white-collar work, it is clear that Westfield residents have access to higher-paying and more desirable occupations because nearly everyone there has completed high school and went on to college. These professions encompass corporate, sales, and business management, as well as financial operations. Plainfield residents, on the other hand, do not have the same opportunities because nearly one-fourth of them did not finish high school. Most Plainfield residents work in manual labor that does not require a degree or advanced specialization.

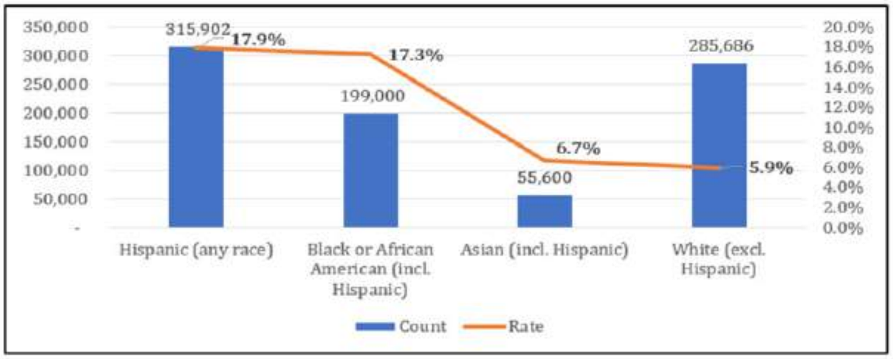
In terms of education, a comparison of the two high schools reveals that Plainfield ranks near the bottom in the state, whereas Westfield consistently ranks in the top 50. Plainfield's population and geographic dimensions indicate that it is a densely populated community with little access to open space. Finally, the substantial disparity in the number of violent crimes suggests that the two communities' safety and security characteristics are at opposing ends of the spectrum.

Impact of Wealthier Residents

The ability to attract wealthy residents has a direct impact on the town. The benefits to the town are numerous. High-income workers pay more in taxes while receiving fewer benefits. These funds, when judiciously applied, boost the town's profile. It may entail hiring extra police officers to make the town safer. Some of this money can be utilized to provide tax breaks for new firms, increasing the town's economic viability.

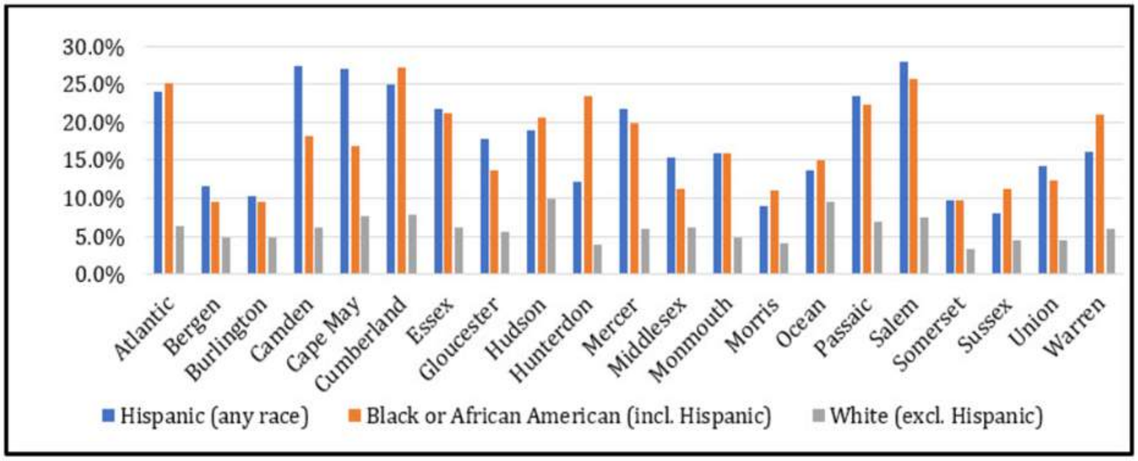
Table 1 shows that 80% of Plainfield's population is Black or Hispanic. Whites make up approximately 78% of Westfield's population. Given the stark racial differences between the two communities, it is not surprising that they have significantly distinct economic profiles. The graph in Figure 1.1 depicts the poverty rate in New Jersey by ethnicity. Whites have a rate of 5.9%, whereas Blacks and Hispanics are at nearly three times that figure. More specifically, the Union County data (Figure 1.2), which includes both Plainfield and Westfield, shows a similar gap in poverty levels.

Figure 1.1: Individuals living in poverty (count and rate) by race/ethnicity: NJ in 2019



Source: U.S. Census Bureau, 2019 (see endnote 3). Prepared by R/ECON™.

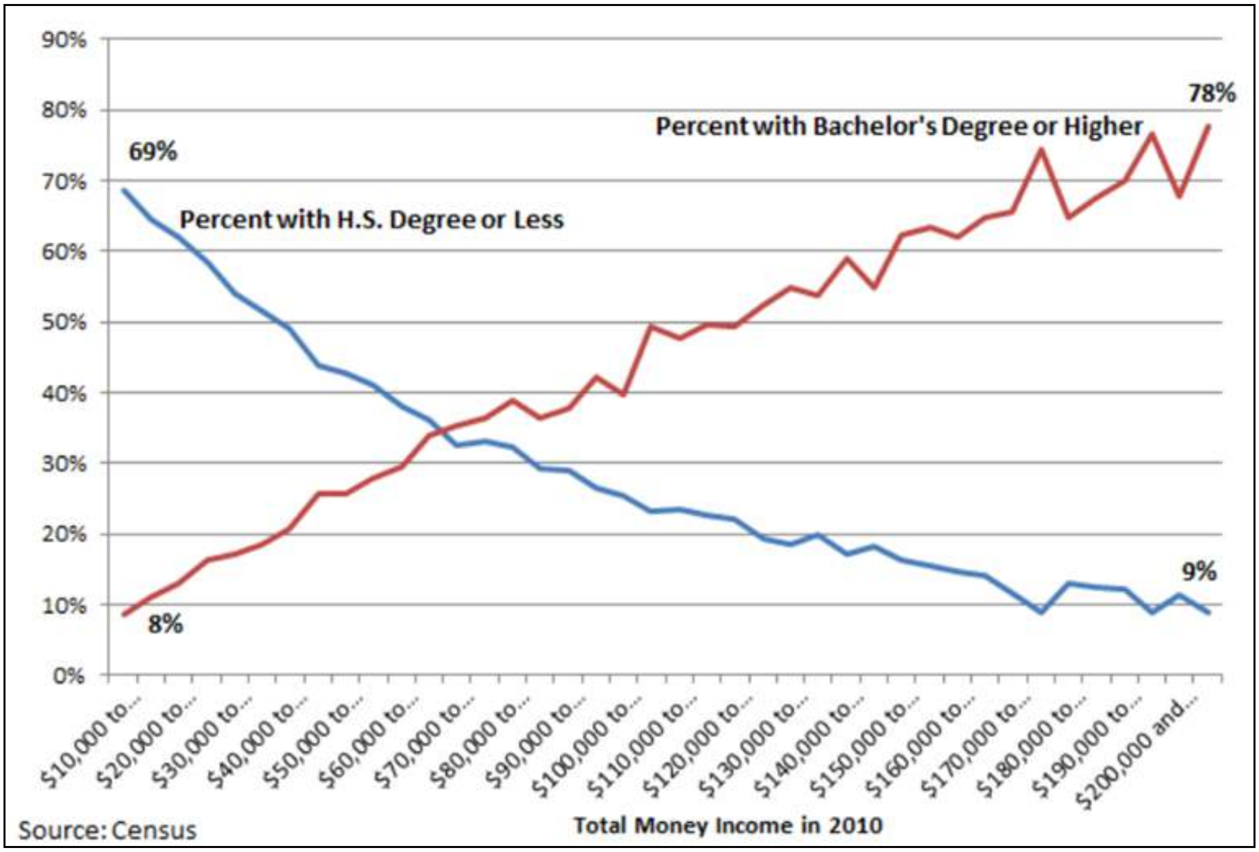
Figure 1.2: Percent living in poverty by race/ethnicity: NJ Counties, 2019



Source: U.S. Census Bureau, 2019 (see endnote 3). Prepared by R/ECON™.

While the people that make up a town are one of the most essential aspects of a community's economic success, I wanted to see the level of their impact. More specifically, I wanted to examine whether the difference in economic success between the two communities was simply due to individual wealth. My first thought was to compare the average income per person in different towns. Looking at Table 1 in the opening section of this paper, we can see that a person in Westfield earns over $40K more on average than an individual in Plainfield.

Figures 1 and 2 depict the distribution of household income by band. This view allows for a more detailed comparison of each town's total population. The spike in Plainfield is between $ 75K and $100K. Whereas, in Westfield, approximately half of the population earns $200,000 or more. This significant economic disparity may be due in part to the education divide between the two communities. Table 1 highlights the education divide, with more residents in Westfield having completed higher education than in Plainfield. Graph 1 supports the association between educational attainment and income.



Source: Census

Graph 1: 2010 Census data comparing US income versus education gap

Higher individual income correlates with a municipality producing more wealth, as shown in Table 1 by the property taxes collected by Plainfield and Westfield. For starters, wealthy people are less thrifty since they can afford more services and other goods, whereas people with less wealth are more careful about how they spend their limited disposable income.

Another advantage of wealthy citizens is that they can demand improvements in schools by becoming more involved in groups such as the PTA (Parent Teacher Association). They are often more interested in school activities and advocate for improvements to academic and athletic programs.

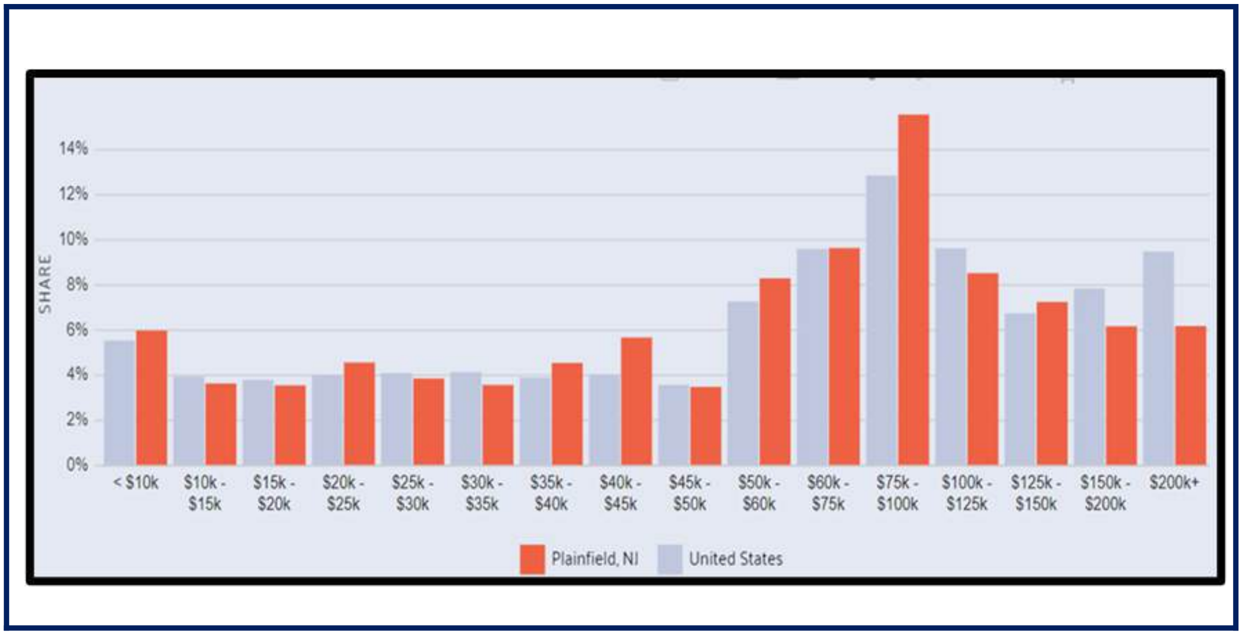


Figure 1: Average household income in Plainfield, New Jersey compared to the US in 2021.[3]

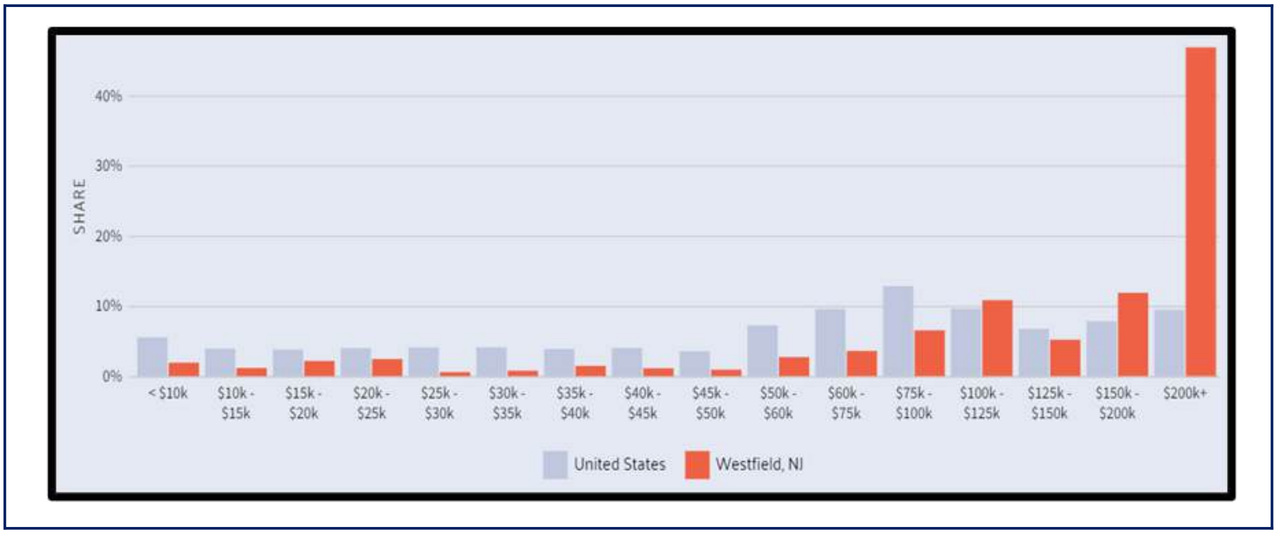


Figure 2: Average household income in Westfield, New Jersey compared to the United States in 2021.[4]

Economic Policies

Economic policy is one of the foundations upon which any municipality bases its financial decisions. Having strong and clear policies that are adequately communicated and followed can substantially benefit a municipality. On the other hand, having unfocused and poorly enforced policies could lead to a shaky financial foundation. Comparing the policies of both towns and investigating how they are implemented may provide additional insights into how the policies are affected by the towns' economies, which in turn dictate the priorities for the respective town officials.

Westfield's Economic Policies

The Town of Westfield has articulated a strategy for its future that promotes wise growth and innovation, enhances the downtown business climate, and successfully manages redevelopment.[6] An emphasis on smart growth in communities entails prioritizing long-term planning above short-term goals, with a particular emphasis on walkability and bicycling, mixed-use development, establishing a sense of place, and broadening the range of transit, employment, and housing options.

Plainfield Economic Policies

Plainfield's Office of Economic Development's declared goals are to retain, attract, and expand businesses. Its mandate also includes focusing on promoting minority businesses and the development of real estate in each of the wards.

Westfield is already economically prosperous; its priorities include downtown access, mixed-use real estate development, and so on. Over the last few years, the town has already invested in programs such as new parking spaces and walkway development. There is also a thriving program focused on clean energy. The town has launched a scheme to encourage people to switch to electric vehicles and solar panels for residences.[15]

Plainfield, on the other hand, appears to have a very generic, high-level policy geared toward attracting new firms and retaining existing ones. There is a brief mention of inexpensive housing for locals, which is a major concern in the area.[7] Many families have been evicted because they were unable to pay their mortgage or rent.[8] Plainfield's policy may have limitations since the town does not have the resources to handle concerns other than the ones that are most basic and urgent.

The Role that Businesses Play in a Local Economy

The town's economy benefits greatly from a thriving business ecosystem. For example, successful neighborhood stores result in increased local employment and better tax income for local governments.[14] According to Civic Economics' Andersonville Study of Retail Economics, for every $100 spent at locally owned companies, $68 remains in the community.[17] Businesses make major contributions to all taxes, including income, property, and employment taxes.[9] More enterprises in the area can increase tax revenue for local governments, bringing in more funds for road repairs, school development, and other public services.

Businesses with strong local links contribute to community-focused political campaigns via funding, lobbying, letter-writing, and other political measures to persuade politicians on issues that affect the neighborhood.[13] A local firm, for example, may be able to induce a political candidate to deliver a speech in the community during the campaign by contributing to fundraising. One of the most obvious personal benefits of local businesses is an increase in employment. Local employment promotes a sense of community, which can help a local business's reputation.

Gentrification and its Effects

Gentrification is the process by which more affluent inhabitants alter an existing neighborhood.[2] There are clear reasons why gentrification may be good. In theory, the approach dramatically improves an area that has poor property prices, empty properties, and high crime rates. Gentrification, according to AMI House Builders, has been credited with saving troubled cities by promoting commercial growth, improving economic opportunity, increasing property values, improving infrastructure, and reducing crime.[10] Furthermore, communities primarily rely on property taxes to fund public services such as utilities, roads, schools, police stations, and fire departments.[12] The higher the property values, the more the city collects in taxes, which can be used for the benefit of residents.

Gentrification can result in the loss of inexpensive housing, perhaps displacing long-term inhabitants, who are often minorities, by directly or indirectly creating an environment that is no longer affordable or accessible to them. While some citizens can remain in newly gentrified districts and benefit from increased funding, resources, and amenities, gentrification also has the disadvantage of displacing many people. 2017 Georgetown Law research supports the idea that gentrification typically results in negative consequences such as forced displacement, the promotion of discriminatory conduct by those in positions of power, and an emphasis on areas that exclude low-income individuals and people of color.[5]

Plainfield's mayor was asked during the 2023 State of the City address why the city continues to build apartments and encourages gentrification.[18] He responded that no old homes had been removed to make space for new residential developments. Every new development took place in abandoned and unsightly regions. He promised everyone that the introduction of new people would not displace the current ones. Finding a balance between the requirements of a community, its citizens, and companies is a difficult but worthwhile undertaking. For example, several cities and municipalities require affordable housing to be included in new developments.[11] Creating policies that accommodate both existing and incoming residents can result in an environment that benefits everyone.

Conclusion

A community's success is determined by a variety of things. The economic profile is one of the most important measures for determining success. This study paper examines various factors that contribute to the huge variations in financial standing between Westfield and Plainfield. However, the racial gap between the two communities is a major element that influences other factors to varying degrees. Westfield is largely White and Plainfield's population is primarily black or Hispanic.

Town authorities and politicians must make a long-term commitment to see the policy objectives executed and implemented. Plainfield will undoubtedly benefit from an injection of cash in the form of new, wealthier inhabitants and businesses. Plainfield can become more diverse by encouraging new residents to move there, which has a knock-on effect of drawing more people and money. There is a need to be vigilant that gentrification does not result in a loss of affordability and culture for current residents. Similarly, Westfield can benefit from providing cheaper homes and encouraging ethnic and economically disadvantaged families to relocate there. This might not only bring prospective talent to the area, but it could also help downtown businesses, some of whom are struggling due to expensive rents.

Although racial disparities dictate the towns' economic characteristics, adopting proactive initiatives to make communities more diverse can assist in closing the economic gap between Plainfield and Westfield. Many adjacent towns and communities across New Jersey and the country can benefit from the information presented in this report.

Work Cited

1. *New Jersey*. FBI. https://ucr.fbi.gov/crime-in-the-u.s/2018/crime-in-the-u.s.-2018/tables/table-8/table-8-state-cuts/new-jersey.xls (accessed 2023-12-19).
2. *Oxford Languages and Google - English | Oxford Languages*. https://languages.oup.com/google-dictionary-en/ (accessed 2023-12-19).
3. *Plainfield, NJ | Data USA*. https://datausa.io/profile/geo/plainfield-nj#demographics (accessed 2023-12-19).
4. *Westfield, NJ | Data USA*. https://datausa.io/profile/geo/westfield-nj#demographics (accessed 2023-12-19).
5. Chong, E. *Examining the Negative Impacts of Gentrification*. https://www.law.georgetown.edu/poverty-journal/blog/examining-the-negative-impacts-of-gentrification/ (accessed 2023-12-19).
6. *Westfield's Smart Growth Plan* https://www.westfieldnj.gov/839/Westfields-Smart-Growth-Plan (accessed 2023-12-19).
7. *Plainfield's new council president emphasizes the city's need for more affordable housing.* https://www.mycentraljersey.com/story/news/local/union-county/2024/01/10/plainfield-affordable- housing/72160741007/ (accessed 2023-12-19).
8. David Rutherford *Plainfield Evictions* https://www.tapinto.net/towns/plainfield/sections/government/articles/plainfield-evictions-we-know-more-than-ever-yet-questions-remain/ (accessed 2023-12-19).
9. Scott Hodge *U.S. Businesses Pay or Remit 93 Percent of All Taxes Collected in America* https://taxfoundation.org/data/all/federal/businesses-pay-remit-93-percent-of-taxes-in-america/ (accessed 2023-12-19).
10. Coldwell Banker Commercial *Gentrification: The Good, the Bad, and Its Impact on Real Estate Markets* *https://*www.cbcworldwide.com/blog/gentrification-the-good-the-bad-and-its-impact-on-real-estate-markets (accessed 2023-12-19).
11. *Planning & Zoning - Affordable Housing* https://westfieldnj.gov/FAQ.aspx?QID=176 (accessed 2023-12-19).
12. Janelle Fritts *Where Do People Pay the Most in Property Taxes?* https://taxfoundation.org/data/all/state/property-taxes-by-state-county-2022/ (accessed 2023-12-19).
13. Miranda Morley *Business Ideas for Lobbying* https://smallbusiness.chron.com/business-ideas-lobbying-35168.html (accessed 2023-12-19).
14. David Ingram *The Advantages of Businesses in the Local Economy* https://smallbusiness.chron.com/advantages-businesses-local-economy-3289.html (accessed 2023-12-19).

15. *Community Solar* https://www.westfieldnj.gov/933/Community-Solar (accessed 2023-12-19).
16. *Plainfield, Westfield NJ | Data* https://joeshimkus.com/NJ-Tax-Rates.aspx (accessed 2023-12-19)
17. Glen J. Dalakian Sr. *Shopping Local Helps Neighbors and Economy* https://www.thejournalnj.com//columns/shopping-local-helps-neighbors-and-economy/ (accessed 2023-12-19).
18. *Mayor's State Of The City Address 2023* https://www.plainfieldnj.gov/government/communications/mayor_s_state_of_the_city_address_2023.php (accessed 2023-12-19).

Lycium Barbarum as a Treatment for Type 2 Diabetes Mellitus By Shubhay Mishra

Abstract

Lycium barbarum, commonly known as goji berry, has been used in traditional Chinese medicine for centuries for its therapeutic properties. This review examines its potential as a treatment for type 2 diabetes mellitus (T2DM), a condition characterized by chronic hyperglycemia. The databases ScienceDirect, ProQuest, and PubMed were searched for relevant articles published between 2014 and 2024. The inclusion criteria focused on studies investigating the physiological effects of *L. barbarum*, particularly those involving preclinical or clinical trials. Key active compounds, including polysaccharides, phenols, and carotenoids, were identified and analyzed for their hypoglycemic properties. Major preclinical and clinical trials were reviewed to evaluate the efficacy of *L. barbarum* in managing T2DM. Although promising results were observed in both animal models and human trials, the studies were limited by small sample sizes and short durations. This review highlights the need for more extensive research to confirm the potential of *L. barbarum* as a cost-effective and natural treatment option for T2DM.

Methods

The databases ScienceDirect, ProQuest, and PubMed were used to obtain information. Articles were searched for by entering keywords, including but not limited to type 2 diabetes, *L. barbarum*, combination therapy, food therapeutics, and glycemia. The publication years of articles ranged from 2014 to 2024 to ensure the information was up-to-date. In terms of inclusion criteria, studies that reviewed or directly investigated changes in physiological profile of studies that performed preclinical or clinical trials were included. Studies that did not focus on *L. barbarum* or were not peer-reviewed were excluded from the review. The gathered information was synthesized by topic: Active compounds, study summaries, comparison with other treatments, and limitations.

Introduction

Lycium barbarum, commonly known as goji berry or wolfberry, is a widely cultivated fruit and has been a staple in traditional Chinese medicine for thousands of years. It is rich in flavonoid, carotenoids, and polysaccharides, leading to a multitude of salutary effects including neuroprotection, cytoprotection, anticancer, and antioxidation (Islam et al., 2017; Yang et al., 2022). These properties have prompted extensive research into its potential therapeutic applications, especially in the context of chronic diseases such as Parkinson's, Alzheimer's, and nonalcoholic fatty liver disease (Song et al., 2022; Zhou et al., 2020; Gao et al., 2021).

Type 2 diabetes mellitus (T2DM) is characterized by chronic hyperglycemia, or excessively high glucose. It damages a variety of internal and external organs, including the heart, kidneys, and nerves (Ta, 2014). The disease symptomizes as excessive urination, increased hunger and thirst, and numbness. In more severe cases, diabetic ketoacidosis (DKA) can also occur (Balaji et al., 2019). The CDC estimated that 38.4 million people in the United States had diabetes in 2021, and the number is expected to increase (*National Diabetes*, 2021). Current diabetes treatments focus on managing blood sugar levels and preventing complications through lifestyle changes and medication (Nathan, 2015). However, as T2DM becomes more prevalent, the search for novel, cost-effective treatments has intensified. With consumers finding plant-based treatments more appealing than synthetic options, *L. barbarum* could be an effective option for treating T2DM (Jiang et al., 2021).

This paper aims to provide a brief overview of the current research on *L. barbarum* as a treatment for T2DM, elucidating its mechanisms of action, evaluating its efficacy in previous studies, and discussing the challenges and future prospects of this treatment. This review contributes to the understanding of how traditional medicinal plants like *L.barbarum* can be integrated into modern treatment strategies for diabetes.

1 Active Compounds in *Lycium Barbarum*

1.1 Polysaccharides

The main active compounds in *L. barbarum* are water-soluble glycoconjugate polysaccharides, more than 33 of which have been identified (Tian et al., 2019). Polysaccharides compose about 5-8% of dry *L. barbarum* by weight (Ma et al., 2022).The monosaccharides within these polysaccharides are arabinose, rhamnose, galactose, glucose, mannose and xylose (Liu et al., 2022). They are industrially extracted via a number of methods, most notably water extraction method, enzyme-assisted extraction method, microwave-assisted extraction method, and ultrasonic-assisted extraction method (Wang et al., 2024). Polysaccharides account for antioxidant and antitumor activity of *L. barbarum* (Tian et al., 2019).

1.2 Phenols

In addition to the berries themselves, phenols can be found in the roots and leaves of *L. barbarum* and can be extracted through aqueous ultrasonic extraction. There are an extensive variety of phenols in *L. barbarum*, 84 of which have been discovered (Jiang et al., 2021). The phenols that can be found include flavonoids, coumarins, lignans, and phenolic acids (Ma et al., 2022). Phenols function as antioxidants (Magiera & Zareba, 2015).

1.3 Carotenoids

The main carotenoids present in *L. barbarum* are zeaxanthin dipalmitate (ZD) and its two isomers (Kan et al., 2020). ZD accounts for antioxidant, anticancer, and antifibrotic properties (Murillo et al., 2019). ZD has been shown to be beneficial for a number of diseases, especially liver-related disorders ((Bahaji Azami & Sun, 2019). Carotenoids appear in red goji berries in a concentration of about 233 µg/g (de Freitas Rodrigues).

2 Major Preclinical and Clinical Trials

2.1 Zhou et al., 2022

In this randomized study performed on 16 male diabetic mice, 8 were assigned to a control group and 8 were assigned to a *L. barbarum* polysaccharide (LBP) group. The testing period was 15 weeks. The researchers examined the changes in the gut microbiota profile of the mice, as well as diabetes biomarkers, namely fasting plasma glucose level and Hba1C index. Compared to the control group, the LBP group showed a significantly lower fasting plasma glucose level (p<0.05). Hba1C index, a value that reflects the average blood glucose level over time, was also significantly lower for the LBP group compared with the control group (p<0.05). The main change in gut microbiota was a change in concentration of GLP-1, a gut hormone that regulates insulin secretion among other roles, which seriatim helps regulate glucose. Compared to the control group, which had an average GLP-1 concentration of 5.02 pM, the LBP group had a significantly greater (p<0.01) concentration at 6.43 pM. Taken together, the results indicate that *L. barbarum* can serve as ingredients in functional food that target gut microbiota for treatment of hyperglycemia in T2DM (Zhou et al., 2022).

2.2 Zhao et al., 2016

In this randomized study performed on T2DM mice, 8 were assigned to a control group, 8 were assigned to a high concentration (10 mg/kg·day) *L. barbarum* polysaccharide group (LBP-H), and 8 were assigned to a low concentration (5 mg/kg·day) *L. barbarum* polysaccharide group (LBP-L). The control group rats received a saline solution, while the experimental group rats were given LBP-4a, a polysaccharide from L. barbarum, via gavage. This supplementation occurred daily for a period of four weeks. The researchers aimed to determine whether *L. barbarum*'s antidiabetic effects acted in a dose-dependent manner. They recorded the impact of *L. barbarum* supplementation on biomarkers of hyperglycemia and hyperinsulinemia. Compared to the LBP-L group, which had a median blood glucose level after 4 weeks of 18.57 mM, the LBP-H group had a significantly lower (p<0.05) level at 15.21 mM. Furthermore, compared to the LBP-L group, which had a median insulin level after 4 weeks of 6.08 mU/l, the LBP-H group had a significantly lower (p<0.05) level at 5.01 mU/l. The researchers also examined changes in metabolism biomarkers (fat tissue weight), as reduced metabolism is characteristic of T2DM. Compared to the T2DM control, with a final median fat tissue weight of 1.41 g, the LBP-H group had a significantly lower (p<0.05) value of 1.07 g. Notably, there was no significant difference between the LBP-L group's median fat tissue weight and that of the T2DM control. The study reveals how LBP supplementation drastically alleviates T2DM in a dose-dependent manner and further embellishes *L. barbarum* as a plausible dietary treatment. This study is limited due to the small sample size, small time period, and presence of only two levels of dosage (Zhao et al., 2016).

2.3 Pollini et al., 2020

In this study, phenolic acids derived from *L. barbarum* leaves were analyzed for porcine α-amylase inhibitory activity. α-amylase is a digestive enzyme which acts on starch to produce glucose, and T2DM is characterized by hyperglycemia. Therefore, the aforementioned activity is a proxy for antidiabetic activity. 8 main phenolic acids were determined in the leaf extract: Syringic acid, chlorogenic acid, salicylic acid, caffeic acid, vanillic acid, p-coumaric acid, sinapic acid, and vanillin. Chlorogenic acid and salicylic acid were the main constituents. IC50 values were determined for the inhibitory activity of each phenolic acid, with chlorogenic acid and salicylic acid being particularly salient at 0.5 mg/ml and 1.8 mg/ml, respectively. The overall leaf extract had an IC50 value of 25.4 mg/ml. These results suggest that phenolic compounds found in *L. barbarum* have a hypoglycemic effect, helping to alleviate symptoms of diabetes. This effect is bolstered when combined with effects from *L. barbarum* polysaccharides, also having a hypoglycemic effect. Porcine α-amylase is highly similar to human α-amylase, so these results can reasonably be extrapolated to humans (Pollini et al., 2020).

2.4 Cai et al., 2015

In this randomized, double-blind study, 67 patients with T2DM were assigned to either *L. barbarum* polysaccharide (LBP) (n=37) supplementation or placebo (n=30). Patients in the experimental group were administered LBP at a rate of 300 mg/day for 3 months. Patients in the placebo group were administered cellulose at a rate of 300 mg/day. Biomarkers for T2DM were recorded at the beginning and end of the three-month period in both groups. Compared to the placebo group, which had a median change in glucose AUC of +1.61 mmol·h·L-1, the LBP group had a significantly greater change ($p<0.05$) at -7.86 mmol·h·L-1. Furthermore, compared to the placebo group, which had a median change in insulinogenic index of -0.98, the LBP group had a significantly different ($p<0.01$) median change of +0.04. Oddly, insulin AUC change was not significantly different between groups ($p<0.47$) despite the insulinogenic index increase. The researchers also recorded AUC changes of lipids, such as cholesterol and triglycerides, but no significance was determined, which contrasts with preclinical data. These results indicate an improvement in β-cell responsiveness, and confirms the efficacy of LBP as a treatment for T2DM. This study's results are hampered by small sample size and short intervention time, which may have led to false negatives for the lipid results (Cai et al., 2015).

3 Comparative Analysis

Although *L. barbarum* has shown promising effects alone, particularly through its polysaccharides and phenolic compounds, it is possible that it is more effective when in tandem with existing medication. In a second experiment by Cai et al., 37 T2DM patients were assigned to either a *L. barbarum* without hypoglycemic medications (LBP) (n=17), or *L. barbarum* with hypoglycemic medications (LBP-Hy) (n=20). Biomarkers for diabetes were recorded before and after the 3 month intervention period, including glucose AUC and insulin AUC. In the LBP group, change in glucose AUC, change in insulin AUC, and change in homeostasis model assessment index of insulin resistance (HOMA-IR) was significantly different ($p<0.05$). In the LBP-Hy group, no biomarkers were significantly different ($p>0.05$). Notably, glucose AUC decrease and HOMA-IR decrease was observed, but to a lesser extent than the former group.

This suggests that LBP has a negative effect on hypoglycemic medication (Cai et al., 2015). However, further clinical trials need to address the hypoglycemic efficacy of LBP versus existing medications, in addition to concurrent usage.

4 Challenges and Limitations

Presumably due to being a novel research area, there were very few research papers that passed inclusion criteria for this review. The majority of studies were preclinical trials conducted on animal models. Although these studies provide valuable insights, the results may not be wholly translatable to humans due to physiological differences or other confounding variables. For example, the metabolic processes in mice or rats may respond differently to *L. barbarum* compared to humans, potentially skewing efficacy and safety profiles. Despite this, the few clinical data available shows that *L. barbarum* is safe for consumption in individuals with T2DM. All studies reviewed were limited in scope. They have small sample sizes, short durations, and lack diversity in subject populations. The single clinical trial conducted by Cai et al. did not include long-term follow-up, which makes it challenging to determine the prolonged efficacy and safety of L. barbarum. If *L. barbarum* is ever going to be used as a treatment for T2DM, studies that address these limitations must be conducted.

Conclusion

Thus far, *Lycium barbarum* has demonstrated potential as a safe and effective treatment for type 2 diabetes mellitus, primarily through its polysaccharides, phenolic compounds, and carotenoids. Preclinical and clinical studies indicate its ability to lower blood glucose levels, improve insulin sensitivity, and regulate metabolism. However, the current body of research is limited by small sample sizes, short study durations, and a lack of diverse subject populations. Future studies should aim to address these limitations by conducting larger, longer-term trials with varied demographics. If these challenges are overcome, *L. barbarum* could become a valuable addition to the range of treatments available for managing T2DM, aligning with the growing consumer preference for plant-based therapies.

Works Cited

Bahaji Azami, N. L., & Sun, M. (2019). Zeaxanthin dipalmitate in the treatment of liver disease. *Evidence-Based Complementary and Alternative Medicine*, *2019*, 1-14. https://doi.org/10.1155/2019/1475163

Balaji, R., Duraisamy, R., & Kumar, M. P. (2019). Complications of diabetes mellitus: A review. *Drug Invention Today*, *12*(1).

Cai, H., Liu, F., Zuo, P., Huang, G., Song, Z., Wang, T., Lu, H., Guo, F., Han, C., & Sun, G. (2015). Practical application of antidiabetic efficacy of Lycium barbarum polysaccharide in patients with type 2 diabetes. *Medicinal Chemistry*, *11*(4), 383-390. https://doi.org/10.2174/1573406410666141110153858

de Freitas Rodrigues, C., Ramos Boldori, J., Valandro Soares, M., Somacal, S., Emanuelli, T., Izaguirry, A., Weber Santos Cibin, F., Rossini Augusti, P., & Casagrande Denardin, C. (2021). Goji berry (Lycium barbarum l.) juice reduces lifespan and premature aging of caenorhabditis elegans: Is it safe to consume it? *Food Research International*, *144*, 110297. https://doi.org/10.1016/j.foodres.2021.110297

Gao, L.-L., Ma, J.-M., Fan, Y.-N., Zhang, Y.-N., Ge, R., Tao, X.-J., Zhang, M.-W., Gao, Q.-H., & Yang, J.-J. (2021). Lycium barbarum polysaccharide combined with aerobic exercise ameliorated nonalcoholic fatty liver disease through restoring gut microbiota, intestinal barrier and inhibiting hepatic inflammation. *International Journal of Biological Macromolecules*, *183*, 1379-1392. https://doi.org/10.1016/j.ijbiomac.2021.05.066

Islam, T., Yu, X., Badwal, T. S., & Xu, B. (2017). Comparative studies on phenolic profiles, antioxidant capacities and carotenoid contents of red goji berry (Lycium barbarum) and black goji berry (Lycium ruthenicum). *Chemistry Central Journal*, *11*(1). https://doi.org/10.1186/s13065-017-0287-z

Jiang, Y., Fang, Z., Leonard, W., & Zhang, P. (2021). Phenolic compounds in Lycium berry: Composition, health benefits and industrial applications. *Journal of Functional Foods*, *77*. https://doi.org/10.1016/j.jff.2020.104340

Kan, X., Yan, Y., Ran, L., Lu, L., Mi, J., Zhang, Z., Li, X., Zeng, X., & Cao, Y. (2020). Evaluation of bioaccessibility of zeaxanthin dipalmitate from the fruits of lycium barbarum in oil-in-water emulsions. *Food Hydrocolloids*, *105*, 105781. https://doi.org/10.1016/j.foodhyd.2020.105781

Liu, H., Cui, B., & Zhang, Z. (2022). Mechanism of glycometabolism regulation by bioactive compounds from the fruits of lycium barbarum: A review. *Food Research International*, *159*, 111408. https://doi.org/10.1016/j.foodres.2022.111408

Liu, J., Pu, Q., Qiu, H., & Di, D. (2021). Polysaccharides isolated from lycium barbarum L. by integrated tandem hybrid membrane technology exert antioxidant activities in mitochondria. *Industrial Crops and Products*, *168*, 113547. https://doi.org/10.1016/j.indcrop.2021.113547

Ma, R.-H., Zhang, X.-X., Thakur, K., Zhang, J.-G., & Wei, Z.-J. (2022). Research progress of lycium barbarum L. as functional food: Phytochemical composition and health benefits. *Current Opinion in Food Science*, *47*, 100871. https://doi.org/10.1016/j.cofs.2022.100871

Magiera, S., Zaręba, M. Chromatographic Determination of Phenolic Acids and Flavonoids in *Lycium barbarum* L. and Evaluation of Antioxidant Activity. *Food Analysis Methods, 8*, 2665–2674 (2015). https://doi.org/10.1007/s12161-015-0166-y

Murillo, A., Hu, S., & Fernandez, M. (2019). Zeaxanthin: Metabolism, properties, and antioxidant protection of eyes, heart, liver, and skin. *Antioxidants*, *8*(9), 390. https://doi.org/10.3390/antiox8090390

Nathan, D. M. (2015). Diabetes: advances in diagnosis and treatment. *Jama*, *314*(10), 1052-1062.

National diabetes statistics report. (2021, May 15). Centers for Disease Control and Prevention. https://www.cdc.gov/diabetes/php/data-research/index.html

Pollini, L., Riccio, A., Juan, C., Tringaniello, C., Ianni, F., Blasi, F., Mañes, J., Macchiarulo, A., & Cossignani, L. (2020). Phenolic acids from lycium barbarum leaves: In vitro and in silico studies of the inhibitory activity against porcine pancreatic α-Amylase. Processes, 8(11), 1388. https://doi.org/10.3390/pr8111388

Song, J., Liu, L., Li, Z., Mao, T., Zhang, J., Zhou, L., Chen, X., Shang, Y., Sun, T., Luo, Y., Jiang, Y., Tan, D., Tong, X., & Dai, F. (2022). Lycium barbarum polysaccharide improves dopamine metabolism and symptoms in an mptp-induced model of parkinson's disease. *BMC Medicine*, *20*(1). https://doi.org/10.1186/s12916-022-02621-9

Wang, J., Li, S., Zhang, H., & Zhang, X. (2024). A review of lycium barbarum Polysaccharides: Extraction, purification, structural-property relationships, and bioactive molecular mechanisms. *Purification, Structural-Property Relationships, and Bioactive Molecular Mechanisms*. https://doi.org/10.2139/ssrn.4840573

Yang, T., Hu, Y., Yan, Y., Zhou, W., Chen, G., Zeng, X., & Cao, Y. (2022). Characterization and evaluation of antioxidant and anti-inflammatory activities of flavonoids from the fruits of lycium barbarum. *Foods*, *11*(3), 306. https://doi.org/10.3390/foods11030306

Zhao, R., Gao, X., Zhang, T., & Li, X. (2016). Effects of Lycium barbarum. polysaccharide on type 2 diabetes mellitus rats by regulating biological rhythms. Iranian journal of basic medical sciences, 19(9), 1024–1030.

Zhou, Y., Duan, Y., Huang, S., Zhou, X., Zhou, L., Hu, T., Yang, Y., Lu, J., Ding, K., Guo, D., Cao, X., & Pei, G. (2020). Polysaccharides from lycium barbarum ameliorate amyloid pathology and cognitive functions in app/ps1 transgenic mice. *International Journal of Biological Macromolecules*, *144*, 1004-1012. https://doi.org/10.1016/j.ijbiomac.2019.09.177

Zhou, W., Yang, T., Xu, W., Huang, Y., Ran, L., Yan, Y., Mi, J., Lu, L., Sun, Y., Zeng, X., & Cao, Y. (2022). The polysaccharides from the fruits of lycium barbarum L. confer anti-diabetic effect by regulating gut microbiota and intestinal barrier. Carbohydrate Polymers, 291, 119626. https://doi.org/10.1016/j.carbpol.2022.119626

Feasibility of Range Improvement Using Vertical Axis Wind Turbines on Agricultural Drones By Sanjay Ravishankar

Abstract

Agricultural drones benefit farmers immensely with numerous applications; however, these drones can be difficult to obtain due to the initial and operational costs. One potential method to decrease the operational cost is to use vertical axis wind turbines as these turbines have improved the range of electric vehicles, such as cars and trucks. Because there is a gap in the current literature regarding the use of wind turbines on drones, this experiment explored the feasibility of applying vertical axis wind turbines to drones as a method to increase the range and improve the efficiency of the drone. The experiment used a positive control group with no additional materials, a negative control group with six wind turbines, and an experimental group with six wind turbines attached to a portable battery. The independent variable was the use of wind turbines while the dependent variables were time in air and voltage generated, if applicable. The statistical tests conducted on the data show that the use of wind turbines does not have a statistically significant impact on the duration of the flight since the ANOVA *p*-value was 0.36 and the *t*-test *p*-value was 0.089. This means that the drag caused by the wind turbines was not significant, and the turbines do not negatively impact the flight of the drone. The typical voltage generated is greater than 6 volts, which is enough to charge the battery given that the current produced is at least one ampere. Therefore, there is a potential positive impact since the wind turbines are able to charge the battery of the drone. Overall, the experiment concluded that using wind turbines on a drone is feasible, and further research could experimentally quantify the impact on its range.

Introduction

Recently, there has been much interest surrounding wind energy and wind turbines. Wind turbines are devices that have the ability to convert the motion of the wind to electrical energy that can be used for power. First, the incoming wind exerts a torque on the blades of the turbines, which causes the turbine to begin spinning. This leads to an increase in the rotational kinetic energy of the wind turbine; this kinetic energy is then converted into electrical energy using a generator (Douak et al., 2018). One major aspect of aerodynamics that is extremely useful when examining wind turbines is computational fluid dynamics. Computational fluid dynamics plays a crucial role in aerodynamics studies by enabling researchers to simulate and analyze complex airflow patterns. Additionally, computational fluid dynamics allows for the detailed modeling of fluid flow, turbulence, and pressure distribution, providing insights into the aerodynamic performance without the need for costly and time-consuming wind tunnel experiments. This computational approach aids in the design and optimization of more efficient and aerodynamically sound vehicles, ultimately leading to advancements in aviation, automotive engineering, and various other industries (Li et al., 2023).

Specifically, vertical axis wind turbines have been utilized to harness wind energy. Unlike the more common horizontal-axis wind turbines, these turbines have an axis of rotation that is perpendicular to the wind. This means that such turbines can utilize wind that comes from any direction. Some examples of vertical axis wind turbines include Savonius wind turbines, Darrieus wind turbines, and hybrid Savonius-Darrieus wind turbines. There are numerous benefits of vertical axis wind turbines; in an experiment that tested seven different types of wind turbines, the optimal turbine was found to be the Savonius-Darrieus, with the Savonius as a close second (Wilberforce et al., 2023). Because vertical axis wind turbines are oriented perpendicular to the wind, they are able to capture wind from any direction. This quality makes them preferable in various different environments. For instance, vertical axis wind turbines, especially the Savonius-Darrieus turbine, reign supreme in both low wind quality and low wind speed conditions (Wilberforce et al., 2023). Similarly, their ability to capture wind from any direction makes them useful when the wind is strong and turbulent as turbulent wind is prone to constantly change direction (Redchyts et al., 2023). In addition to this, vertical axis wind turbines can be beneficial due to their noise and size. It has been elucidated that when compared to their horizontal axis counterparts, vertical axis wind turbines produce significantly less noise (Li et al., 2022). They are also more compact, which allows them to be installed close together. Because of this, these turbines are superior in urban regions with low space availability (Eftekhari et al., 2021). Vertical axis wind turbines provide multiple benefits, including their capability to reap wind energy in different environments, their relatively low levels of noise, and their compact sizes.

Many previous studies have attempted to optimize systems of vertical axis wind turbines in order to increase the power of the turbines. To begin with, it has been found that increasing the number of wind turbines generally leads to an increase in the efficiency of each turbine (Hassanpour and Leila, 2021). Similarly, the use of an upstream deflector, which deflects the airflow towards the blades of the turbines, can improve the performance of the wind turbine, particularly in unsteady wind conditions (Zhao et al., 2021). However, there are also some features of the vertical axis wind turbines themselves that have been optimized. The optimal rotor radius ratio of a hybrid turbine is 0.5, which means that the radius of the inner rotor is half of the radius of the outer rotor (Irawan et al., 2023). The ratio between the wind speed and the speed of the tips of the blades, known as the tip speed ratio, should be 2.6 in the ideal vertical axis wind turbine. Additionally, the maximum performance of the turbine is achieved when the pitch angle is four degrees. The turbine solidity – which is the ratio between the area of the blades and the sweeping area – should be approximately 0.315 (Elsakka et al., 2022). When investigating the use of vertical axis wind turbines, all of these parameters must be considered for an optimal performance.

With this knowledge of vertical axis wind turbines as well as wind turbines in general, the integration of wind turbines on electric vehicles has been the focus of extensive research in recent years. Such an integration is visualized in Figure 1. The results of these generally have highly encouraging outcomes; both the range and efficiency of the vehicles have improved through these additions in various studies. First of all, the inclusion of micro wind turbines on electric cars have resulted in an increase in the range by up to 6.8%. This is a significant development as it can help reduce the charging time of such cars (Ebaid et al., 2023). A similar study found that electric cars could experience an increase in range of nearly 23% due to regeneration from the wind turbines (Gupta and Naveen, 2021). Similar conclusions were drawn when vertical axis wind turbines were used on cars and trucks on the highway. The power coefficient of the turbines was found to be a maximum of 24%, which shows that it can easily be utilized to extend the range of these vehicles since the power coefficient is a measure of the efficiency of a wind turbine (Hu et al., 2022). While there has been research on the use of wind turbines on electric cars and trucks, there has not yet been similar research on drones, which creates a gap in the current literature.

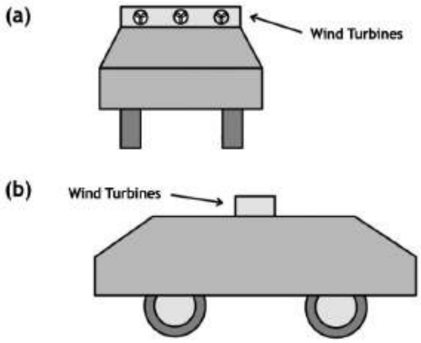


Fig 1: Application of wind turbines on an electric vehicle from (a) the front view, and (b) the side view.

Understanding the physics and aerodynamics of drones, or unmanned aerial vehicles, is vital for research that aims to optimize performance. Specifically, the analysis of the lift and airflow of a drone must be considered. As aerial vehicles, drones are affected by the four fundamental forces of aerodynamics. As shown in Figure 2, these forces are the upwards lift force, the downwards weight force, the forwards thrust force, and the backwards drag force. While the lift force causes the drone to move upwards, the weight force creates a balance that allows the drone to maintain a constant altitude. In the case of a drone, the process of generating lift involves the rotation of the propellers, which creates an upward force. Owing to the fact that the propellers are a vital factor in the performance of a drone, there are various categories of drones based on the number and type of propellers, each of which has advantages and disadvantages. The generation of lift is especially important in the takeoff stage of a drone's flight as the lift force must be greater than the opposing weight force. Since Newton's Second Law states that an acceleration is caused by a net force in the same direction, the net force must be upwards for the drone to take off (Monteiro et al., 2022).

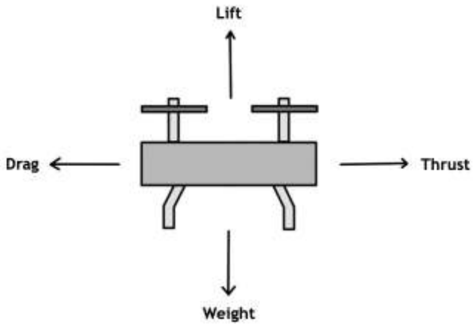


Fig 2: Four fundamental forces of aerodynamics on a drone.

Another important topic to consider is the airflow that surrounds a drone. This is due to the fact that the airflow essentially dictates the performance of the drone. Efficient airflow management is crucial for generating lift, minimizing drag, and maintaining stable flight. The shape and design of the drone are optimized to control the pressure distribution over their surfaces, ensuring that the airflow creates the necessary lift to counteract gravity and achieve stable flight (Ghirardelli et al., 2023). Additionally, flow control techniques are often used to manipulate the airflow over the drone's surfaces to improve lift and reduce drag (Greenblatt and Williams, 2022). As drone applications continue to expand in countless industries, utilizing the aerodynamics and airflow of a drone can greatly improve the performance and efficiency of the drone.

Recently, one such application of drones has been in the agricultural industry. Drones have proven to be valuable tools for farmers in recent years. Equipped with advanced sensors and cameras, drones can capture high-resolution images of crops, soil, and vegetation, allowing farmers to assess crop health and detect pests and diseases with unrivaled accuracy (Chin et al., 2023). Likewise, by analyzing images captured by drones, farmers can make informed decisions about irrigation, fertilization, and pesticide application; this targeted approach to crop management increases yields and reduces environmental impact by minimizing the use of chemicals (Hafeez et al., 2023). Additionally, drones contribute to precise crop mapping and analysis, leading to improved yield prediction and optimized cultivation practices (Rejeb et al., 2022). Not only are drones useful in the collection of images and data, but they also utilize technology to enable targeted and controlled spraying of fertilizers and pesticides. This reduces the need for excessive chemical usage and promotes sustainable, eco-friendly farming practices (Hafeez et al., 2023). Drones offer significant time and cost savings compared to traditional methods of field monitoring due to their ability to cover large areas quickly and efficiently. Agricultural drones, however, can be inaccessible for some farmers. Not only do certain regulations prevent rural workers from purchasing drones, but these vehicles are also extremely expensive (Rejeb et al., 2022). The farmers who need help maintaining their farms are those who cannot afford an expensive drone. This inequity in the agricultural industry is detrimental to the economies of developing nations as well (Ayamga et al., 2021). Therefore, based on the innumerable benefits that agricultural drones bring to farms, it is imperative that farmers obtain access to these drones. One potential solution is to improve the quality of cheaper drones to have similar qualities as more expensive drones. The advancement of the agricultural industry likely depends upon the ability to improve existing technology in order to extend the reach of such technology to those who currently cannot afford it.

In order to address the gap concerning the application of wind turbines to drones as well as mitigate the current problem with technology in the agricultural industry, this research investigated the feasibility of using Savonius vertical axis wind turbines on agricultural drones to improve the range of the drone. Because the airflow of a drone causes an increase in drag if wind turbines were to be added to the drone, this experiment is considered successful if the decrease in time due to the turbines was not statistically significant. In other words, the difference in flight duration was not impacted to a large degree by the addition of the vertical axis wind turbines. Additionally, the voltage generated by the wind turbines must be greater than six volts as calculations show that a minimum of six volts is required to charge the drone, given that the current is at least one ampere. Therefore, this research focused on the improvement of ordinary drones by increasing the range using relatively cheap materials in an effort to make agricultural drones more affordable and accessible.

Materials and Methods

The drone that was used in the experiment was the Bwine F7 Drone, model F7GB2 3B (sourced from Amazon, Shenzhen Coolle Chaowan Technology Co., Ltd, Shenzhen, China). Additionally, six QX Electronics vertical axis wind turbines (sourced from Amazon, Shenzhen, China) were used as the wind turbines. A generic stopwatch and digital multimeter were used to collect measurements, and a spare battery from the drone was used to store the energy from the wind turbines. This experiment was conducted indoors in order to protect privacy. To ensure safety, eye and body protection was utilized. This included safety goggles to safeguard the eyes and multiple layers of clothing for safety in the case of a collision.

The independent variable of this experiment was use of the Savonius vertical axis wind turbines on the drone. Based on this independent variable, there were three groups used in this experiment: a positive control group, which consisted of a drone without any additional materials; a negative control group, which consisted of a drone with six wind turbines in order to test the drag; and an experimental group, which consisted of a drone with six wind turbines attached to the battery in order to collect the energy from the wind turbines. Figure 3 displays the three different groups whereas Table 1 provides the similarities and differences between the groups. The two dependent variables of this experiment were the time that the drone spent in the air, in seconds, and the voltage generated by the wind turbines, in volts. First, the time that the drone was able to hover for the entirety of its battery life was quantified. This provides insight on the effect of the wind turbines on the time in air as well as the stability of the drone. Because the drone used in this experiment did not allow mid-flight charging, the voltage produced was directed to a spare battery from the drone. This voltage, which reveals the ability of the wind turbines to charge the drone, was measured using a multimeter. The constants in the experiment included the wind speed of 3 meters per second, the altitude of 3 meters, the temperature of 20° Celsius, and the pressure of 1 atmosphere.

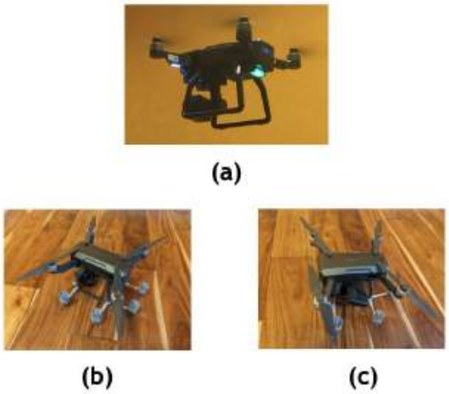


Fig 3: Side views of drones in (a) Group #1, (b) Group #2, and (c) Group #3.

Feature	**Group #1:** Positive Control Group	**Group #2:** Negative Control Group	**Group #3:** Experimental Group
Drone	Yes	Yes	Yes
Portable Battery	No	No	Yes
Wind Turbines	0	6	6
Additional Mass (kg)	0	0	1

Table 1: Overall schematic of groups.

The overall purpose of this research was to test the feasibility and effectiveness of using vertical axis wind turbines to harness wind energy on a drone to improve the range of the vehicle. This was tested through ten trials of each group, resulting in a total of thirty trials, in which the drag caused by the wind turbines, the additional weight of the battery, and the voltage generated by the wind turbines was measured. Since the drag and weight directly correlate to the decrease in the range of the drone after the addition of wind turbines, the time that the drone is in the air was measured in the experiment. If there is a significant decrease in time after the turbines are added, this would mean that it is not feasible to use these wind turbines on a drone. Thus, a successful result includes a reduction of time that is not statistically significant and a voltage that is greater than six volts.

Procedure

Figure 4, which presents the overall procedure, begins with obtaining and preparing the necessary materials. First of all, the drone was prepared, and the vertical axis wind turbines were attached to DC motors. These wind turbines were placed on wooden spokes that were attached to the drone in the trials for the negative control group and experimental group. The wind turbines were strategically positioned to be directly in the path of the airflow generated by the drone propellers. This ensures that the wind turbines are activated both when the drone is hovering or moving through the air.   Since the wind that the drone passes through and the airflow of the propellers would have components that are in the same direction, the wind generated by the propellers will always complement the initial wind. Despite the fact that propellers generally generate wind in a predominantly vertical direction, there is a component of the airflow that is horizontal, and since this wind velocity causes the wind turbines to spin, it can be approximated to be at least 3 meters per second as that is the conservative value at which wind turbines begin to spin. The reason why vertical axis wind turbines were used instead of horizontal axis wind turbines is due to the fact that while most of the airflow surrounding a propeller is downwards, the wind from the two propellers that are on either side of the wind turbine would cancel each other since both would point downwards on different sides of the turbine. Therefore, vertical axis wind turbines, which take in wind from all directions, were preferable for this situation. Additionally, unlike horizontal axis wind turbines, these turbines produce a limited amount of wind in the vertical direction, which means that the wind generated by the turbines would not disrupt the rotation of the propellers.

The first phase of the experiment was to test the positive control group. This included ten trials of a drone with no additional materials in order to measure the time in air. The next phase was to test the negative control group in a similar fashion. This time, the six vertical axis wind turbines were attached to the drone through the spokes to identify the effect on the drag of the drone by measuring the time in air over ten trials and comparing these values with the values from the positive control group. After that, ten trials of the experimental group were completed. Because this included a portable battery that was attached to the wind turbines, it involved measuring the voltage of the battery in addition to the time in air. This was used to determine if the wind turbines provided enough energy to charge the drone. The final phase of the experiment was data analysis. After collecting all of the data from the trials, PRISM GraphPad was utilized to compute the statistical summaries and apply the necessary tests.

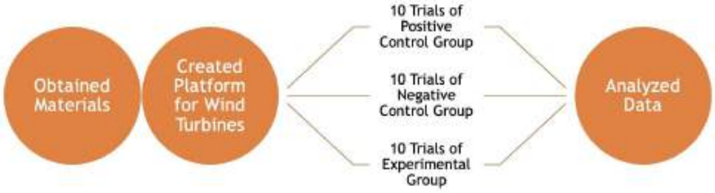

Obtained Materials
Created Platform for Wind Turbines
10 Trials of Positive Control Group
10 Trials of Negative Control Group
10 Trials of Experimental Group
Analyzed Data

Fig 4: Diagram showing the overall procedure for this experiment.

Challenges

The actual experiment included some challenges that required trial-and-error. First of all, horizontal axis wind turbines were used on the drone, but due to the problems mentioned in the procedure, these turbines did not spin. After using vertical axis wind turbines for the first two groups, the circuitry required for the experimental group caused numerous other problems at first. Originally, the wind turbines that were used were equipped with AC motors, which was not compatible with the multimeter and battery. Therefore, it was necessary to adjust the motors, which involved attaching the wind turbines to a DC motor rather than the original AC motor. This worked well except for a problem when combining multiple wind turbines into one source. It became a series circuit, which means that if the first wind turbine was rotating, instead of directing the energy generated to the multimeter along with the others, the energy was directed to the second wind turbine. This means that when the first turbine began to rotate, the second one began to rotate without any wind. In order to fix this problem, a diode was required to prevent the charge from moving in a certain direction. In addition, the drag caused by the platform in one of the designs was extremely high, which caused the drone to be unable to fly high. The design of the drone and the wind turbines changed multiple times due to the numerous problems that showed up over time.

Results

After obtaining the time and voltage measurements, the data were recorded, and several statistical tests were conducted. This included statistical summaries, standard deviation and standard error of the mean calculations and a one-way ANOVA test, and a *t*-test. The standard deviation of a dataset is a measure of variability of the data while the standard error of the mean is a measure of precision of the mean. The time in air and voltage were analyzed separately using these tests.

Statistic	**Mean**	**Median**	**SD**	**IQR**	**SEM**
Group #1 Positive Control	19.36	19.11	3.02	4.30	0.96
Group #2 Negative Control	18.72	19.19	2.71	3.27	0.86
Group #3 Experimental	17.51	17.56	2.89	5.13	0.91

Table 2: Statistical summaries for time measurements.

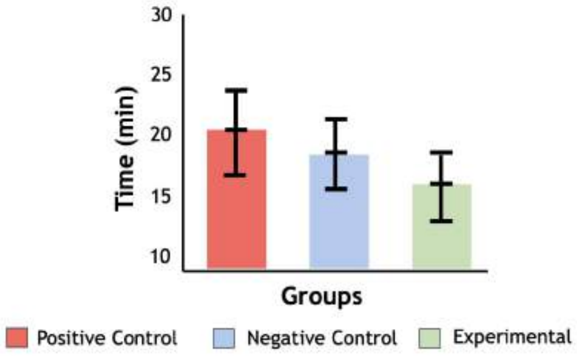


Fig 5: Bar graph of time data points.

First of all, the statistical summaries were calculated for the time data, resulting in the values in Table 2. Additionally, the raw data was graphed as a bar graph, as shown in Figure 5, which displayed the mean and 95% confidence interval for each group. A one-way ANOVA test was conducted on all three groups, which included a summary with important information, a Brown-Forsythe test, and a Bartlett's test. The final test that was conducted on the time measurements was a *t*-test. Because the positive control group and the experimental group have the largest difference, the *t*-test was conducted on these two groups.

Value	ANOVA Summary	Brown-Forsythe test	Bartlett's test
***p*-value**	0.36	0.88	0.95
Statistically different (*p* < 0.05)?	No	No	No

Note: The design of this experiment proposed that if the design was successful, then there would not be a statistically significant difference.

Table 3: Results of ANOVA test.

Statistic	Hypothesized Mean Difference	Degrees of Freedom	*t*-statistic	*p*-value	Statistically different (*p* < 0.05)?
***t*-test**	0	18	1.400	0.089	No

Note: The design of this experiment proposed that if the design was successful, then there would not be a statistically significant difference.

Table 4: Results of *t*-test.

The voltage data was analyzed as well. Because there was only one group with voltage data, only numerical summaries and figures were used. In addition to the statistical summaries, as in Table 5, the raw data was graphed as a violin plot, as in Figure 6.

Statistic	Mean	Median	SD	IQR	SEM
Group #3 Experimental	6.02	6.02	0.14	0.25	0.04

Table 5: Statistical summaries for voltage measurements.

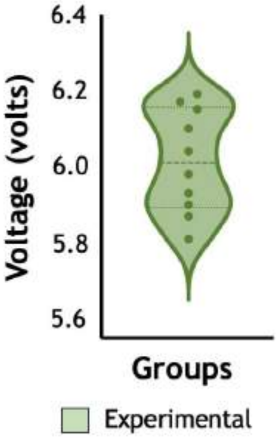


Fig 6: Violin plot of voltage data points.

Data Analysis

The statistical summaries that were calculated were used to investigate the impact of wind turbines and additional weight on the time spent in air. As demonstrated on the bar graph, the mean time was observed to decrease slightly as the wind turbines and weights were added. However, the decrease was not significant. Various statistical tools, such as the ANOVA test and the *t*-test, can be used to support the claim that the effect on time was negligible. In the ANOVA summary, the *p*-value is 0.36; similarly, the *p*-values of the Brown-Forsythe test and the Bartlett's test are 0.88 and 0.95, respectively. Because the condition for the *p*-value for all of these tests is 0.05, all of the *p*-values are greater than the 0.05 standard, which elucidates that the difference in times is not statistically significant. In addition, the *p*-value of the *t*-test was calculated to be 0.089. Similar to the ANOVA test, the standard for the *p*-value is 0.05, which means that the difference in mean times between the two most extreme groups is not statistically significant. Both of these indicate that the decrease in time was due to chance rather than the use of wind turbines.

It is also important to interpret the results of the voltage measurements. The statistical summaries demonstrate that the typical value for the voltage is greater than six volts. Although not all of the data values meet this criterion, both the mean and the median are slightly greater than six volts. Additionally, since the current was at least one ampere at all times during the experiment, the voltage was successful the majority of the time.

Thus, the two data that were collected in this experiment demonstrate that there is little negative impact on the time in air by the wind turbines and that most of the time, the voltage generated by the wind turbines is greater than the minimum amount required to charge the drone.

Error Analysis and Future Research

One possible source of error is the wind conditions during the experiment. While the experiment was conducted indoors, it is possible that the wind speed was not always exactly three meters per second due to variations in the airflow. This could have caused disruptions in the flight time of the drone in various trials, which would have been attributed to the independent variable. In addition, randomization of the trials would have likely decreased error. In the experiment, the first ten trials were performed using the positive control group, followed by ten trials of the negative control group, and then ten trials of the experimental group. However, it is possible that the drone could have worn down over time, meaning that the experimental group would have included a slightly worse drone than the positive control group. If the group used in each trial was randomly assigned, then this error could have been resolved.

Some significant limitations of this experiment were the lack of variety in drones and wind turbines. The experiment used one drone and one type of vertical axis wind turbine rather than multiple types. Even though these were not the independent variables in this experiment, it could have been useful if the same experiment was performed on varying types of drones and used varying types of vertical axis wind turbines. If similar results were achieved when different drones were used, then it can be concluded that the use of vertical axis wind turbines can be applied to a wide variety of unmanned aerial vehicles rather than only the drone used in the experiment. In addition to this, the experiment could have compared different types of wind turbines, including Savonius turbines, Darrieus turbines, and Savonius-Darrieus turbines. While this experiment only used Savonius wind turbines, an experiment involving multiple different types of vertical axis wind turbines could identify the most effective type for this application. Another limitation of the research was the fact that the drone used in the experiment did not allow mid-flight charging since the battery could not be charged while it was discharging. If this were allowed, then the effect of the wind turbines on the range of the drone could have been quantified. Instead, the impact must be approximated using logical assumptions. Overall, this research can be expanded to other types of drones and different types of vertical axis wind turbines, both of which have unique benefits.

There are a variety of potential experiments that can build upon this research. For instance, future research could vary the vertical axis wind turbines. Two possible experiments that could be conducted are those that vary the placement of the wind turbines and those that vary the type of the wind turbines. It is possible that the wind turbines would have a greater effect if they were located on a different part of the drone. In order to test this, an experiment would need to be conducted with wind turbines in varying positions. Similarly, the effect of deflectors or other tools could be considered in a future experiment. The most impactful research, though, would likely be a comparison of various different types of vertical axis wind turbines. This current experiment utilized Savonius vertical axis wind turbines. Other popular types include Darrieus wind turbines and hybrid Savonius-Darrieus wind turbines, both of which are extremely different in terms of design and efficiency. If a future experiment were to have the type of wind turbine as the independent variable, it is possible that this idea could be further developed and improved upon. Finally, the type of drone could be changed in a future experiment in order to allow mid-flight charging. This would likely result in an estimate of the true increase in range that a drone would experience due to the addition of vertical axis wind turbines. Therefore, the factors that cause the limitations of this experiment, namely the type of drone and the type of vertical axis wind turbine, could be used as points of further research in the future.

Potential Impact

Currently, the majority of the rural workers in developing countries face barriers to affordable drone technology due to either cost or regulations. The high price of agricultural drones is increasingly becoming a problem in this industry as farmers are unable to obtain the help that they need to maintain their farms (Rejeb et al., 2022; Ayamga et al., 2021). This research provides a positive impact in the agricultural industry by increasing the overall affordability of agricultural drones through the improvement in the quality of cheaper drones to match that of more expensive drones, starting with the range.

In order to approximate the impact that the wind turbines would have on an agricultural drone, some calculations must be performed. These calculations are not experimental results; they are simply mathematical approximations given the results of this experiment as well as some assumptions of a typical drone. First, the average charging time of a drone is generally between one and two hours, with most drones having a charging time of 90 minutes. The rate of charge is calculated by finding the ratio of the total amount of charge gained to the time taken for the gain in charge. The total charge gained is 100% due to charging the battery over the charging time of 90 minutes, which means that the average rate of charge is approximately 1.11% per minute. The total charge gained from the use of wind turbines can be calculated assuming an optimal rate of charge. It can be stated that the charge gained is equal to the rate of charge multiplied by the total time by rearranging the previous equation. Because the rate of charge for a typical drone is given to be 1.11% per minute as calculated previously, and the mean time in air for the negative control group was 18.72 minutes, it is possible to calculate the total gain in charge over the flight of the drone. The reason why the negative control group is considered in this case is because this scenario involves a drone with mid-flight charging capabilities, which means that a portable battery is unnecessary, and the wind turbines would be the only addition to the drone. This mirrors the setup of the negative control group. In order to calculate the total charge gained, the mean time in air and the rate of charge must be multiplied since the drone would be charging at that rate over the entire range of the drone. The result is a total increase in charge of more than 20%; this is a significant increase in the range of the drone because the range of the drone could potentially be increased by approximately 3.75 minutes.

This indicates that there can be a potential impact that is caused by the use of vertical axis wind turbines on a drone. This impact can be applied to the agricultural industry in the form of affordable drones that are of higher quality. If these vertical axis wind turbines and other advanced sensors were to be added to a cheaper drone, it is possible that the result would be an agricultural drone that is of similar quality as a much more expensive drone. Specifically, the major application of this result is in the mapping and surveillance since the drones that are used for these purposes are more easily able to sustain the use of wind turbines due to their lower weight and technological complexity. However, these drones are also more necessary for farmers without access to much technology. While hybrid drones may include a camera, these drones are extremely expensive, which means that improving the quality of cheaper drones with a single purpose would have a greater impact for many farmers. The increase in the use of technology in the agricultural industry would lead to an increase in productivity and efficiency. With this increase, there would be an advancement in this sector; it could have significant implications in the long run (Jayne and Sanchez, 2021).

While the major aim of this experiment was to apply this idea to the agricultural industry, drones with wind turbines could be applied to any field. One such example is in the military and espionage sectors, where a drone with a much higher range would be very useful; an improvement in the range of such drones would lead to more applications (Emimi et al., 2023). Another example is in the industrial and delivery sectors. As companies have started to use drones to deliver packages, it would be a great enhancement if these drones were to significantly increase their range (Eskandaripour and Boldsaikhan, 2023; Garg et al., 2023). Even though these are specific examples, the main idea of this experiment can be applied to any field, industry, or sector where drones are useful.

Overall, the positive impact that comes from the results of this experiment can be attributed to the dramatic increase in charge caused by the vertical axis wind turbines. Calculations demonstrate that the total increase could be as high as 20% over the 18.72-minute range of the drone. This extension of the range can be valuable in the agricultural industry due to the current lack of affordable technology in this industry. Additionally, this idea can be applied to other fields and sectors to provide a larger positive impact in the long run.

Cost Analysis

A cost analysis can be performed on both an agricultural drone with no wind turbines and an agricultural drone with wind turbines in order to quantify the reduction in the cost of operations. The values that are assumed for this cost analysis are provided in Table 6. The number of cycles was calculated by dividing the side length of the farm by the side length of the field of view, and the total distance traveled was calculated by multiplying the number of cycles by the side length of the farm. Similarly, the time taken by the drone can be found using the distance traveled and the velocity.

	Farm Size [m]	Drone Velocity [m/s]	Drone Field of View [m]	Inspection Frequency
Assumptions	1200 x 1200	20	3.333 x 3.333	1 per day
	Number of Cycles	**Distance [m]**	**Time Taken [hrs]**	**Charging Time [hrs]**
Assumptions	360	432000	6	1.5

Table 6: Assumptions for cost analysis.

Group	Flight Time [min]	Batteries Required	Power per Battery [W]
Drone Without Wind Turbines	20	18	150
Drone With Wind Turbines	24	15	150
Group	**Energy Required per Inspection [kW-hrs]**	**Energy Required per Year [kW-hrs]**	**Total Cost per Year [$]**
Drone Without Wind Turbines	4.05	1478.25	442.475
Drone With Wind Turbines	3.375	1231.875	369.5625

Table 7: Calculations for cost analysis.

Table 7 displays the calculated values that are based on the previous assumptions. The typical flight time of a drone is 20 minutes, and the approximate 20% increase calculated previously causes the new drone to have a flight time of 24 minutes. Using this flight time, the number of batteries was found by dividing the total time taken by the flight time of the drone with one battery. Since a typical battery is charged using 25 volts and 6 amperes, the power required for one battery is the product of these values, or 150 watts. The energy required for one battery is the power required multiplied by the charging time, and the energy required for one entire inspection is the result multiplied by the number of batteries. This is utilized when calculating the energy required for one year since the inspection frequency is one inspection per day for 365 days. Finally, the total cost for one year is calculated using a typical electricity cost of 30 cents per kilowatt-hour.

This means that the total decrease in cost would be $73.9125, or 16.667%. The set-up of the wind turbines would cost $12 for six wind turbines, $5 for the DC motors, and $5 for the mount bracket and wiring. This means that the initial cost would be $22. Therefore, for the first year, the decrease in cost would only be $51.9125, but for subsequent years, the decrease in cost would be $73.9125. Thus, it can be concluded that the use of vertical axis wind turbines on an agricultural drone helps to decrease the electricity cost of the drone, which means that it makes the drone more accessible to farmers.

Conclusion

The data collected from the experiment and the statistical tests that were conducted on this data provided useful results. Due to the statistical summaries, graphs, ANOVA test, and the *t*-test, it can be concluded that the effect of the wind turbines and the additional weight on the time that the drone spent in the air is not statistically significant. Therefore, the difference in time between the three groups could be attributed to chance. This is due to the fact that all of the *p*-values in the ANOVA test and *t*-test are much greater than 0.05, which is the standard for significance. This implies that the use of vertical axis wind turbines on a drone is feasible due to the negligible drag that they generate. Since the vertical axis wind turbines interfere with the airflow surrounding the propellers of the drone, it is expected that there is additional drag due to the wind turbines. However, the statistically insignificant decrease in time caused by the addition of the wind turbines indicates that this drag is negligible, and therefore, wind turbines can be used on a drone. Additionally, it can be concluded from the data that the average voltage generated by the wind turbines is greater than six volts. Since it requires six volts to charge the portable battery for a current of at least one ampere, this result can be seen as a success. The final result of this experiment is that it is feasible to use wind turbines to improve the range of a drone. In the end, the experimental hypothesis was supported, and the null hypothesis was rejected.

However, further research is necessary in order to quantify the effect that the use of wind turbines would have on the range of an agricultural drone. This experiment demonstrated that the use of wind turbines can generate sufficient energy to overcome the increase in drag; this could have significant effects on farmers by providing them with helpful technology that can be produced for a relatively cheaper price. Experiments that utilize drones with mid-flight charging must be performed to support the findings of this research.

Acknowledgments

I would like to acknowledge Ms. Shritha Jagadheeswaran and Mr. Wen Xuan Wong from Wichita State University, who provided feedback on the aerospace engineering portions of my research. I would also like to acknowledge Dr. Michael Young from Santa Barbara City College, who provided feedback on the physics portions of my research.

Works Cited

Ayamga, Matthew, et al. 'Exploring the Challenges Posed by Regulations for the Use of Drones in Agriculture in the African Context'. *Land*, vol. 10, no. 2, MDPI AG, Feb. 2021, p. 164, DOI: 10.3390/land10020164.

Chin, Ruben, et al. 'Plant Disease Detection Using Drones in Precision Agriculture'. *Precision Agriculture*, vol. 24, no. 5, Springer Science and Business Media LLC, Oct. 2023, pp. 1663–1682, DOI: 10.1007/s11119-023-10014-y.

Douak, M., et al. 'Wind Energy Systems: Analysis of the Self-Starting Physics of Vertical Axis Wind Turbine'. *Renewable and Sustainable Energy Reviews*, vol. 81, Elsevier BV, Jan. 2018, pp. 1602–1610, DOI: 10.1016/j.rser.2017.05.238.

Ebaid, Munzer S. Y., et al. 'Feasibility Studies of Micro Wind Turbines Installed on Electric Vehicles as Range Extenders Using Real-Time Analytical Simulation with Multi Driving Cycles Scenarios'. *Advances in Mechanical Engineering*, vol. 15, no. 4, SAGE Publications, Apr. 2023, p. 168781322311659, DOI: 10.1177/16878132231165964.

Eftekhari, Hesam, et al. 'Aerodynamic Performance of Vertical and Horizontal Axis Wind Turbines: A Comparison Review'. *Indonesian Journal of Science and Technology*, vol. 7, no. 1, Universitas Pendidikan Indonesia (UPI), Dec. 2021, pp. 65–88, DOI: 10.17509/ijost.v7i1.43161.

Elsakka, Mohamed Mohamed, et al. 'Response Surface Optimisation of Vertical Axis Wind Turbine at Low Wind Speeds'. *Energy Reports*, vol. 8, Elsevier BV, Nov. 2022, pp. 10868–10880, DOI: 10.1016/j.egyr.2022.08.222.

Emimi, Mohamed, et al. "The Current Opportunities and Challenges in Drone Technology." *International Journal of Electrical Engineering and Sustainability (IJEES)*, vol. 1, no. 3, July 2023, pp. 74–89, https://ijees.org/index.php/ijees/article/view/47.

Eskandaripour, Hossein, and Enkhsaikhan Boldsaikhan. 'Last-Mile Drone Delivery: Past, Present, and Future'. *Drones*, vol. 7, no. 2, MDPI AG, Jan. 2023, p. 77, DOI: 10.3390/drones7020077.

Garg, Vipul, et al. 'Drones in Last-Mile Delivery: A Systematic Review on Efficiency, Accessibility, and Sustainability'. *Transportation Research. Part D, Transport and Environment*, vol. 123, no. 103831, Elsevier BV, Oct. 2023, p. 103831, DOI: 10.1016/j.trd.2023.103831.

Ghirardelli, Mauro, et al. 'Flow Structure around a Multicopter Drone: A Computational Fluid Dynamics Analysis for Sensor Placement Considerations'. *Drones*, vol. 7, no. 7, MDPI AG, July 2023, p. 467, DOI: 10.3390/drones7070467.

Greenblatt, David, and David R. Williams. 'Flow Control for Unmanned Air Vehicles'. *Annual Review of Fluid Mechanics*, vol. 54, no. 1, Annual Reviews, Jan. 2022, pp. 383–412, DOI: 10.1146/annurev-fluid-032221-105053.

Gupta, Abhishek, and Naveen Kumar. 'Energy Regeneration in Electric Vehicles with Wind Turbine and Modified Alternator'. *Materials Today: Proceedings*, vol. 47, Elsevier BV, 2021, pp. 3380–3386, DOI: 10.1016/j.matpr.2021.07.164.

Hafeez, Abdul, et al. ‘Implementation of Drone Technology for Farm Monitoring & Pesticide Spraying: A Review’. *Information Processing in Agriculture*, vol. 10, no. 2, Elsevier BV, June 2023, pp. 192–203, DOI: 10.1016/j.inpa.2022.02.002.

Hassanpour, Mahsa, and Leila N. Azadani. ‘Aerodynamic Optimization of the Configuration of a Pair of Vertical Axis Wind Turbines’. *Energy Conversion and Management*, vol. 238, no. 114069, Elsevier BV, June 2021, p. 114069, DOI: 10.1016/j.enconman.2021.114069.

Hu, Wenyu, et al. ‘Modified Wind Energy Collection Devices for Harvesting Convective Wind Energy from Cars and Trucks Moving in the Highway’. *Energy (Oxford, England)*, vol. 247, no. 123454, Elsevier BV, May 2022, p. 123454, DOI: 10.1016/j.energy.2022.123454.

Irawan, Elysa Nensy, et al. ‘The Effect of Rotor Radius Ratio on the Performance of Hybrid Vertical Axis Wind Turbine Savonius-Darrieus NREL S809’. *Journal of Energy and Power Technology*, vol. 05, no. 01, LIDSEN Publishing Inc, Jan. 2023, pp. 1–12, DOI: 10.21926/jept.2301001.

Jayne, Thomas S., and Pedro A. Sanchez. ‘Agricultural Productivity Must Improve in Sub-Saharan Africa’. *Science (New York, N.Y.)*, vol. 372, no. 6546, American Association for the Advancement of Science (AAAS), June 2021, pp. 1045–1047, DOI: 10.1126/science.abf5413.

Li, Shoutu, et al. ‘Experimental Investigation on Noise Characteristics of Small Scale Vertical Axis Wind Turbines in Urban Environments’. *Renewable Energy*, vol. 200, Elsevier BV, Nov. 2022, pp. 970–982, DOI: 10.1016/j.renene.2022.09.099.

Li, Yan, et al. ‘A Review on Numerical Simulation Based on CFD Technology of Aerodynamic Characteristics of Straight-Bladed Vertical Axis Wind Turbines’. *Energy Reports*, vol. 9, Elsevier BV, Dec. 2023, pp. 4360–4379, DOI: 10.1016/j.egyr.2023.03.082.

Monteiro, Martín, et al. ‘Simple Physics behind the Flight of a Drone’. *Physics Education*, vol. 57, no. 2, IOP Publishing, Mar. 2022, p. 025029, DOI: 10.1088/1361-6552/ac484a.

Redchyts, Dmytro, et al. ‘Aerodynamic Performance of Vertical-Axis Wind Turbines’. *Journal of Marine Science and Engineering*, vol. 11, no. 7, MDPI AG, July 2023, p. 1367, DOI: 10.3390/jmse11071367.

Rejeb, Abderahman, et al. ‘Drones in Agriculture: A Review and Bibliometric Analysis’. *Computers and Electronics in Agriculture*, vol. 198, no. 107017, Elsevier BV, July 2022, p. 107017, DOI: 10.1016/j.compag.2022.107017.

Wilberforce, Tabbi, et al. ‘Wind Turbine Concepts for Domestic Wind Power Generation at Low Wind Quality Sites’. *Journal of Cleaner Production*, vol. 394, no. 136137, Elsevier BV, Mar. 2023, p. 136137, DOI: 10.1016/j.jclepro.2023.136137.

Zhao, Peidong, et al. ‘Investigation of Fundamental Mechanism Leading to the Performance Improvement of Vertical Axis Wind Turbines by Deflector’. *Energy Conversion and Management*, vol. 247, no. 114680, Elsevier BV, Nov. 2021, p. 114680, DOI: 10.1016/j.enconman.2021.114680.

The Psychological and Neurological Effects of Polycystic Ovary Syndrome By Aarna Jain

Abstract

This paper explores the broad impacts of PCOS, focusing on its psychological and neurological effects. Women with PCOS are at a higher risk of depression and anxiety, influenced by factors such as hormonal imbalances, stress, and vitamin D deficiency. Neurologically, PCOS is associated with disruptions in neurotransmitter functions and conditions like Central Sensitivity Syndrome (CSS), which contribute to pain sensitivity and altered stress responses. The daily lives of women with PCOS are significantly affected, with symptoms impacting their physical, emotional, and social well-being. By highlighting these challenges, this paper aims to raise awareness and advocate for comprehensive care strategies to better support women living with PCOS.

Introduction

Polycystic Ovary Syndrome (PCOS) is a complex endocrine disorder that affects around 10% of women. The physiological impacts of PCOS extend beyond reproductive health, often including metabolic complications such as insulin resistance, obesity, and type 2 diabetes (Legro et al., 2013). Additionally, the psychological and neurological implications of PCOS, such as the increased prevalence of mood disorders and neurological impairments, are gaining research attention (Ehrmann, 2005). This paper first describes PCOS and then highlights research on both psychological and neurological effects observed in women with PCOS. Finally, the review will highlight how this affects the daily lives of women with PCOS, with the overall goal of increasing awareness of PCOS and helping to address better the health needs of women living with the syndrome.

What is PCOS?

Polycystic ovary syndrome (PCOS) is a complex endocrine hormonal disorder affecting women of reproductive age, characterized by hormonal imbalances, metabolic disturbances, and reproductive dysfunctions. It is also characterized by irregular menstrual cycles, elevated levels of male hormones, and the presence of cysts on the ovaries (DuRant and Leslie, 2018). Diagnosis typically requires meeting at least two of three criteria: irregular ovulation or lack thereof, signs of increased male hormones, and polycystic ovaries visible on ultrasound. PCOS can lead to fertility issues, irregular periods, weight gain, insulin resistance, and heightened risks of Type 2 diabetes and cardiovascular disease. Early detection and management are crucial for symptom control and reducing associated health risks. These health risks are not just limited to those described above, but can also include increased risk of psychological disorders.

Psychological Effects on Women from PCOS

Depression

Emerging evidence suggests a significant association between PCOS and depression, highlighting the heightened risk of depressive symptoms among affected women (Chaudhari et al., 2018). The high prevalence rates are evidenced through a comprehensive study examining anxiety and depression in 70 women diagnosed with PCOS, aged between 18 and 45 years (Chaudhari et al., 2018). Using standardized rating scales, the authors reported prevalence rates of anxiety and depression at 38.6% and 25.7%, respectively, within the study cohort (Chaudhari et al., 2018). Notably, the presence of acne, a common cutaneous manifestation of PCOS, was found to be associated with an increased likelihood of depression, suggesting a potential link between dermatological symptoms and mental health outcomes in PCOS(Chaudhari et al., 2018).

Additional research has been done to understand factors associated with depression in women with PCOS. For example, Naqvi et al. (2015) conducted an extensive study investigating predictors of depression among women with PCOS. Their cross-sectional analysis of 114 women with PCOS revealed that disturbed sleep and a family history of depression were significant risk factors for depression, as assessed by the Personal Health Questionnaire (Naqvi et al., 2015). Additionally, the study highlighted the impact of vitamin D deficiency on the severity of depressive symptoms, albeit without influencing the likelihood of depression diagnosis (Naqvi et al., 2015). These findings underscore the multifactorial nature of depressive symptoms in PCOS, encompassing both psychosocial and physiological determinants.

Hormone dysregulation and stress have also been found to play a role in depression in women with PCOS (Hollinrake et al. [2007]). Research by Hollinrake et al.(2007) elucidated the role of immune system dysregulation and pro-inflammatory marker activity during stressful periods as contributing factors to the development of depressive symptoms in women with PCOS. The study, which was conducted at a university hospital, aimed to estimate the prevalence of depressive disorders in women with Polycystic Ovary Syndrome (PCOS) and evaluate the correlation between depression, hyperandrogenism, and metabolic markers. The study included 103 women diagnosed with PCOS based on the Rotterdam criteria and 103 control subjects without PCOS, seen for annual exams. Using the Primary Care Evaluation of Mental Disorders Patient Health Questionnaire (PRIME-MD PHQ) and the Beck Depression Inventory, the study found that women with PCOS had a significantly higher risk of depressive disorders compared to controls (21% vs. 3%; odds ratio 5.11; 95% CI 1.26-20.69; $P<.03$). This increased risk was independent of obesity and infertility, with an overall risk of 4.23% (95% CI 1.49-11.98; $P<.01$). Depressed PCOS subjects had a higher body mass index (BMI) and evidence of insulin resistance ($P<.02$) compared to nondepressed PCOS subjects. The study concludes that women with PCOS are at a significantly increased risk of depressive disorders and recommends routine screening for depression in this population.

The findings from these studies collectively emphasize the intricate relationship between PCOS and depression, shedding light on both the prevalence and predictors of depressive symptoms in affected women. The co-occurrence of PCOS and depression poses significant challenges for healthcare providers in managing the holistic well-being of their patients. The complex causes of depression in PCOS require a comprehensive approach to assessment and intervention, including psychosocial support, lifestyle changes, and appropriate medication.

Anxiety

Recent studies have also highlighted a significant association between PCOS and anxiety, with affected individuals experiencing heightened levels of anxiety compared to the general population. For example, a study by Elsenbruch et al. (2011), conducted with a cohort of 70 female participants, found that 38.6% of study participants with PCOS also met criteria for an anxiety disorder, as assessed by standardized diagnostic criteria. Logistic regression analysis revealed a relationship between PCOS symptoms and the probability of anxiety, with variables such as alopecia, hirsutism, acne, obesity, irregular menstruation, acanthosis, and infertility (limited to married females) identified as significant predictors of anxiety (Elsenbruch et al., 2011).

Furthermore, another study by Laura G Cooney et al. explored the co-involvement of psychological and neurological abnormalities in infertility associated with PCOS. This study compared anxiety levels between women with PCOS and healthy controls, revealing a higher prevalence of anxiety among women with PCOS compared to the control group. These findings emphasize the multifactorial nature of anxiety in PCOS, encompassing both physiological and psychosocial determinants.

In addition to the work above, researchers have also relied on insights from clinical observations and patient narratives better to understand the complex relationship between PCOS and anxiety (Dason et al.). Through listening to patient experiences, it was clear that stressors such as disease symptoms, concerns about fertility, and body image issues all contribute to heightened anxiety levels. These observations highlight the importance of adopting a holistic approach to PCOS management, integrating mental health screening and interventions into routine care practices.

The findings from these studies collectively underscore the significant burden of anxiety experienced by women with PCOS, emphasizing the need for comprehensive assessment and management strategies. Similar to depression, the nature of anxiety in women with PCOS appears to be multifactorial, encompassing both physiological and psychosocial determinants, necessitating a multidisciplinary approach to care. Healthcare providers can optimize patient outcomes and improve overall quality of life by addressing both physical and mental health aspects of PCOS.

Neurological effects on women with PCOS

The neuroendocrine imbalance in Polycystic Ovary Syndrome (PCOS) involves dysregulation of several hormonal axes, including the hypothalamic-pituitary-gonadal (HPG) axis, the hypothalamic-pituitary-adrenal (HPA) axis, and insulin resistance. In PCOS, disruptions in the HPG axis lead to abnormal levels of gonadotropins (such as luteinizing hormone and follicle-stimulating hormone), which in turn affect ovarian function and contribute to menstrual irregularities and infertility. Hyperandrogenism, a hallmark of PCOS, results from both ovarian and adrenal sources, further exacerbating hormonal imbalances. Additionally, dysregulation of the HPA axis contributes to the pathogenesis of PCOS, with increased levels of cortisol observed in affected individuals. This dysregulation is linked to both metabolic dysfunction and psychological symptoms commonly associated with PCOS, such as anxiety and depression. Insulin resistance is another key feature of PCOS, contributing to hyperinsulinemia and compensatory hyperglycemia. Insulin resistance further exacerbates hyperandrogenism and disrupts ovarian function, contributing to the pathophysiology of PCOS. (Ruddenklau & Campbell, 2019). Furthermore, insulin resistance can affect neurotransmitter levels by altering glucose metabolism in the brain and influencing the transport of amino acids that are precursors to neurotransmitters.

PCOS is also associated with alterations in several additional hormones, including androgens (such as testosterone), and sex hormone-binding globulin (SHBG), which can affect neurotransmitter function. For instance, elevated testosterone in PCOS can influence neurotransmitter systems involved in mood regulation, such as serotonin and dopamine. (Chaudhari, N. & Nampoothiri, L., 2017). Additionally, changes in sex hormone-binding globulin (SHBG), which binds to and regulates the activity of sex hormones, can affect the availability and activity of these hormones, further impacting neurotransmitter function. (Chaudhari, N. & Nampoothiri, L., 2017)

Central Sensitivity Syndrome (CSS) is a neurological condition in women with Polycystic Ovary Syndrome characterized by increased sensitivity to pain, altered stress response, and various other symptoms. It involves dysregulation of the central nervous system, particularly the hypothalamus-pituitary-adrenal (HPA) axis and the autonomic nervous system (ANS). In women with PCOS, CSS manifests as heightened sensitivity to pain, such as increased perception of pressure pain and reduced pain threshold. (Miller et al., 2020). This heightened sensitivity may be linked to dysfunctions in the HPA axis and ANS, which regulate stress response and pain perception. The dysregulation of these systems in PCOS may lead to alterations in neurotransmitter levels, particularly serotonin and dopamine, which play key roles in modulating pain perception and stress response. Additionally, changes in hormone levels, such as increased cortisol and insulin resistance, may also contribute to the development of CSS in PCOS. (Miller et al., 2020).

How PCOS affects the daily lives of women

Unsurprisingly, given the research described above, PCOS can have a tremendous effect on the daily lives of women living with the syndrome. The distressing physical symptoms, such as irregular menstrual cycles, excessive hair growth, acne, and changes in metabolism (i.e., weight gain and difficulty losing weight), can significantly affect a woman's self-image and confidence, and increase the risk of mental difficulties, including anxiety, depression, and low self-esteem. (DuRant and Leslie, 2018). Additionally, PCOS can impair cognitive function, leading to concentration difficulties, memory problems, and decreased cognitive performance [need citation], further hindering daily functioning and productivity. The combination of physical, emotional, and cognitive symptoms can interfere with social activities and relationships, causing women to feel isolated or ashamed, and leading to withdrawal from social interactions. This social withdrawal can result in a diminished quality of life. PCOS is also associated with increased risks of other health conditions, such as Type 2 diabetes, cardiovascular disease, and sleep apnea, which can further complicate the management of the syndrome and contribute to a greater overall health burden. Failing to address PCOS can lead to many issues like infertility and an overall burden in life.

A case example of what living with PCOS might look like in real life can be seen in a *Psychology Today* article written by Georgia Witkin, which presents the story of Julie, a 27-year-old woman who went her whole life not knowing the reason for the unbearable symptoms she was experiencing. Julie didn’t experience the common external symptoms such as weight gain, excessive hair growth, hair loss on her head, or acne. Instead, her symptoms were all internal. During puberty, she began suffering from headaches and migraines and was diagnosed with anxiety and depression. She also had sleep problems and was put on birth control pills to manage painful periods and regulate her menstrual cycle. Despite these symptoms, no one considered PCOS. While in college, Julie experienced breakthrough bleeding despite being on birth control. Her doctor changed her prescription and attributed her anxiety and depression to dating and school stress, again missing the connection to PCOS. Years later, when she had breakthrough bleeding again, Julie knew something was wrong. She realized that simply changing her birth control pills wasn't the solution and might even worsen her anxiety and depression. She confided in a colleague, who encouraged her to seek a full medical evaluation, and at her next doctor's appointment, Julie requested comprehensive testing. The doctor discovered 15 egg follicles on each ovary during a transvaginal ultrasound, confirming PCOS (Witkin).

Conclusion

PCOS is a complex disorder that significantly affects many aspects of a woman’s life. Metabolic issues like insulin resistance and high levels of hormones not only impact physical health but also contribute to mental health problems such as anxiety, depression, and neurological difficulties. The hormonal imbalances in PCOS disrupt both the endocrine and nervous systems, leading to a wide range of complications. Managing PCOS requires a well-rounded approach that includes medical care, support for mental health, and lifestyle adjustments. While significant strides have been made in understanding and managing PCOS, many studies have small sample sizes, focus more on physical aspects, and lack consistency in diagnostic criteria and treatment protocols. Future research should aim to include larger, more diverse populations, adopt holistic approaches, standardize diagnostic and treatment guidelines, conduct long-term studies, and focus on both physical and mental health impacts. By addressing these gaps, healthcare providers can significantly improve the quality of life for women with PCOS, offering them the support and resources they need to manage this complex disorder.

Works Cited

Chaudhari, Nishant, and Lakshmi Nampoothiri. "Neurotransmitter Alterations in PCOS: Role of Insulin and Androgen Imbalance." *Journal of Endocrinological Investigation*, vol. 40, no. 10, 2017, pp. 1085-1092.

Chaudhari, Nishant, et al. "Anxiety and Depression in Women with Polycystic Ovary Syndrome in Comparison to Women with Infertility: A Cross-Sectional Study." *Indian Journal of Psychological Medicine*, vol. 40, no. 2, 2018, pp. 140-146.

DuRant, Juliet, and Nancy Leslie. "Polycystic Ovary Syndrome: Beyond Reproductive Health." *Journal of Women's Health*, vol. 27, no. 5, 2018, pp. 603-612.

Ehrmann, David A. "Polycystic Ovary Syndrome." *New England Journal of Medicine*, vol. 352, no. 12, 2005, pp. 1223-1236.

Elsenbruch, Susanne, et al. "Impaired Quality of Life, Increased Anxiety and Depression in Patients with Polycystic Ovary Syndrome." *Human Reproduction*, vol. 18, no. 6, 2003, pp. 1383-1389.

Hollinrake, Erin, et al. "Increased Risk of Depressive Disorders in Women with Polycystic Ovary Syndrome." *Fertility and Sterility*, vol. 87, no. 6, 2007, pp. 1369-1376.

Legro, Richard S., et al. "Diagnosis and Treatment of Polycystic Ovary Syndrome: An Endocrine Society Clinical Practice Guideline." *Journal of Clinical Endocrinology & Metabolism*, vol. 98, no. 12, 2013, pp. 4565-4592.

Miller, Kristen E., et al. "Central Sensitivity Syndromes in Women with Polycystic Ovary Syndrome: A Systematic Review." *Pain Medicine*, vol. 21, no. 1, 2020, pp. 105-115.

Naqvi, Syed Haider, et al. “Predictors of Depression in Women with Polycystic Ovary Syndrome.” *Archives of Women's Mental Health*, vol. 18, no. 1, Sept. 2014, pp. 95–101, https://doi.org/10.1007/s00737-014-0458-z.

Ruddenklau, Amy, and Rebecca E. Campbell. "Neuroendocrine Impairment in Polycystic Ovary Syndrome." *Steroids*, vol. 144, 2019, pp. 85-88.

Witkin, Georgia. “PCOS: The Mental, Emotional, and Physical.” *Psychology Today*, 5 Sept. 2018

Hollinrake E, Abreu A, Maifeld M, Van Voorhis BJ, Dokras A. Increased risk of depressive disorders in women with polycystic ovary syndrome. Fertil Steril. 2007 Jun;87(6):1369-76. doi: 10.1016/j.fertnstert.2006.11.039. Epub 2007 Mar 29. PMID: 17397839.

Cooney, Laura G. et al. “High prevalence of moderate and severe depressive and anxiety symptoms in polycystic ovary syndrome: a systematic review and meta-analysis.” Human reproduction (Oxford, England) vol. 32,5 (2017): 1075-1091. doi:10.1093/humrep/dex044

Dason, Ebernella Shirin, et al. “Diagnosis and Management of Polycystic Ovarian Syndrome.” CMAJ, vol. 196, no. 3, 29 Jan. 2024, pp. E85–E94, www.cmaj.ca/content/196/3/E85, https://doi.org/10.1503/cmaj.231251.

Building Bridges: The History and Future of Jewish-Muslim Collaboration By Tohar Liani

Abstract

This study delves into the potential future of collaboration and dialogue between Jewish and Muslim groups. It explores historical periods of teamwork, influential leaders, and organizations, as well as the impact of present-day initiatives. By assessing achievements and hurdles, this paper aims to underscore the necessity for ongoing dialogues and cooperation between Jewish and Muslim communities. It also discusses strategies for strengthening this relationship, focusing on education, interfaith efforts, and community involvement.

Introduction

The connection between Muslim communities has a rich history with moments of collaboration alongside conflict. Understanding the evolution of their collaboration and dialogue is essential for fostering peace and mutual respect -=in today's world. This paper seeks to delve into the background of notable individuals, and contemporary endeavors that have influenced Jewish-Muslim relations while proposing ways to advance these initiatives in the future.

The importance of this exploration lies in its potential to nurture an inclusive society through interfaith comprehension and collaboration. Over time interactions between Muslim communities have varied from intellectual exchanges and cultural partnerships to confrontations and misunderstandings. The rich history shared by these two faith communities offers lessons and insights that can guide current and future endeavors to connect them.

This paper aims to uncover the factors that enabled collaborations and address the challenges ahead by exploring specific historical periods and notable instances of cooperation. Beginning with a focus on the era particularly in places like Al Andalus, where Jewish and Muslim scholars collaborated, contributing to a vibrant intellectual and cultural atmosphere. This era, known as the Golden Age of Jewish-Muslim relations serves as an example of potential coexistence and mutual enrichment.

Progressing through time the research delves into Muslim relations in the Ottoman Empire highlighting the millet system that allowed for religious autonomy and fostered peaceful coexistence. The 20th century introduced opport=unities and obstacles for collaboration, between Jews and Muslims. In the aftermath of World War II and the Holocaust, there were instances of individuals and communities supporting Jews demonstrating acts of courage and solidarity during challenging times. In decades increased interfaith dialogues and organizations emerged as part of efforts to bridge divides and cultivate understanding.

The groundwork laid by these initiatives has set the stage for efforts that adapt to the changing dynamics of our world. Today there are grassroots movements and organizations committed to nurturing collaboration between Jewish and Muslim communities. These endeavors encompass projects addressing community needs and global gatherings uniting leaders and scholars from both faiths. Education and engaging the youth are pivotal in shaping mindsets and attitudes. Moreover, political and social advocacy have emerged as tools for addressing shared concerns and countering discrimination.

In the discussion section, we explore the challenges and opportunities ahead for Muslim collaboration. While political tensions and historical grievances present hurdles, they also offer chances for deeper dialogue and mutual understanding. We delve into how technology facilitates communication, emphasizing platforms' potential to bridge communities across distances. The paper also suggests pathways to enhance Jewish-Muslim relations, stressing education, grassroots initiatives, and leadership in fostering lasting collaboration.

The conclusion recaps the paper's findings underscoring the importance of sustained efforts to connect Jewish and Muslim communities. By drawing lessons from history and embracing strategies we can work towards a more inclusive and harmonious tomorrow. The article highlights the importance of promoting peace, mutual respect, and understanding through interfaith dialogue and collaboration in our interconnected world.

Historical Context

The discussion on Muslim cooperation delves into historical contexts, such as Al Andalus during the medieval era. Al Andalus, also called Muslim Spain was a region where Jewish and Muslim scholars collaborated harmoniously contributing to an intellectual and cultural landscape. This era, often referred to as the Golden Age of Jewish-Muslim relations serves as an example of how coexistence and mutual learning can thrive.

The Muslim rulers of Al Andalus under the Umayyad Caliphate's rule were known for their relatively inclusive policies towards religious diversity. Jews, Christians, and Muslims lived together peacefully. Engaged in scholarly exchanges that benefited all communities involved. One prominent figure from this period is Moses Maimonides, a century-Jewish philosopher, theologian and physician. Maimonides, also known as Rambam, was born in Cordoba within Al Andalus before relocating to Egypt. His writings like "The Guide for the Perplexed" attest to his engagement with Islamic philosophy and showcase the intellectual cross-pollination that characterized that time. Maimonides' ideas in philosophy were greatly impacted by the teachings of Muslim thinkers like Al Farabi and Avicenna. This intellectual exchange thrived due to the coexistence and mutual respect prevailing during that era.

The prosperous era of Al Andalus concluded with the Reconquista, when Spain was reclaimed by Christians leading to Granada's fall in 1492. The expulsion of Jews and Muslims signified the end of this period of shared achievements. Nevertheless, the influence of Al Andalus endured, shaping Muslim ideologies for years to follow. The cultural accomplishments from this era highlight the potential for collaboration and mutual growth when communities unite with respect and curiosity.

Looking ahead, the interactions between Jews and Muslims in the Ottoman Empire present another example of harmony and cooperation. Spanning from 1299 to 1922 the Ottoman Empire encompassed populations including Jews and Muslims. The Ottoman millet system allowed religious communities like Jews to govern themselves according to their laws and uphold their traditions. This system fostered coexistence among different religious groups, within the empire.

Jewish communities flourished in cities like Istanbul, Salonika, and Smyrna, playing vital roles in trade, governance, and cultural affairs. Among these cities, Salonika stood out as home to one of the most dynamic Jewish populations globally. The Jewish residents of Salonika engaged in economic pursuits such as trade, finance, and industry. Additionally, the cultural landscape of the community in Salonika flourished with numerous synagogues, schools, and cultural establishments.

The Ottoman Empire's tolerant stance toward religious minorities fostered a level of coexistence and cooperation that was uncommon during that era in other regions. However, it is important to acknowledge that this harmonious coexistence faced challenges at times. Periods of tension and conflict arose despite the empire's tolerant policies toward religious minorities. Nonetheless, the Ottoman millet system provided a structure for coexistence and mutual respect that allowed both Jewish and Muslim communities to prosper.

The 20th century presented opportunities as well as trials for collaboration between Jews and Muslims. Following World War II and the Holocaust, there were instances where Muslim individuals and communities extended support to Jews demonstrating acts of courage and solidarity, amidst adversity. During the time of the Holocaust, several countries with a majority of population and individuals extended help and refuge to Jews who were escaping persecution. In Albania, which is mostly a Muslim nation many Muslims bravely risked their lives to provide shelter and protection to Jewish families. The Albanian custom of "besa," focusing on honor and taking care of guests, played a role in these acts of courage.

Another remarkable instance is seen at the Grand Mosque in Paris, where Jews found refuge during the Holocaust. The mosque's leader, Si Kaddour Benghabrit, issued documents confirming Jewish identity to help them evade capture and deportation by the Nazis. These courageous and united efforts showcase the potential for collaboration between faiths even in humanity's darkest periods.

In the part of the 20th century there was an increase in interfaith dialogues and organizations aiming to bridge gaps and promote understanding between Jewish and Muslim communities. One notable milestone was the creation of the World Congress of Imams and Rabbis for Peace in 2005. This initiative brings together leaders from both religions to advocate for peace and mutual understanding. It serves as a platform for discussions and joint efforts on shared concerns towards objectives.

Contemporary Efforts

Today numerous grassroots movements and organizations are actively working towards promoting collaboration between Muslim communities. Various projects are being implemented, ranging from community endeavors to global conferences that bring together leaders and scholars from both faith communities. For instance, the Jewish Conference (MJC) is a notable annual event that aims to foster trust and understanding among young leaders from Jewish and

Muslim backgrounds. The conference serves as a platform for discussions where participants can share their experiences, challenge stereotypes, and collaborate on joint projects.

Another impactful initiative is the Sisterhood of Salaam Shalom, an organization that unites Muslim women to cultivate personal connections and advocate for social justice. By creating spaces for dialogue, promoting mutual respect, and addressing issues like anti-Semitism and Islamophobia, the organization strives to forge enduring bonds between women of both faiths. Through meetings, collaborative activities, and community service endeavors, the Sisterhood of Salaam Shalom seeks to empower Jewish and Muslim women to work together toward positive change.

Education and youth involvement are components of these initiatives as they influence the beliefs and perceptions of future generations. Interfaith educational programs such as interfaith schools and exchange initiatives provide people with opportunities to learn about each other's cultures and religions, in a supportive environment that encourages inclusivity. These educational endeavors help break down stereotypes and foster mutual respect from an age.

One instance of such a project is the Interfaith Youth Core (IFYC) a group collaborating with colleges and universities to foster interfaith harmony and comprehension. IFYC initiatives inspire students to partake in interfaith discussions, engage in community service undertakings, and cultivate leadership abilities. By nurturing connections and promoting cooperative efforts IFYC strives to cultivate a new generation of leaders dedicated to advancing interfaith understanding and collaboration.

Political and social advocacy has also emerged as a means of addressing shared concerns and combating prejudice. Collaborative advocacy endeavors by Muslim organizations have made a notable impact on public opinion and policy. In the United States for example Jewish and Muslim advocacy groups have cooperated on issues like combating hate crimes, safeguarding liberties, and advocating for immigration reform. These joint endeavors showcase the potential of cooperation in effecting positive societal changes.

The Jewish Muslim Alliance of Advocacy and Service (JMAAS) is an organization focusing on social advocacy. JMAAS united Muslim leaders and activists to tackle mutual concerns such as countering anti-Semitism and Islamophobia, advocating for social justice, and championing marginalized communities. Through its efforts, J MAAS aims to amplify the voices of both communities while presenting a united front against discrimination and injustice.

Challenges & Opportunities

While there have been achievements in fostering collaboration between Jewish and Muslim communities there are still obstacles to overcome. Political conflicts, particularly related to the Palestinian situation often strain the relationship between these two groups. The ongoing Palestinian conflict has been a source of tension and division affecting the dialogue and partnership between followers of both faiths. Nevertheless, it is crucial to acknowledge that this conflict also offers opportunities for understanding and communication. By addressing the underlying issues of the conflict and striving toward a peaceful resolution, Jewish and Muslim communities can lay down a solid groundwork for ongoing cooperation.

The significance of technology in facilitating communication and collaboration is further explored, emphasizing how virtual platforms can bridge communities separated by distance. In today's era, social media, online conferences, and educational resources on the internet can enhance global dialogue and teamwork between Jewish Muslim communities. These platforms enable individuals from regions to connect, exchange ideas, and collaborate on joint endeavors. The digital realm also opens avenues for interfaith projects like online discussions forums, webinars and digital storytelling ventures.

An exemplary instance of a virtual interfaith project is the Abrahamic Family Houses series of online dialogues. This initiative gathers scholars and leaders from Muslim and Christian backgrounds to engage in conversations, on shared topics of interest and concern. The digital platform enables a range of people to take part and interact, connecting with individuals who may not be able to join physical gatherings. Through the use of technology, the Abrahamic Family House seeks to encourage a dialogue on interfaith collaboration and empathy.

In the realm of improving relations between Muslim communities, the focus is on education, grassroots movements, and leadership to nurture ongoing cooperation. Education stands out as a factor in fostering collaboration between Jews and Muslims. Educational institutions should integrate teachings on interfaith matters. Encourage student exchange programs to cultivate understanding from an early age. Interfaith educational endeavors that highlight values, cultural appreciation, and conflict resolution can lay a solid groundwork for mutual respect and joint efforts.

Grassroots initiatives are pivotal in promoting collaboration between the Muslim communities. Backing and expanding movements can establish a sturdy base for continuous cooperation. Local projects addressing community concerns like food security, education, and social justice can serve as blueprints for larger-scale endeavors. Through work towards shared objectives, Jewish and Muslim groups can foster trust and bring about positive transformations at the local level.

It is crucial to urge religious leaders to persist in advocating for interfaith dialogue. Endorsements, from figures and their active participation, can spur broader community engagement while lending credibility to collaborative ventures. Political and religious leaders wield platforms with significant influence; their backing of interfaith initiatives can greatly impact public perception and policy decisions. By championing dialogue across faiths and fostering collaboration, leaders contribute to creating an atmosphere of mutual respect and empathy.

Conclusion

The history of collaboration and dialogue between the Muslim communities is filled with instances of respect and shared accomplishments. Despite facing challenges there is potential for future collaboration. By drawing from experiences embracing modern initiatives and focusing on education and grassroots movements both communities can continue to forge bonds of understanding and cooperation. This ongoing endeavor plays a role in promoting peace and respect in an interconnected world.

In summary, the exploration of Muslim collaboration emphasizes the importance of sustaining efforts to nurture comprehension and harmony between these two faith groups. Lessons learned from instances of coexistence and partnership, such as those seen in Al Andalus and the Ottoman Empire, offer valuable insights for present-day endeavors. The emergence of interfaith dialogues and organizations combined with grassroots activities and educational programs showcases the potential for transformations.

By confronting obstacles head-on and capitalizing on opportunities brought forth by technology and global connectivity Jewish and Muslim communities can collaborate toward establishing an inclusive society marked by harmony. Upholding a dedication to interfaith dialogue is key to shaping a future where mutual understanding prevails. Looking ahead it remains vital to continue investing in education supporting grassroots projects and rallying as well as religious leaders to champion interfaith unity. The opportunity for collaboration and communication between Muslim communities holds great promise with the positive outcomes reaching far beyond just those directly involved. Through fostering peace, empathy, and mutual regard both groups can play a role in creating a fairer and kinder world. The ongoing process of connecting these two faith traditions highlights the dedication to fostering dialogue and working together showcasing the lasting impact of cooperation.

Works Cited

Abitbol, Michel. *The Jews of North Africa during the Second World War*. Wayne State University Press, 1989.

Benbassa, Esther. *The Jews of France: A History from Antiquity to the Present*. Princeton University Press, 1999.

Firestone, Reuven. *An Introduction to Islam for Jews*. The Jewish Publication Society, 2008.

Hussain, Amir. *Muslims and the Making of America*. Baylor University Press, 2016.

Maimonides, Moses. *The Guide for the Perplexed*. Translated by M. Friedlander, Dover Publications, 1956.

Sauer, Joshua. "Interfaith Dialogue in the Modern World." *Journal of Religious Studies*, vol. 45, no. 2, 2019, pp. 123-145.

The Association of ESG Scores with Firm Profitability: A Case Study On Technology Stocks By Darren Robbins

Abstract

The concepts of environmental, social, and governance (ESG) investing and sustainable business have emerged as key focuses for corporate strategies. This research paper examined the relationship between ESG scores and a company's profitability. I seek to answer the research question: How does a higher ESG score impact technology firms' profitability? In my analysis, I performed a simple linear regression to test whether there is a positive correlation between ESG score and firm profitability. My variables were the overall ESG score, further stratified into an environmental, social, and governmental score, and firm profit margin (%). No significant relationship existed between a company's overall ESG score and profit margin. However, there is evidence of a positive relationship between a company's social score and profit margins. This suggests that putting a greater emphasis on ESG metrics in operations may drive positive financial returns.

1. Introduction

The concept of responsible investing began in 1970 with the initiative of socially responsible investing (SRI), which allowed investors to represent their portfolios based on their beliefs (Krantz, 2024). Nevertheless, it was not until 1990 that the term ESG (environmental, social, and governance) was introduced into the investing world. Gradually, investors began to understand that a company's operating results could benefit if that business focused on ESG matters. Initially, investors were focused on headline environmental factors. The effort to quantify ESG metrics led to the creation of the Global Reporting Initiative (GRI) in 1997. In 2000, the Millennium Development Goals (MDGs) were created, which outlined eight international development goals to be achieved by 2015. That same year, the Carbon Disclosure Project (CDP) was the first organization to ask companies to report on their environmental impact. The MDGs laid the groundwork for the first ESG reporting practices, which gained popularity around 2002. In the following decade, ESG reporting agencies expanded, and companies emphasized ESG metrics more heavily. Some notable reporting agencies are the Climate Disclosure Standards Board (CDSB), the Principles for Responsible Investment (PRI), and the Sustainability Accounting Standards Board (SASB). Now, ESG scores are widely used across many industries and corporations.

Currently, ESG scores measure a company's efforts to contribute to a more sustainable and ethical environment. They provide insight into the metrics companies use in their operations to improve their financial output and become more socially responsible. In recent years, ESG ratings and indices have become more common. For instance, Morgan Stanley Capital International (MSCI) provides various ESG indices, enabling investors to monitor companies' ESG performance. These indices are favored by investors seeking to incorporate ESG factors into their portfolios. Given the growing concerns about climate change and social issues, ESG considerations will remain pivotal in how companies and investors assess their performance and operations. However, some companies do not value the importance of ESG ratings as they do not believe it drives shareholder value. My research project investigates how a firm's profit margin correlates to its ESG scores, focusing on technology stocks.

ESG scores encapsulate the company's efforts to manage environmental, social, and governance issues for the future and measure their sustainability and ethics. These scores are important for investors as they provide insight and evidence of a company's long-term performance. In finance, these scores help make investment decisions for socially responsible investors. Furthermore, investors are progressively integrating these aspects into their analysis procedures to recognize significant risks and potential avenues for growth, extending beyond financial considerations. By considering environmental and social impacts, investors can identify companies more likely to sustain growth and avoid risks associated with unsustainable practices. The constituents of ESG scores are shown in **Table 1**.

Environmental issues	**Social issues**	**Governance issues**
Carbon footprint	Labor practices	Board diversity and structure
Energy efficiency	Pro-diversity rights	Executive compensation
Renewable energy usage	Human rights	Shareholder rights
Water usage	Community relations	Business ethics
Pollution	Health and Safety	Risk management
Waste management		Supply chain management
Biodiversity impact		

Table 1. This table lists the issues associated with each section of ESG.

2. Conceptual Framework

A higher ESG score is attributed to how sustainable and ethical a company is in its operations and how well it manages ESG-associated risks. The stock market should reward companies for their efforts to contribute to a more sustainable future. Additionally, consumers are more likely to support a company that has more environmentally friendly products and whose operations are more ethical, increasing their revenue. The companies can mitigate any incremental expenses arising from stricter compliance and additional mitigation efforts related to ESG matters, resulting in increased profit margins.

Given that consumers are more aware of companies' ethical practices and behaviors, they will likely be more willing to support the business (Martins, 2024). I hypothesize that a higher ESG score will increase profit margins for companies because consumers may change their purchasing behaviors depending on the reputation of the company they are purchasing from. This is driven by the idea that integrating environmental, social, and governance considerations into business strategies may allow companies to benefit financially in the long run. Additionally, companies with strong ESG performances may consequently mitigate various risks, including regulatory fines, lawsuits, reputational damage, and operational disruptions. By addressing environmental and social issues and adopting good governance practices, companies can minimize these risks, which may translate into cost savings and improved profitability over time.

There is a growing body of research that seeks to quantify the linkage between adherence to ESG goals and improved firm performance. A recent research study conducted by Rebecca Doherty, a writer for McKinsey & Company, showed that "companies that achieve better growth and profitability than their peers while improving sustainability and ESG outgrow their peers and exceed them in shareholder returns." (Doherty et al.). Furthermore, a study led by Mahmut Aydoğmuş and performed by the World Federation of Exchanges and Borsa Istanbul suggests "...that overall ESG combined score is positively and significantly associated with firm value…and firm profitability." ESG scoring enables investors, stakeholders, and customers to assess a company's sustainability initiatives and ethical conduct. More specifically, sustainable investing is growing in popularity. According to a Global Sustainable Investment Alliance (GSIA) report, "global sustainable investment assets reached $35.3 trillion in 2020, up 15% from 2018." Businesses that perform well on ESG scoring are more likely to attract investment and secure favorable financing terms, as investors and lenders increasingly prioritize sustainability and ethical considerations in their decision-making.

The impact of greenwashing, a term related to providing investors or the public with false information regarding the environmental impact of a company's product or operation (Hayes, 2024), is a significant limitation associated with ESG scores. Companies may misrepresent or exaggerate their efforts to become more sustainable. Self-reporting ESG scores may be less reliable than an actual reporting agency. According to a Forbes report by Benjamin Laker, an analysis conducted by Scientific Beta suggests, "ESG ratings present a paradoxical nature, measuring relative progress rather than absolute impacts." Due to this self-reporting, the company's efforts to be sustainable may have no impact and not factor into their financials. To uncover if my hypothesis is true, I performed a regression analysis on ESG scores to see if there is a correlation between a company's ESG score and profit margins.

3. Methodology and Data

Method

The variables in this research project include ESG scores as the independent variable and financial metrics, specifically profit margins, as the dependent variable. The companies used in this research study are all information technology sector companies in the S&P 500. All data used in this research was recorded from Yahoo Finance and each company's financial statistics. Each company was assigned an overall ESG score, as determined by Sustainalytics, encapsulating environmental, social, and governance metrics. Sustainalytics calculates ESG scores as an inverse of ESG risk. ESG risk is determined by a company's exposure to industry-specific material ESG risks and how well a company is managing those risks. All ESG scores for each company are made publicly available on Yahoo Finance, which is where I collected them as variables to use in the regression. Additionally, a specific governance, social, and environmental score was provided for each respective company. Each company's profit margin was recorded to analyze the company's growth.

Each company contains a sustainability section, which includes its environmental, social, and governance (ESG) risk ratings, with an overall total ESG risk score and a score for each section. The ESG Risk Rating quantifies a company's significant ESG risks and their management of those risks. The underlying source of the data displayed by Yahoo Finance is from Sustainalytics, a Morningstar Company. The profit margin is based on the financial statements of companies' SEC filings aggregated by Morningstar and is the source of the information shown in Yahoo Finance. The profit margin is calculated by dividing the gross profit by the year's total revenue and multiplying by 100.

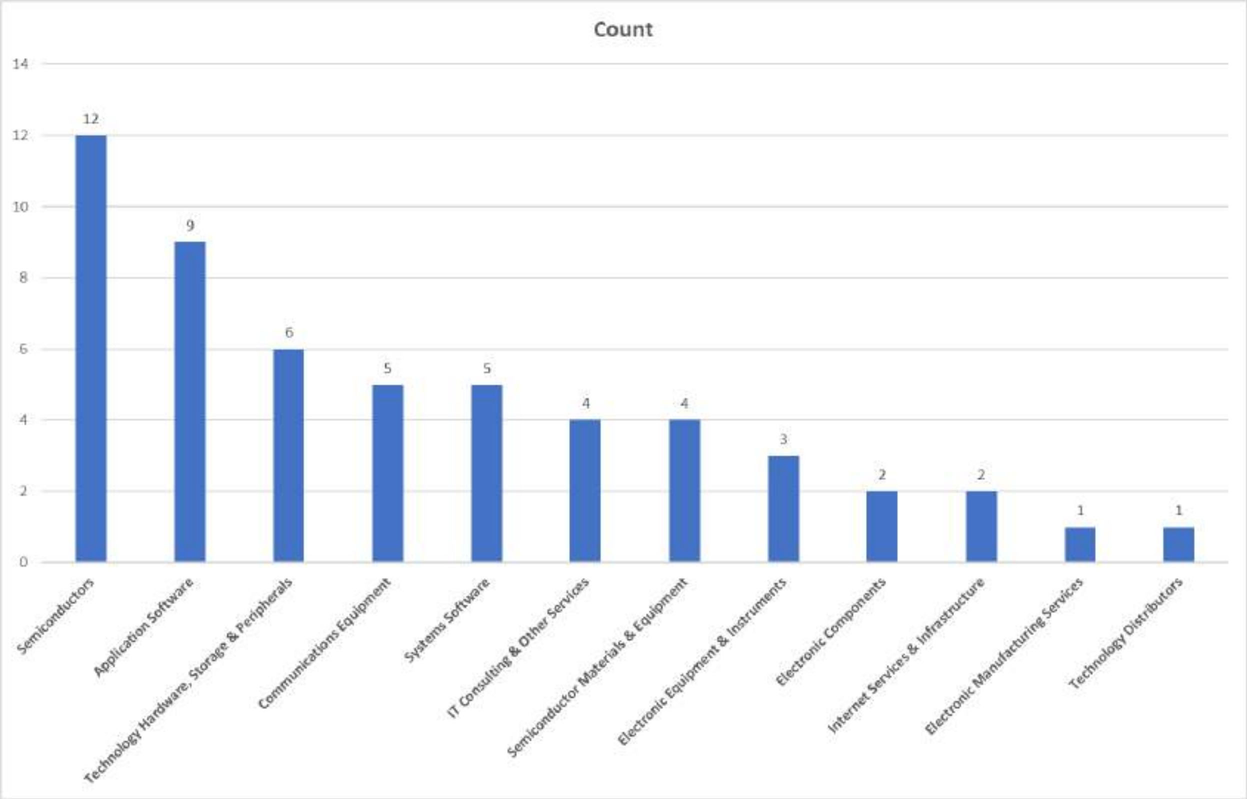


Figure 1. The bar chart shows the total number of companies involved in this research project. Information tech companies are grouped by their specific industry. Semiconductors have the largest representation, with 12 companies, while only one company represents the Electronic Manufacturing Services industry, and one represents the Technology Distributors industry.

Data Analytic Plan

Regression analysis was performed to analyze and discover relationships between ESG score components and profitability. Regression was conducted on the overall ESG score and each environmental, governance, and social section. Regression analysis is a set of statistical methods to estimate relationships between a dependent variable and one or more independent variables. It is described by the equation Y = a + bX + ϵ, where y is the dependent variable, a is the y-intercept, b is the coefficient (slope), x is the independent variable, and ϵ represents the residual error. In this research, regression analysis was used to see if there is a correlation between a company's ESG score and respective profit margin.

4. Results

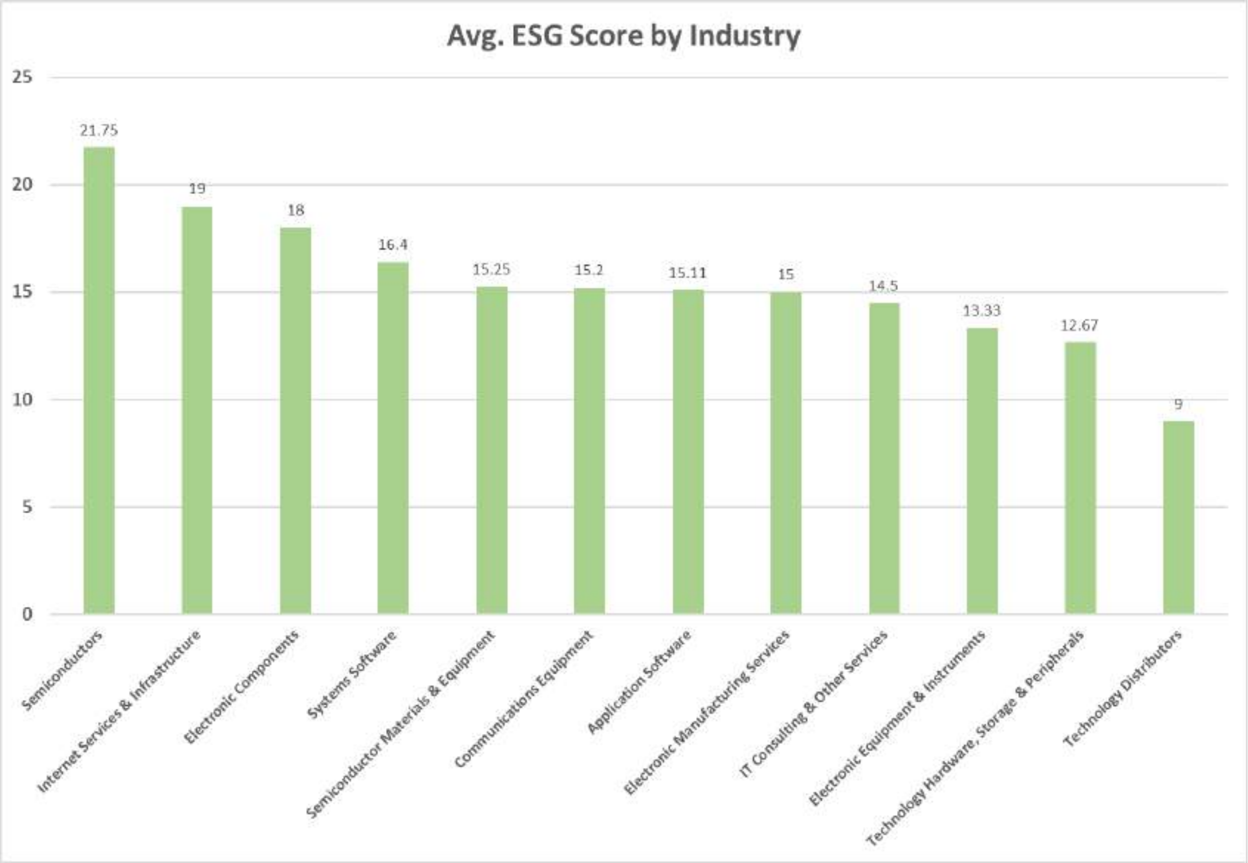


Figure 2. This chart illustrates the average ESG score separated by tech Industry. Semiconductors, including companies such as NVIDIA, AMD, and Intel, produced an average ESG score of 21.75. However, technology distributors had an average ESG score of only 9. This is a representation of each company's overall composite ESG score.

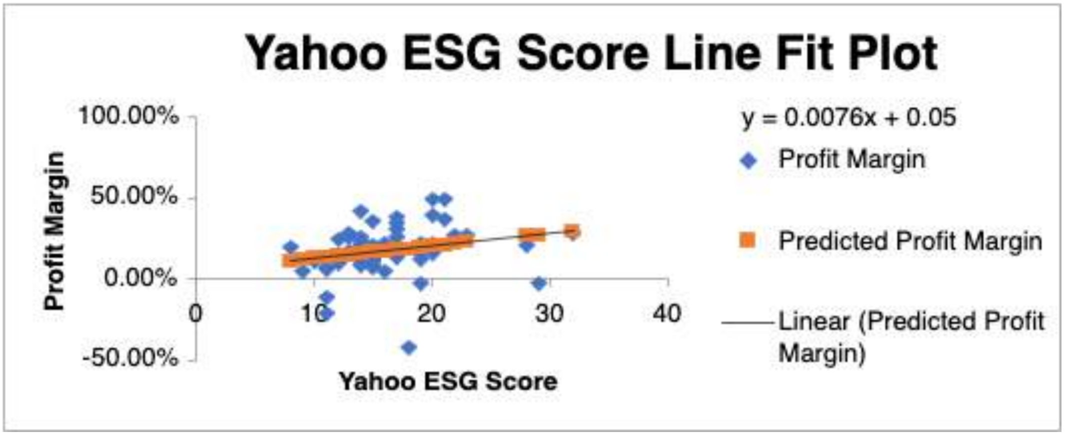


Figure 3 shows the regression analysis graph of ESG score and profit margin. The line indicates a positive correlation with a coefficient of 0.0076. This means that for every increase in ESG score, my data shows that a company's profit margin will increase by 0.76%.

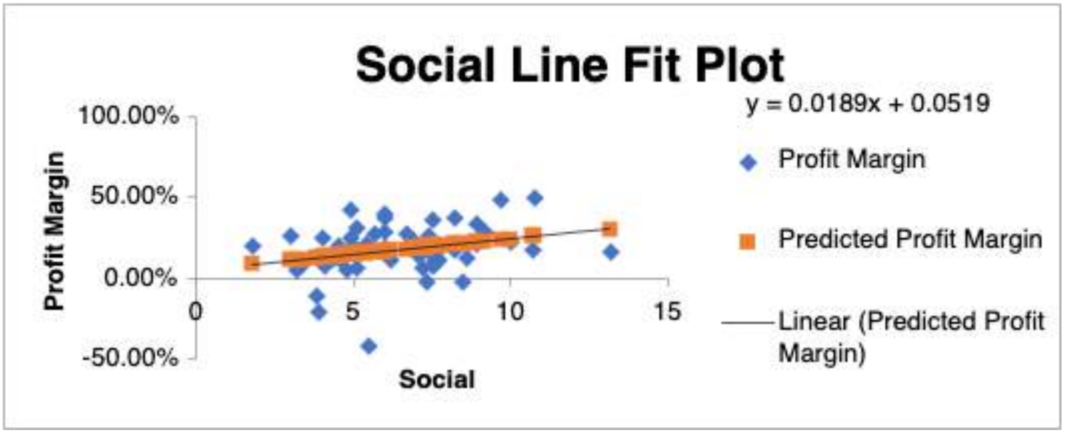


Figure 4 shows the regression analysis graph of social score and profit margin. The line indicates a positive correlation with a coefficient of 0.0189. This means for every increase in social score, the profit margin increases by 1.8%.

Regression analysis was performed to analyze the data, showing the correlation between ESG scores and profit margin and their impact on company success. R^2captures what percent of the movement in the independent variable is described by the dependent variable. For the overall ESG score, the R^2 is 0.049, meaning ESG scores can describe 4.9% of the movement in profit margins. The coefficient for the overall ESG score is 0.007586. This means that for every increase in ESG score, a company's profit margin will increase by 0.7586%. When examining each category, we can see which influences profitability most. The p-value for environmental is > 0.05, meaning that it fails to reject the null hypothesis at a significance level of α=0.05 and that there is no real correlation between the score and profit margin. Additionally, the p-value for governance is > 0.05, indicating that it fails to reject the null hypothesis at a significance level of $\alpha=0.05$. However, the p-value for social is at 0.05, a confidence of 95%, indicating that there is a correlation between a higher social score and increased profit margins. The coefficient is 0.018. This means that for every single social ESG score, a company's profit margin increases by 1.8%.

Furthermore, this implies that if a company can increase its social score by one point from a five to a six, it can increase its profit margin by 2%. For example, a company such as Advanced Micro Devices (AMD) with revenue of $22.7 billion can increase profits by approximately $450 million with a one-point increase in their social score while keeping expenses constant. Additionally, the R^2 value is 0.07, indicating that the social ESG score can describe 7% of all movement in profit margins. The data shows that the ESG score with the greatest impact on a company's profit margin is Social.

5. Discussion

The main findings from my data show that ESG score is positively associated with firm profit margin. The regression shows that there is a correlation between a company's social score and firm profitability. The social score had the largest impact on firms' profitability, with a coefficient of 0.0189. Additionally, social has a higher R^2 value than the other two sections combined, indicating that social can explain more of the movement in the dependent variable than the others. Environmental scores had the lowest impact on tech companies' profit margins and could explain the least amount of movement of the dependent variable, approximately seven basis points.

This research is essential for the investing community because it provides a new tool by which they can measure a company and help forecast future potential. Additionally, this research is key for company management, specifically technology executives, because it provides a more modern view of factors that should be considered in running the business as it relates to the firm's contributions to society. This research highlights that the two noble goals of increasing profits and creating a more sustainable and just society are not in conflict with one another and are symbiotic.

The limitation associated with my data is that the ESG scores are only provided to public companies; therefore, this data was centralized specifically for public tech companies. The companies in the technology sector are in different industries and focus on different metrics. Additionally, the time variable is a major limitation as my data is only measured from one period of time. This restricts me from seeing exactly how ESG scores have impacted profit margins over time. Furthermore, I cannot determine how the economy performed during this period. There could be idiosyncratic events at certain companies that occurred over the past twelve months, skewing their profit margin away from longer-term averages.

Further research should be to gather more data across multiple sectors and compare the impact of ESG on the company's financials. Additionally, performing multiple regressions and analyzing a company's ESG score over time could provide more accurate evidence of ESG scores' impact on profit margin. Other possible research advancements would be to quantify the degree by which different companies are investing into ESG initiatives. Another research study would be to investigate the specific ESG metrics that drive the observed correlation with profitability. Also, one could analyze which social factors, in particular, have the most significant impact on profit margins. Understanding the specific components of ESG that drive financial performance can inform targeted ESG strategies for companies.

6. Conclusion

This paper aimed to address the research question: How does a higher ESG score impact technology firms' profitability? I hypothesized that increasing ESG scores would benefit companies by increasing profit margins. The data shows that there is no correlation between overall ESG score and firm profitability. There is a positive association that an increase in ESG score may yield an increase in profit margin. The category that most impacted profit margin was social, with the largest coefficient and R^2 value. This means that it can describe more movement in profit margin than the other sections. Surprisingly, environmental scores did not have a marginal impact on the firm's profitability even though they are considered the main component of a sustainable future. This could partially be because the study concentrated on only technology stocks, which don't factor in as many environmental components in their operations. When looking specifically at social scores, there is statistically significant evidence of a positive correlation between an increase in social score and increased profit margin. This work may motivate business leaders to prioritize positive social change at their companies.

Work Cited

Admin, ESGRI. “The Complete Guide to ESG Scoring.” *ESG Reporting Intelligence*, 21 June 2023, esgri.com/the-complete-guide-to-esg-scoring/.

Aydoğmuş, Mahmut, and AbstractIn this study. “Impact of ESG Performance on Firm Value and Profitability.” *Borsa Istanbul Review*, Elsevier, 17 Nov. 2022, www.sciencedirect.com/science/article/pii/S221484502200103X.

Doherty, Rebecca, et al. “The Triple Play: Growth, Profit, and Sustainability.” *McKinsey & Company*, McKinsey & Company, 9 Aug. 2023, www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/the-triple-play-growth-profit-and-sustainability.

The ESG Risk Rating: connect.sustainalytics.com/hubfs/SFS/Sustainalytics ESG Risk Rating - FAQs for Corporations.pdf. Accessed 4 Apr. 2024.

“ESG Risk Ratings.” *Sustainalytics.Com*, www.sustainalytics.com/corporate-solutions/esg-solutions/esg-risk-ratings. Accessed 3 Apr. 2024.

Hayes, Adam. “How to Tell If a Company Has High ESG Scores.” *Investopedia*, Investopedia, 2023, www.investopedia.com/company-esg-score-7480372.

Hayes, Adam. “What Is Greenwashing? How It Works, Examples, and Statistics.” *Investopedia*, 22 Jan. 2024, www.investopedia.com/terms/g/greenwashing.asp#:~:text=Greenwashed%20products%20might%20convey%20the,energy%20or%20pollution%20reduction%20efforts.https://www.investopedia.com/terms/g/greenwashing.asp#:~:text=Greenwashed%20products%20might%20convey%20the,energy%20or%20pollution%20reduction%20efforts.

Krantz, Tom. “The History of ESG: A Journey towards Sustainable Investing.” *IBM Blog*, 8 Feb. 2024, www.ibm.com/blog/environmental-social-and-governance-history/.

Laker, Benjamin. “Greenwashing Unmasked: A Critical Examination of ESG Ratings and Actual Environmental Footprint.” *Forbes*, Forbes Magazine, 7 Aug. 2023, www.forbes.com/sites/benjaminlaker/2023/08/04/navigating-the-mirage-unraveling-the-disconnect-between-esg-ratings-and-real-environmental-impact/?sh=631f7451f8b7.

Martins, Andrew. “Most Consumers Want Sustainable Products and Packaging.” *Business News Daily*, 28 Mar. 2024, www.businessnewsdaily.com/15087-consumers-want-sustainable-products.html.

Shifflett, Shane. *Wall Street’s ESG Craze Is Fading - WSJ*, 19 Nov. 2023, www.wsj.com/finance/investing/esg-branding-wall-street-0a487105.

“What Is ESG Investing and Analysis?” *CFA Institute*, www.cfainstitute.org/en/rpc-overview/esg-investing#:~:text=ESG%20stands%20for%20Environmental%2C%20Social,material%20risks%20and%20growth%20opportunities. Accessed 3 Apr. 2024.

Addressing Traditional Finance Fund Limitations with Blockchain-Based Open-Ended Funds By Kartavchenko Mikhail, Pismenskaya Sofia, Kartavchenko Alexander, Svirskii Nikita, Illarionova Evelina, Kolotushkina Lyubov, Krivonosov Semen

Abstract

This research explores the potential of blockchain-based on-chain open-ended funds to resolve issues inherent in traditional finance funds, such as lack of transparency, restricted accessibility, and operational inefficiencies. Implemented as a student-managed investment project, this study evaluates the practical application of such funds. Findings indicate significant benefits, including enhanced transparency, broader accessibility, and efficient management through Uniswap's interface. Challenges like complex setup and security remain. These insights highlight blockchain technology's transformative potential in modern financial management, validated through our real-life student-managed fund experiment.

Introduction

Traditional investment funds have long been a staple in wealth management, offering professional management and potential economies of scale. However, they suffer from notable drawbacks, including opacity in operations, limited accessibility for smaller investors, and operational inefficiencies. For instance, investors often find it challenging to access real-time data about their investments, and high fees can erode returns. Blockchain technology, with its decentralized, transparent, and secure nature, promises to mitigate these issues. Specifically, on-chain open-ended funds can address these limitations by leveraging smart contracts, which automate fund operations and ensure transparency.

On-chain open-ended funds operate through smart contracts on the blockchain. Investors contribute cryptocurrency to the fund’s smart contract, which mints fund tokens based on the current token price. This price is dynamically calculated from the fund's portfolio value. Tokens can be redeemed for a proportional share of the fund's assets, with management and profit fees deducted as specified. This setup promises enhanced transparency and accessibility but also introduces challenges, such as complex setup and potential security risks. This research examines the viability of blockchain-based on-chain open-ended funds compared to traditional finance funds. By implementing a student-managed fund as a practical example, we demonstrate the benefits and challenges of this innovative financial instrument.

Methods

Participants: The on-chain open-ended fund was managed by high-school students who actively participated in its setup and management decisions.

Study Design: This study focuses on the functionality of the on-chain open-ended fund. Comparative analysis of functionalities, advantages, and limitations was based on real-world implementation and management experiences.

Materials: The on-chain fund utilized the Polygon blockchain and smart contract technology for contributions, token minting, and fund operations. Additionally, funds were managed directly through the Uniswap interface, enabling seamless integration and liquidity management.
Procedure: The research involved setting up the on-chain fund, collecting data on its operation, and analyzing key metrics such as transparency, accessibility, and operational efficiency. Practical challenges during setup and management were documented to provide insights into the fund's functionality.
Fund: We deployed our smart contract on the Polygon blockchain due to its low transaction fees. Investors could contribute USDT by interacting with the fund's smart contract, which automatically minted fund tokens based on the current price. A student managing board from supporting schools made fund management decisions. The Uniswap Interface was utilized for management, enabling secure cryptocurrency trading directly from the fund's smart contract, making it virtually impossible to steal funds. Withdrawals were processed directly through the smart contract by submitting a request, fulfilled after 24 hours based on the current token price.

Results

Transparency: The on-chain fund significantly improved transparency through blockchain-based operations, recording all transactions and fund movements on the blockchain for real-time visibility. Traditional finance funds often lack this level of transparency, making it difficult for investors to track their investments in real-time.
Accessibility: The on-chain fund allowed unrestricted entry, enabling investments regardless of citizenship, income, or location, unlike traditional funds that often impose specific criteria on investors. Traditional funds typically have high minimum investment thresholds, excluding smaller investors.
Efficiency: The on-chain fund exhibited operational efficiency through automated processes managed by smart contracts. Using the Uniswap interface, we managed liquidity and trading directly from the smart contract. Traditional funds, on the other hand, suffer from inefficiencies due to manual processes, delayed transaction settlements, and higher administrative costs.
Limitations included the complexity of creating and managing smart contracts and their vulnerability to hacking due to poor design. Additionally, integrating bridges for cross-blockchain functionality and incorporating other DeFi instruments like lending and staking further increased the complexity of the setup, necessitating multiple interfaces to be connected to the smart contract. Migrating to updated smart contracts is cumbersome, requiring careful handling of existing investor tokens.

Discussion

This study reveals that on-chain open-ended funds offer substantial advantages in transparency and accessibility compared to traditional finance funds. Blockchain-based operations ensure visibility and immutability of transactions, enhancing investor trust. Additionally, the lack of entry barriers democratizes investment opportunities for a broader range of investors.

Significant limitations exist. Setting up a secure and adaptable smart contract is complex, requiring detailed planning and expertise. Poorly designed smart contracts are vulnerable to hacking, undermining blockchain's security benefits. Managing a diverse portfolio within a smart contract presents operational challenges due to the need for integration with numerous DeFi interfaces. Additionally, accessing traditional financial instruments from the blockchain remains problematic, further complicating fund management.

Conclusion

Blockchain-based on-chain open-ended funds present a promising alternative to traditional finance funds, particularly in terms of transparency and accessibility. Despite notable challenges, future research should focus on developing robust and adaptable smart contract frameworks and exploring solutions for integrating traditional financial assets into blockchain systems. This study underscores the potential for blockchain technology to transform financial management, provided the associated complexities and risks are effectively managed.

References

- Buterin, V. (2014). A Next-Generation Smart Contract and Decentralized Application Platform.
- Gomber, P., Kauffman, R. J., Parker, C., & Weber, B. W. (2018). On the Fintech Revolution: Interpreting the Forces of Innovation, Disruption, and Transformation in Financial Services.
- Christidis, K., & Devetsikiotis, M. (2016). Blockchains and Smart Contracts for the Internet of Things.
- Li, X., Jiang, P., Chen, T., Luo, X., & Wen, Q. (2020). A Survey on the Security of Blockchain Systems.

Acknowledgments

We would like to express our sincere gratitude to the faculty and staff of Governor's Physics and Mathematics Lyceum No. 30 for their invaluable support and guidance throughout this project. Special thanks to the student managing board for their dedication and hard work in managing the fund.

Additionally, we extend our heartfelt thanks to Capestone for their support and collaboration in our research. Their contribution was instrumental in the success of this project. We would also like to thank Alexander Kartavchenko from TopTal LLC for his expert consultation on coding, which was crucial for the setup of our smart contract.

Finally, we extend our thanks to all participants who invested in the fund, contributing to the success and practical validation of our research. This project would not have been possible without the collective efforts and collaboration of everyone involved.

Student Management Board participants:

- Mikhail Kartavchenko - Governor's Physics and Mathematics Lyceum No. 30
- Popov Mikhail - Governor's Physics and Mathematics Lyceum No. 30
- Svirsky Nikita - Dar Essalam American School
- Krivonosov Semen - Presidential Physics and Mathematics Lyceum No. 239
- Kolotushkina Lyubov - Gymnasium No. 74
- Grigoriev Aleksey - Governor's Physics and Mathematics Lyceum No. 30

Exploring the relationship between sleep, lifestyle, and academic or work performance in medical trainees and healthcare professionals By Alec Alarakhia, Maggie Chivilicek, Nicole Ackerson, Nasir Alarakhia

ABSTRACT

Sleep quality plays a critical role in the well-being and cognitive functioning of individuals (Morris et al.; Marvaldi et al.; Mølgaard et al.; Machado-Duque et al.; Maheshwari and Shaukat; Scammell; Corrêa et al.; Gordon et al.). In the high-stress medical field, healthcare professionals and medical students often will not sleep an adequate amount (typically seven or more hours per night)(Watson et al.), whether due to difficulty finding the time to sleep an adequate amount or struggling to find good quality sleep(Morris et al.; Marvaldi et al.; Mølgaard et al.; Machado-Duque et al.; Maheshwari and Shaukat; Scammell; Corrêa et al.; Alhusseini et al.). This review paper summarizes the literature on the profound effect of poor sleep quality on medical students and healthcare professionals, shedding light on sleep levels' multifaceted consequences on their physical and mental health, job performance, and overall quality of life. Specifically, this paper analyzes the relationship between duration and quality of sleep and the quality of various aspects of the life of medical students and healthcare professionals, including social relationships, mental health, and academic/job performance. We examine how sleep affects the social relationships of those in the medical field, focusing on familial and romantic relationships, and we highlight the effect of sleep deprivation on mental health, including the impact of poor sleep on depression and anxiety rates in healthcare professionals and medical students. Research has found that the hours an individual spends sleeping have numerous effects on their quality of life (Marvaldi et al.; Machado-Duque et al.; Gordon et al.; Alhusseini et al.; Kecklund and Axelsson; Scott et al.; Coelho et al.; Papp et al.; Zhang et al.), research must be conducted to explore the importance of sleep further to understand better the relationship between sleep and various aspects of life. Those who work and study in the medical field are of particular interest in the discussion about sleep's effects on lifestyle because these individuals have been found to have the most demanding workweeks on average. As such, many medical students and healthcare workers are sleep-deprived, with 44.2% of healthcare workers being sleep-deprived (Shaik et al.), far greater than the nationwide average of 35% (Paprocki). Research conducted to explore sleep deprivation is especially interesting to healthcare professionals due to the heightened likelihood of a healthcare professional being sleep-deprived relative to other professions.

1. Introduction

Medical education and healthcare practice are both highly demanding fields in which success often costs one's well-being. To keep pace with the rigor and requirements of the medical field, medical students and healthcare professionals alike frequently sacrifice adequate sleep. With the intense curriculum of medical students and extreme work hours (including overnight shifts) of healthcare workers, establishing a lifestyle with sufficient time spent sleeping is difficult, so individuals studying and working in the medical field must be sure they have a healthy sleep pattern. Essential for a healthy lifestyle, adequate sleep is pivotal in various aspects of life. This paper delves principally into the available research on the impacts of sleep on mental health, academic/work performance, and personal life among medical students and healthcare professionals. In doing so, we offer insights into how significant sleep is in the lives of those working in and studying medicine and examine the potential consequences of failing to achieve adequate sleep.

This literature review examines published research on this subject to elucidate the importance of advocating for interventions and strategies to improve sleep quality among medical students and healthcare professionals. By synthesizing both perspectives and empirical evidence, this paper provides a comprehensive review of the significance of sleep in mental health, academic/work performance, and personal life. The research reviewed indicates the strong correlative relationship between sleep and these aspects of life, demonstrating the importance of optimizing the sleep patterns and habits of those working and studying in the medical field.

2. Background Information

2.1 Sleep and Circadian Rhythms

The sleep-wake is typically divided into stages: wakeness, rapid eye movement sleep (REM), and non-rapid eye movement sleep (NREM) (Scammell). When one first falls asleep, NREM begins. This early stage is followed by REM and NREM sleep cycles, typically in 90-minute periods throughout the night (Scammell). NREM sleep can also be divided into three phases: N1, N2, and N3 (See Figure 1). N1 is the lightest stage of NREM sleep, from which people arise quickly (Scammell). On the other hand, N3 is the stage in which it is hardest to wake an individual (Scammell).

During REM sleep, the cortex is active, and it is harder to awake, generating vivid thoughts and dreams (Scammell). Across the night, episodes of REM sleep become longer (Scammell). REM sleep has been found to play a critical role in emotional processing and cognitive development, and adults need at least two hours of REM sleep per night (See Figure 1)("REM Sleep Revealed"). As a result, adults must get adequate, uninterrupted sleep to achieve longer episodes of REM sleep. NREM sleep also plays a vital role, being the stage of sleep in which the body repairs tissue, builds bone and muscle, and strengthens its immune system (See Figure 1)(*Definition of NREM Sleep - NCI Dictionary of Cancer Terms - NCI*).

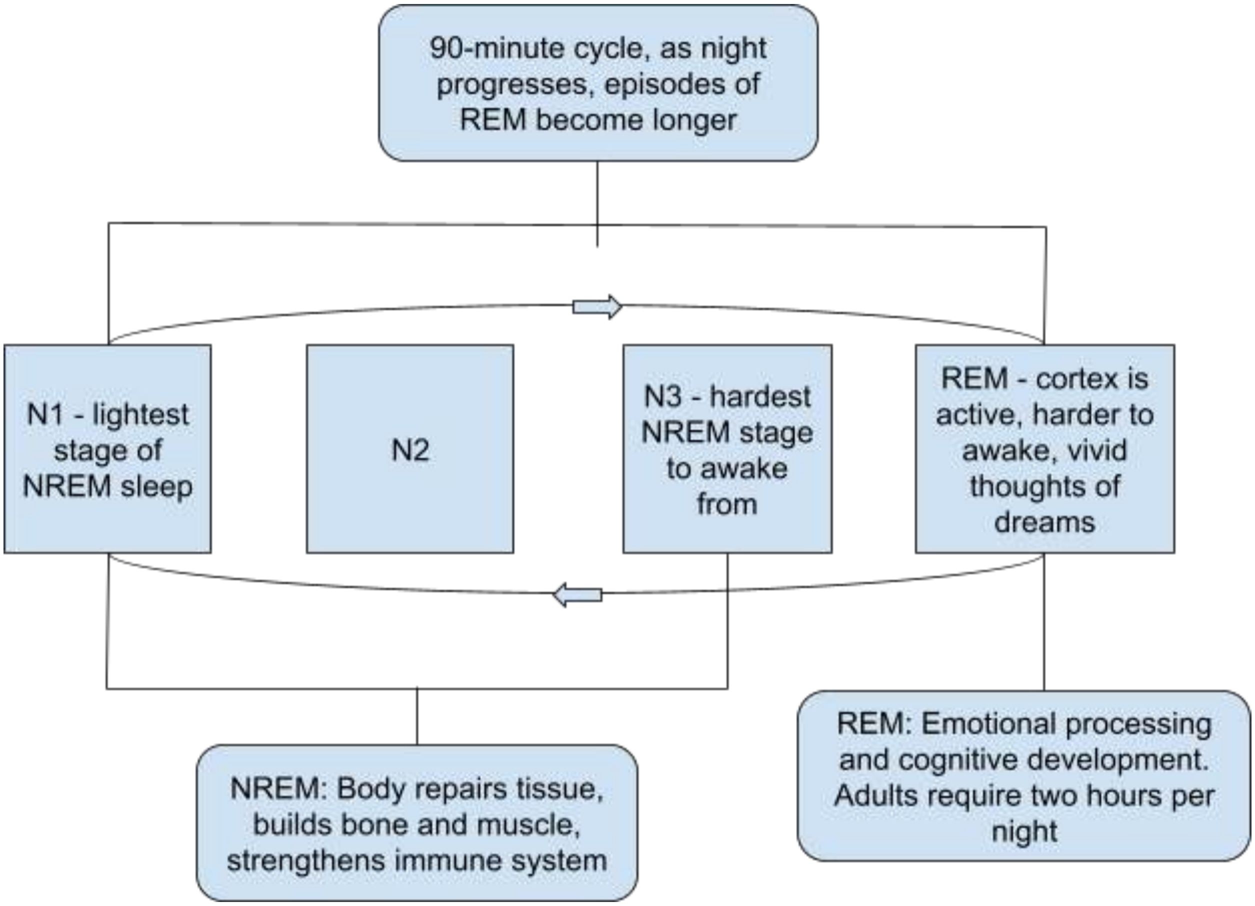

90-minute cycle, as night progresses, episodes of REM become longer
N1 - lightest stage of NREM sleep
N2
N3 - hardest NREM stage to awake from
REM - cortex is active, harder to awake, vivid thoughts of dreams
NREM: Body repairs tissue, builds bone and muscle, strengthens immune system
REM: Emotional processing and cognitive development. Adults require two hours per night

Fig. 1 NREM-REM cycle with stages. The figure serves as a comprehensive model to understand the cyclical nature of the sleep cycle.

The homeostatic sleep drive is a pressure imposed by an individual's body that builds up in their body as the time spent awake increases (Scammell; *Module 2. Sleep Pressure*). Homeostatic sleep drive results from rising adenosine levels, in which the excess adenosine that remains slows the brain processes for awakeness and turns on brain processes for sleepiness (Figure 2). The sleep drive increases during time spent awake, and a "sleep debt" is created when an individual does not obtain adequate sleep (*Module 2. Sleep Pressure*). As a result, medical students and healthcare professionals who accumulate a "sleep debt" will often feel tired due to inadequate sleep, potentially impacting aspects of life such as mental health and school/work performance. Furthermore, as stated earlier, sleeping consistently and regularly is much more valuable than sleeping episodically (*Module 2. Sleep Pressure*). As a result, healthcare workers who work long, interrupted shifts may struggle to maintain a steady schedule, and their sleep drive may be increased.

Circadian rhythms initiate energy storage for metabolic processes and aid synaptic function, memory consolidation, and complex motor system control (Reddy et al.). They are a 24-hour cycle that regulates our cycles of alertness and sleepiness (Reddy et al.). Disruptions and misalignments to the circadian rhythm can impact several systems, including the immune, gastrointestinal, reproductive, endocrine, renal, skeletal, and cardiovascular systems (Morris et al.; Reddy et al.). Circadian rhythms control several identified clock genes, including BMAL1/BMAL2, CLOCK, CRY1/CRY2, and PER1/PER2/PER3 (Figure 2). These genes influence signaling pathways that allow cells to identify the time of day. While the homeostatic sleep drive controls our urge to sleep, circadian rhythms control how alert we remain during the day (Figure 2). To undo sleep debt accumulated by the homeostatic sleep drive, one must sleep for slightly longer nights; however, a adverse effects may arise if an individual oversleeps one night (Haden). Sleeping in too late can throw off an individual's sleep schedule and disrupt their circadian rhythm, making it difficult to fall asleep at bedtime (Haden).

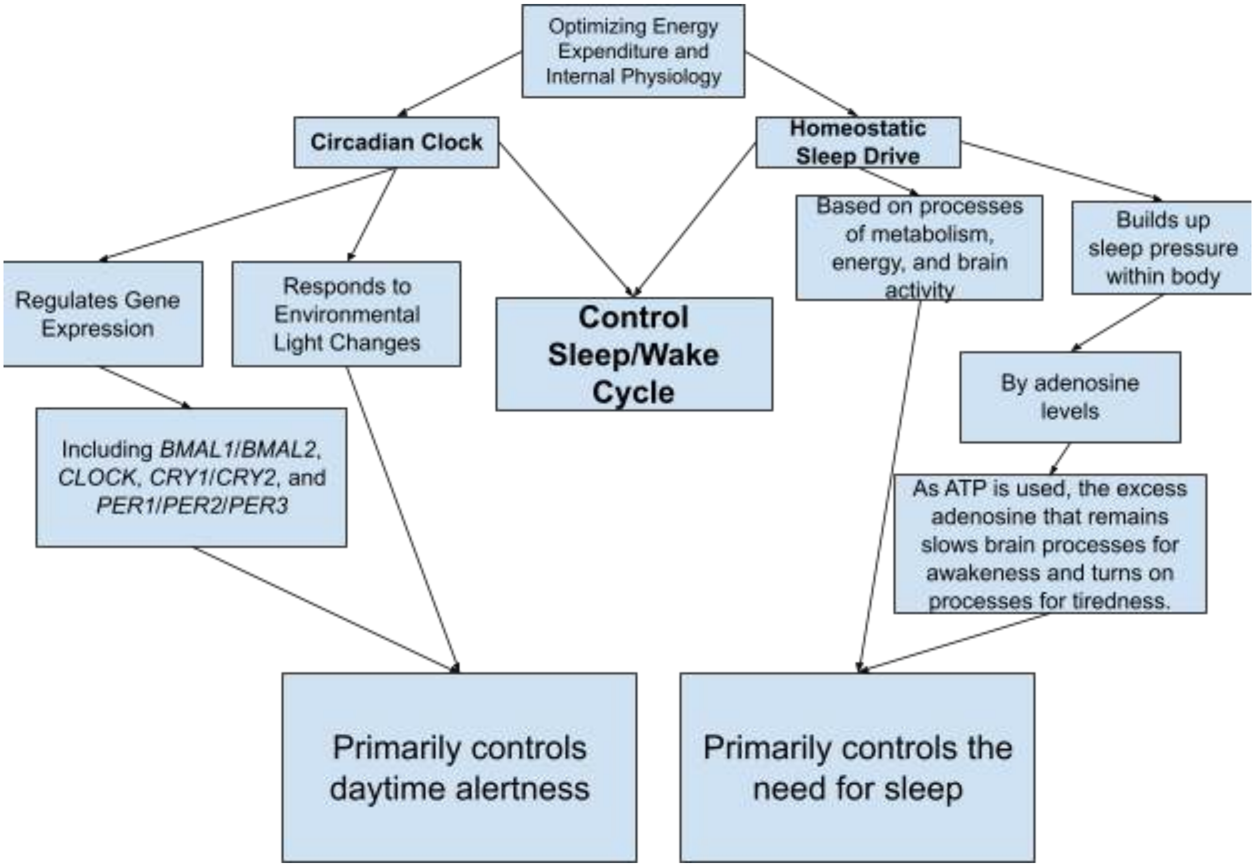

Optimizing Energy Expenditure and Internal Physiology
Circadian Clock
Homeostatic Sleep Drive
Regulates Gene Expression
Responds to Environmental Light Changes
Control Sleep/Wake Cycle
Based on processes of metabolism, energy, and brain activity
Builds up sleep pressure within body
By adenosine levels
Including *BMAL1/BMAL2*, *CLOCK*, *CRY1/CRY2*, and *PER1/PER2/PER3*
As ATP is used, the excess adenosine that remains slows brain processes for awakeness and turns on processes for tiredness.
Primarily controls daytime alertness
Primarily controls the need for sleep

Fig. 2 Comprehensive Sleep/Wake Cycle Model - Authors' Conception. The figure highlights the interplay between the circadian rhythm (clock) and the homeostatic sleep drive, serving as a visual model to conceptualize the two-process sleep model.

2.2 Work Demands of Healthcare Professionals and Medical Trainees

Most physicians work a 40- to 60-hour workweek, with more than a quarter working over 60 hours (*Average Physician Workweek*). This average is significantly higher than the average American adult, who works 38.7 hours per week ("What Is The Average Work Hours Per Week In The US?"). In 2018, the average male physician workweek was 51.89 hours, while female physicians worked 50.46 hours per week on average in 2018 (*Average Physician Workweek*). The average male workload in the US is 40.5 hours, while the female workload is 36.6 hours on average ("What Is The Average Work Hours Per Week In The US?"). This drastic disparity between physician workload and the average American workload (See Figure 3) is primarily caused by the limited number of physicians and the immense demand for patient care (*Average Physician Workweek*).

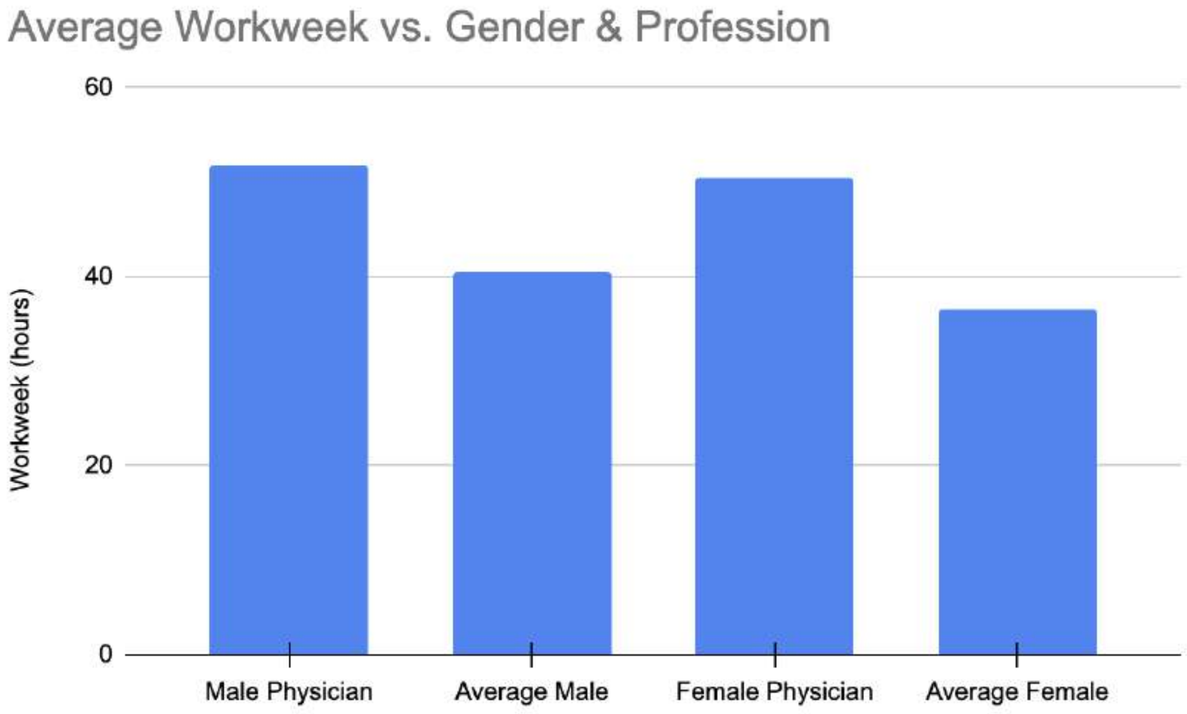


Fig. 3 This figure compares the average workweek of male and female physicians to that of the average adult male and female, demonstrating the lengthy workweek of physicians relative to those working in other professions.

Similar to physicians, medical students have a more challenging and time-demanding workload than the average post-graduate student ("College vs. Medical School Comparison"; Chang; "Student Study Time Matters - CollegiateParent"). Medical students study 6-12 hours per day on average outside of the classroom, whereas the typical law student spends just 4-6 hours per day studying outside of the classroom (Chang; "Student Study Time Matters - CollegiateParent"; Dempsey-Klott). Aside from studying, medical students must spend extra hours in labs and hospitals. As a result, their work continues even outside of the classroom. Furthermore, medical schools are primarily pass-fail. As such, medical students are at a high risk of failing if they lose concentration in school. The high attrition rate of 15.7% to 18.4% for medical students in four-year programs reflects the increased likelihood of failure (Inman). While a student in another type of program could recover their GPA if they lose focus in a class or for a period of time, medical students lack this opportunity, meaning that more students fail to complete medical school than in another program.

Medical students also must pass required, rigorous tests known as USMLE Step 1 and USMLE Step 2 (MD). Step 1 challenges test-takers with a broad range of science knowledge questions over 8 hours (MD). Step 2 hones in further on content directly related to practicing medicine in a 9-hour examination (MD). Medical trainees are required to complete these exams on the path to becoming medical professionals (MD). Few professions have examinations that compare to the duration, difficulty, and preparation required for Steps 1 and 2. This poses yet another challenge for aspiring medical students.

3. Mental Health and Sleep Quality

3.1 Medical Trainees

Sleep quality affects the mental health of those working and studying in the medical field. Research has found that there is a direct correlation between the sleepiness of medical students and their mental health (Alhusseini et al.; Alotaibi et al.; Perotta et al.). For example, in a study of 1350 randomly selected medical students, those who reported higher levels of sleepiness were found to have greater odds of depression and anxiety while in medical school (Perotta et al.). This study used the sleep deprivation index (SDI), which is defined as the difference between mean hours of sleep during weekends and on weekdays. This index can be used to compare the extra hours that an individual sleeps during the school week. An SDI greater than four, for example, would indicate that the student is not sleeping sufficiently on the weekdays. In the questionnaire, it was found that higher SDI groups reported a lower quality of life (Perotta et al.). This survey included questions regarding learning, teachers, educational atmosphere, and overall academic and social self-perception. A lower reported quality of life indicates lower mental health since their perception of their life is worse. It was also found that students with four hours or more of SDI had significantly higher odds of experiencing and reporting depression symptoms in comparison with students with an SDI of less than three (See Figure 5)(Perotta et al.).

Similarly, it was found that as a student's SDI increased, so did their odds ratio (the strength of association between an exposure and an outcome)(Szumilas) for increased anxiety symptoms in comparison with those with a lower SDI. This finding was accurate for both state and trait anxiety (Perotta et al.). State anxiety is defined as anxiety caused by adverse conditions in a particular moment (one's present "state"). In contrast, trait anxiety is anxiety in the form of a trait of personality (Leal et al.). For example, an individual with state anxiety may experience anxiety when they are late for work, but once they arrive on time, they are no longer anxious (Leal et al.; *What Is Trait Anxiety?*). An individual with trait anxiety, though, may experience anxiety regarding being late to work even when there is no reason to do so (Leal et al.; *What Is Trait Anxiety?*). It is likely that depression and sleep are correlated, as it is plausible that depressed people do not sleep as well, and sleep-deprived people feel more depressed. Research has also found that medical students' psychological stress levels are significantly increased when they report poor sleep quality (See Figure 5)(Alhusseini et al.; Perotta et al.). A study conducted at the College of Medicine Imam Muhammad Ibn Saud Islamic University (IMSIU) in Riyadh, Saudi Arabia, used the Kessler Psychological Distress Scale (K10) to demonstrate this (Alotaibi et al.). The K10 contains ten questions that investigate an individual's anxiety and depressive symptoms during the past month (Alotaibi et al.). The study reported poor sleep quality in a majority of medical students, with a prevalence of 77% among participants in the survey (See Study 1). It also found that almost two-thirds of medical students (63.5%) exhibited distress, as defined by K10 (See Figure 5)(Alotaibi et al.). A greater proportion of participants who had poor sleep quality reported distress compared to participants who did not report poor sleep quality.

Interestingly, an increase in distress level is also related to poorer sleep quality, as reported by the medical students included in the study (Alotaibi et al.). Another study suggests that medical students who do not obtain 8-10 hours of sleep per night are more likely to experience stress than medical students who do (Tran et al.). A similar correlation between psychological stress and sleep could be found, as those who are more stressed may not sleep as well, and those who do not sleep as well may be more stressed. Several metrics, including the rate of anxiety, depression, and stress, indicate the negative effect of poor sleep on mental health.

3.2 Healthcare Professionals

Healthcare professionals also experience effects on mental health as a result of poor sleep quality. Similar findings held true regarding the impact of poor sleep quality on the mental health of healthcare professionals (Marvaldi et al.; Coelho et al.; Zhang et al.; Ganesan et al.). One study found that healthcare workers have a higher incidence of depression than the general population. It was clear that sleep levels played a factor in depression levels among healthcare workers, as those who had worse sleep quality (due to sleep duration of six or fewer hours or the presence of sleep disturbances) were more likely to show depression symptoms (See Figure 5)(Zhang et al.). In another study encompassing 4,971 healthcare workers (nurses, nursing assistants, and physicians) from both public and private healthcare facilities in France, poor sleep, burnout, and depression were all prevalent. Of the individuals surveyed, 64.5% reported poor sleep, 56.5% experienced burnout (defined as a form of exhaustion after excessive emotional, physical, and mental fatigue), and 29.8% reported depression (See Figure 5)(Coelho et al.). This depression rate among healthcare workers was more than double the 13.3% rate of adults in the general population who suffer at least one depressive episode, a worrying discovery ("One in Five Young French People Has a Depressive Disorder").

Perhaps more reason for concern is the fact that these rates can be significantly elevated in health crises; for example, it was found that anxiety and depression levels were elevated among healthcare workers during the COVID-19 pandemic (Marvaldi et al.). A meta-analysis that extrapolated data from a variety of countries and different studies on physicians and nurses working during the height of the pandemic supported this finding. Nurses were included in this study because they are another healthcare professional with a similar workweek, though studying for nursing school is less compared to medical school for physicians. Across a total of 51,942 participants, the pooled prevalence of anxiety was found to be 30.0% (See Figure 5)(Marvaldi et al.). The pooled prevalence of depression and depression symptoms was 31.1% for a different analysis of 68,030 participants (See Figure 5)(Marvaldi et al.; Stewart et al.; Weaver et al.).

These findings are both well above average rates of anxiety and depression for healthcare workers, as evidenced by another study that reported 21.6% rates for depression and anxiety among healthcare workers who were not tested during the pandemic (Weaver et al.). Another point of intrigue was the increase in the prevalence of sleep disorders associated with COVID-19: a prevalence of 44.0% existed in a combined total of 12,428 participants (Marvaldi et al.). This rate is higher than the 40.9% prevalence reported in a separate study that did not include healthcare workers employed during the COVID-19 pandemic (Weaver et al.). Anxiety, depression, and sleep disorders were likely more common among healthcare workers during the pandemic because of the added stress and panic that surrounded the global health crisis. In another study, 56.3% of the frontline healthcare workers tested during the COVID-19 pandemic reported burnout (Alotaibi et al.).

Sleep disturbances and nightmares are other results that indicate the impact of sleep deprivation on mental health (Oh et al.). One study reported findings of 6.1 hours mean sleep duration, and 60.9% of healthcare workers tested reported sleep disruptions. Meanwhile, 45.2% of the individuals included in the study reported nightmares at least once per week (Stewart et al.). This is an important finding because research has found a strong correlation between the frequency of sleep disorders in a group of tested individuals and the rates of depression and anxiety (Oh et al.). Since sleep disorders can indicate worse mental health(Oh et al.), it is a further indicator to demonstrate the reduced mental health of healthcare professionals as a result of sleep deprivation.

4. School/Work Performance and Sleep Quality

4.1 Medical Trainees

Poor sleep quality has also been found to have a detrimental effect on medical students' academic performance(Machado-Duque et al.; Maheshwari and Shaukat; Alhusseini et al.; Alotaibi et al.; Jalali et al.). A clear indication of a correlation between sleep and medical students' school results is the stark difference in grade point average (GPA). In a study of almost 800 students – with nearly two-thirds having poor sleep quality – those who were not poor sleepers were found to have significantly higher GPAs (weighted mean of 3.31) than their poor-sleeping counterparts (weighted mean of 2.92)(Maheshwari and Shaukat). Another study found that 36.4% of students with poor sleep quality had a GPA less than or equal to 3.49, while only 27.7% of those tested who did not have poor sleep quality had a GPA less than or equal to the same figure(Alotaibi et al.). This was a significant difference found as a result of a difference in sleep quality.

While sleep quality can be challenging to measure – primarily because it cannot be determined from one metric – studies have been conducted to analyze the link between specific aspects of sleep quality and academic performance. For example, the previously-mentioned study of almost 800 students examined other metrics, including sleep disturbances and habitual sleep efficiency (Maheshwari and Shaukat). While many of these variables tested were subjective, the results that most clearly and objectively demonstrated the effect of sleep on academic performance were those found for sleep latency and sleep duration. Sleep latency, defined as the time it takes to fall asleep after turning off the lights ("How Sleep Latency Impacts the Quality of Your Sleep"), is an easily quantified sleep metric that has been linked to worse academic performance. One study found that for those with a sleep latency less than or equal to 15 minutes, a GPA higher than 3.41 was significantly more common than those with a sleep latency greater than 60 minutes (Maheshwari and Shaukat). It also found that individuals with GPAs lower than 3.41 were more likely to have elevated sleep latencies, and individuals with GPAs greater than 3.41 were more likely to have lower sleep latencies (See Figure 4)(Maheshwari and Shaukat).

In testing sleep duration, the study determined that a GPA greater than 3.41 was more likely among students who slept longer than 7 hours per night on average than those who slept less than 5 hours a night on average (Maheshwari and Shaukat). Furthermore, the study found that those with lower than 3.41 GPA were more likely to sleep fewer hours on average, while those with higher than 3.41 GPAs were more likely to sleep more hours on average (See Figure 4)(Maheshwari and Shaukat).

GPA	Sleep Latency	Duration
>3.41		More hours
<3.41	Greater	

Fig. 4 This figure provides a visual model to conceptualize the difference in sleep latency and sleep duration between those with high and low GPAs.

Another study performed on medical students in Riyadh, Saudi Arabia, analyzed sleep's impact on GPA. The study found that 93.1% of students who reported good sleep quality achieved an unweighted GPA greater than 3.0. In comparison, only 86.89% of students who reported poor sleep quality were able to accomplish an unweighted GPA greater than 3.0 (Alhusseini et al.). This study also related Pittsburgh Sleep Quality Index (PSQI) scores – which assess sleep quality and disturbances over one month – to GPA (Buysse et al.). The PSQI is a useful metric to gauge sleep quality, deprivation, and disturbances. The Riyadh study found that students with lower GPAs (1.50-2.99) had a mean PSQI score higher than students with higher GPAs (greater than 2.99)(Alhusseini et al.). A higher PSQI score indicates poorer sleep quality, so the study concluded that students with higher GPAs had better sleep quality.

Poor sleep quality has also been found to impact resident physicians' performance (Papp et al.). One particular study had residents report many short-term and long-term adverse cognitive effects of sleep loss, including the ability to learn from reading, lectures, and conferences (Papp et al.). Many of these residents connected their poor sleep with a lack of motivation to learn (Papp et al.). Both objective and subjective results indicate that sleep has a notable impact on medical trainees' academic performance and cognitive ability.

While it is clear that sleepiness has some effect on academic performance, the extent to which academic achievement is impacted is unclear. One cross-sectional study, for example, concluded that sleep deprivation had a statistically insignificant impact on academic performance among medical students (Jalali et al.). Though the study included a small sample size (with 102 medical students tested), it provides evidence that much is still to be discovered about the true impact that sleep deprivation can have on medical students' school performance. More research must be done to properly understand the performance implications of sleep deprivation on medical students (Jalali et al.).

4.2 Healthcare Professionals

Many essential functions for optimal work performance are affected as a result of poor sleep quality. Skills such as flexible thinking, decision-making, revision of plans, and attention are all negatively affected when an individual is under increased sleep pressure (Mølgaard et al.). Sleep pressure, or the homeotic sleep drive, is a biological process that builds up a "pressure" in our bodies to sleep when we are sleep-deprived, while the pressure decreases after time spent sleeping (*Module 2. Sleep Pressure*). The effects of poor sleep became evident in the United States in 2011 when the maximum working hours for first-year junior doctors were reduced to 80 hours per week and 16 hours for one shift. Before this change, there were no restrictions on the workweek of young doctors. Across the 15,276 physicians included in the study, the frequency of attention deficits was reduced by 18%, a tremendous decrease after reducing the long working hours of young doctors (Mølgaard et al.). Resident physicians who were included in another study noted adverse effects on their higher-order thinking skills – meaning cognitive abilities and complex thinking – as a result of sleep loss (Papp et al.). Many of the residents noted a reduced ability to diagnose patients and make accurate medical decisions (Papp et al.). Other residents affected by sleep loss mentioned that they were more "following algorithms" instead of "thinking intellectually" while caring for patients (Papp et al.). Another study found that the ability of healthcare workers to make complex rational decisions is reduced with sleep deprivation (Mølgaard et al.). These residents described themselves as inattentive and abrupt when interacting with patients and family members, also reporting that they would act with less patience with families (Papp et al.).

In the medical world, it is vital that healthcare professionals act with the utmost care and respect for the patients and their families. These influences on physicians' attitudes and care toward their patients as a result of poor sleep are detrimental to the performance of healthcare professionals and the patient-physician relationship. Aside from their attitudes toward their work and patients, many residents also described concerns about medical errors occurring due to potential effects of sleep loss and fatigue, including incorrectly entering information into the patient record, prescribing incorrect medications, writing incorrect dosages of drugs, or prescribing for the wrong patient (Papp et al.). In the medical field, physicians must perform their jobs with pinpoint accuracy; there is little room for error and poor sleep quality has been shown to hinder physicians' ability to perform necessary tasks.

5. Physical Health, Relationships, and Sleep Quality

5.1 Medical Students

Many aspects of the personal lives of those learning and working in the medical field are affected by poor sleep. Medical residents have been found to have adverse effects on their interpersonal relationships, ability to perform daily-life tasks, and physical health as a result of sleep loss and fatigue (Papp et al.). If medical students are affected by these negative effects, they will be unable to perform as well in the classrooms, labs, and hospitals. Residents also reported having decreased physical activity and no leisure time (Papp et al.). Another study found that medical students with two or fewer hours of sleep deprivation reported significantly better physical health than those with three to four hours of sleep deprivation (Perotta et al.). Over four hours of sleep deprivation contributed to even worse reported physical health relative to those medical students with sleep deprivation less than four hours (Perotta et al.). The physical health of medical students decreases in correlation with worse sleep quality.

5.2 Healthcare Professionals

Working night shifts also have effects on healthcare workers' physical health. Night shifts are common among medical professionals. One study reported that 91.6% of the junior doctors that they surveyed worked night shifts (Jackson and Moreton), while about 7.4% of America's working population performs any night work (Humans). This type of work schedule can result in circadian misalignment, in which one feels alert and sleepy at inappropriate times (Mølgaard et al.; Jalali et al.). Circadian misalignment leads to increased rates of accidents, Type 2 diabetes, weight gain, cardiovascular diseases, stroke, and cancer (Kecklund and Axelsson). Circadian misalignment also increases risks for hypertension and inflammation (Morris et al.; Mølgaard et al.; Kecklund and Axelsson). Specifically, circadian misalignment has numerous effects on cardiovascular conditions, including an increase in systolic and diastolic blood pressure and increased inflammatory markers(Morris et al.). Systolic blood pressure refers to maximum blood pressure during the contraction of the ventricles, while diastolic pressure refers to the minimum blood pressure just before the vesicle contraction (Buysse et al.). There is a greater risk of stroke and heart disease if high systolic pressure is present than if high diastolic pressure is present.[44,45] One study also determined that cardiometabolic stress is increased by both shift work and sleep loss, indicating a correlation between long-shift work and poor sleep quality (Kecklund and Axelsson). Both night shifts and long shift work can lead to effects on physical health, especially due to cardiovascular diseases.

	Location of Study Particpants	Medical Students or Healthcare	Job Title	Year in Med School	Anxiety	Depression	Stress

		professionals					
Study 1	Sulaiman AlRajhi Colleges, Al-Qassim, Saudi Arabia	Medical students	NA	All years, pre-clinical and clinical	Bad Sleepers: 65.0 Good Sleepers: 31.5	Bad Sleepers: 53.3 Good Sleepers: 22.8	Bad Sleepers: 41.7 Good Sleepers: 11.4
Study 2	China	Healthcare Workers	Clinical Therapists (note during Covid-19)	NA	Bad Sleepers: 91.5 Good Sleepers: 8.5	Bad Sleepers: 88.5 Good Sleepers: 11.5	Bad Sleepers: 94.3 Good Sleepers: 5.7
Study 3	Brazilian Medical Schools	Medical Students	NA	All years	Bad Sleepers: 48.5 Good Sleepers: 41.5	Bad Sleepers: 11.0 Good Sleepers: 7.0	N/A
Study 4	College of Medicine at IMSIU, Riyadh, Saudi Arabia	Medical Students	NA	Pre-clinical years (1st, 2nd, and 3rd years)	N/A	N/A	Bad Sleepers: 68.7 Good Sleepers: 46.2
Study 5	New York City	Healthcare Workers	Hospital Workers	NA	Bad Sleepers: 61.6 Good Sleepers: 39.8	Bad Sleepers: 46.4 Good Sleepers: 25.9	Bad Sleepers: 67.3 Good Sleepers: 52.0
Average	NA	NA	NA	NA	Bad	Bad	Bad

					Sleepers: 66.65 Good Sleepers: 30.33	Sleepers: 49.8 Good Sleepers: 16.8	Sleepers: 68 Good Sleepers: 28.83

Fig. 5 - Data Table. The table demonstrates individual studies and means reporting the impact of sleep on rates of anxiety, depression, and stress.

6. Conclusions

This paper has focused on three main areas that are impacted by sleep deprivation in medical students and workers: mental health, academic/work performance, and overall lifestyle. Sleep deprivation causes adverse effects on the mental health of medical students and healthcare professionals; these effects are apparent when analyzing the increased rates of depression, anxiety, stress, and burnout for those working and studying in the medical field who are sleep deprived. Research has also found that medical students with worse sleep quality or shorter sleep duration have lower GPAs on average, indicating worse academic performance. Similarly, healthcare workers with poor sleep quality or sleep duration report worse work performance and alertness. Finally, medical trainees and healthcare professionals alike have identified and reported adverse effects on interpersonal relationships and physical health as a result of sleep deprivation.

While the effects of sleep deprivation, which is common among medical trainees and healthcare workers, are apparent in mental health, academic/work performance, and personal life, the extent of these effects is unclear. However, to minimize the potential negative consequences of poor sleep, medical students and healthcare professionals should implement solutions to improve the duration and quality of their sleep. Most importantly, though, awareness must be raised regarding sleep's countless impacts on the personal and professional lives of those working and studying in the medical field. Students and physicians alike should be instructed to prioritize rest and not overexert themselves in working shifts, especially late-night ones. Those involved in medical students' paths to becoming physicians, especially teachers of medical students, must take adequate measures and explicate to their students the importance of sleep to improve the sleep quality of the students. Policy changes should also be enacted to support both students and healthcare professionals. For example, the maximum shift hours and the maximum late-night shift hours could be reduced further to ensure that medical students and healthcare professionals get sufficient sleep (8-10 hours of sleep, as reported by one study)(Tran et al.).

Medical students and healthcare professionals have a great responsibility as the upholders of a crucial societal role to protect people's health and well-being through innovation and expertise. Sleep deprivation has been found to counteract the ability of medical students and healthcare professionals to learn, work, and remain mentally and physically healthy. Due to the numerous detriments to medical students and healthcare professionals' quality of life and academic/work performance that arise as a result of sleep deprivation, the medical community must take action to mitigate the adverse risks of poor sleep quality on those working and studying in the medical field.

Works Cited

Alhusseini, Noara Khaled, et al. "Effects of Sleep Quality on Academic Performance and Psychological Distress Among Medical Students in Saudi Arabia." *Health Scope*, vol. 11, no. 2, 2, 2022. *brieflands.com*, https://doi.org/10.5812/jhealthscope-123801.

Alotaibi, Abdullah D., et al. "The Relationship between Sleep Quality, Stress, and Academic Performance among Medical Students." *Journal of Family & Community Medicine*, vol. 27, no. 1, 2020, pp. 23–28. *PubMed Central*, https://doi.org/10.4103/jfcm.JFCM_132_19.

Average Physician Workweek: How Doctors' Hours Are Changing | Staff Care. https://www.amnhealthcare.com/blog/physician/locums/average-physician-workweek-how-doctors-hours-are-changing/. Accessed 14 Mar. 2024.

Brzezinski, Walter A. "Blood Pressure." *Clinical Methods: The History, Physical, and Laboratory Examinations*, edited by H. Kenneth Walker et al., 3rd ed., Butterworths, 1990. *PubMed*, http://www.ncbi.nlm.nih.gov/books/NBK268/.

Buysse, D. J., et al. "The Pittsburgh Sleep Quality Index: A New Instrument for Psychiatric Practice and Research." *Psychiatry Research*, vol. 28, no. 2, May 1989, pp. 193–213. *PubMed*, https://doi.org/10.1016/0165-1781(89)90047-4.

Chang, Edward. "Typical Day of A UCLA Medical Student (Pre-Clinical)." *ProspectiveDoctor*, 20 Jan. 2014, https://www.prospectivedoctor.com/typical-day-of-a-ucla-medical-student-pre-clinical/.

Coelho, Julien, et al. "Sleep Timing, Workplace Well-Being and Mental Health in Healthcare Workers." *Sleep Medicine*, vol. 111, Nov. 2023, pp. 123–32. *PubMed*, https://doi.org/10.1016/j.sleep.2023.09.013.

"College vs. Medical School Comparison." *Med School Insiders*, 17 Apr. 2018, https://medschoolinsiders.com/pre-med/college-vs-medical-school/.

Corrêa, Camila de Castro, et al. "Sleep Quality in Medical Students: A Comparison across the Various Phases of the Medical Course." *Jornal Brasileiro de Pneumologia*, vol. 43, no. 4, 2017, pp. 285–89. *PubMed Central*, https://doi.org/10.1590/S1806-37562016000000178.

Definition of NREM Sleep - NCI Dictionary of Cancer Terms - NCI. 2 Feb. 2011, https://www.cancer.gov/publications/dictionaries/cancer-terms/def/nrem-sleep. nciglobal,ncienterprise.

Dempsey-Klott, Nick. "Law School Daily Study Schedule." *JD Advising*, 18 Aug. 2022, https://jdadvising.com/law-school-daily-study-schedule/.

Ganesan, Saranea, et al. "The Impact of Shift Work on Sleep, Alertness and Performance in Healthcare Workers." *Scientific Reports*, vol. 9, no. 1, Mar. 2019, p. 4635. *PubMed*, https://doi.org/10.1038/s41598-019-40914-x.

Gordon, Amie M., et al. "Sleep and Social Relationships in Healthy Populations: A Systematic Review." *Sleep Medicine Reviews*, vol. 57, June 2021, p. 101428. *PubMed*, https://doi.org/10.1016/j.smrv.2021.101428.

Haden, Rebecca. "What Is Sleep Debt and How Do You Get Rid of It?" *Medical Associates of Northwest Arkansas*, 12 Feb. 2020, https://mana.md/what-is-sleep-debt-and-how-do-you-get-rid-of-it/.

“How Sleep Latency Impacts the Quality of Your Sleep.” *Sleep Foundation*, 26 Aug. 2021, https://www.sleepfoundation.org/how-sleep-works/sleep-latency.

Humans, IARC Working Group on the Identification of Carcinogenic Hazards to. “1. Exposure Data.” *Night Shift Work*, International Agency for Research on Cancer, 2020. *www.ncbi.nlm.nih.gov*, https://www.ncbi.nlm.nih.gov/books/NBK568199/.

Inman, Ryan. “Dropping Out of Medical School: Drop Out Rate + Top Reasons.” *Financial Residency*, 29 Dec. 2022, https://financialresidency.com/dropping-out-of-medical-school/.

Jackson, Emma J., and Adam Moreton. “Safety during Night Shifts: A Cross-Sectional Survey of Junior Doctors’ Preparation and Practice.” *BMJ Open*, vol. 3, no. 9, Sept. 2013, p. e003567. *bmjopen.bmj.com*, https://doi.org/10.1136/bmjopen-2013-003567.

Jalali, Rostam, et al. “The Effect of Sleep Quality on Students’ Academic Achievement.” *Advances in Medical Education and Practice*, vol. 11, July 2020, pp. 497–502. *PubMed Central*, https://doi.org/10.2147/AMEP.S261525.

Kecklund, Göran, and John Axelsson. “Health Consequences of Shift Work and Insufficient Sleep.” *BMJ (Clinical Research Ed.)*, vol. 355, Nov. 2016, p. i5210. *PubMed*, https://doi.org/10.1136/bmj.i5210.

Leal, Pollyana Caldeira, et al. “Trait vs. State Anxiety in Different Threatening Situations.” *Trends in Psychiatry and Psychotherapy*, vol. 39, Aug. 2017, pp. 147–57. *SciELO*, https://doi.org/10.1590/2237-6089-2016-0044.

Machado-Duque, Manuel Enrique, et al. “[Excessive Daytime Sleepiness, Poor Quality Sleep, and Low Academic Performance in Medical Students].” *Revista Colombiana De Psiquiatria*, vol. 44, no. 3, 2015, pp. 137–42. *PubMed*, https://doi.org/10.1016/j.rcp.2015.04.002.

Maheshwari, Ganpat, and Faizan Shaukat. “Impact of Poor Sleep Quality on the Academic Performance of Medical Students.” *Cureus*, vol. 11, no. 4, Apr. 2019, p. e4357. *PubMed*, https://doi.org/10.7759/cureus.4357.

Marvaldi, Maxime, et al. “Anxiety, Depression, Trauma-Related, and Sleep Disorders among Healthcare Workers during the COVID-19 Pandemic: A Systematic Review and Meta-Analysis.” *Neuroscience and Biobehavioral Reviews*, vol. 126, July 2021, pp. 252–64. *PubMed Central*, https://doi.org/10.1016/j.neubiorev.2021.03.024.

MD, Akshay Goel. “Step 1 vs. Step 2 Comparison - Difficulty, Scoring & Knowledge.” *Medlearnity*, 7 Dec. 2021, https://www.medlearnity.com/usmle-step-1-vs-step-2/.

Module 2. Sleep Pressure: Homeostatic Sleep Drive | NIOSH | CDC. 2 Apr. 2020, https://www.cdc.gov/niosh/work-hour-training-for-nurses/longhours/mod2/11.html.

Mølgaard, Jesper, et al. “[Consequences of sleep deprivation on healthcare workers].” *Ugeskrift for Laeger*, vol. 183, no. 26, June 2021, p. V08200579.

Morris, Christopher J., et al. “Circadian Misalignment Increases Cardiovascular Disease Risk Factors in Humans.” *Proceedings of the National Academy of Sciences of the United*

States of America, vol. 113, no. 10, Mar. 2016, pp. E1402-1411. *PubMed*, https://doi.org/10.1073/pnas.1516953113.

Oh, Chang-Myung, et al. "The Effect of Anxiety and Depression on Sleep Quality of Individuals With High Risk for Insomnia: A Population-Based Study." *Frontiers in Neurology*, vol. 10, Aug. 2019, p. 849. *PubMed Central*, https://doi.org/10.3389/fneur.2019.00849.

"One in Five Young French People Has a Depressive Disorder." *Le Monde.Fr*, 14 Feb. 2023. *Le Monde*, https://www.lemonde.fr/en/science/article/2023/02/14/one-in-five-young-french-people-has-a-depressive-disorder_6015640_10.html.

Papp, Klara K., et al. "The Effects of Sleep Loss and Fatigue on Resident-Physicians: A Multi-Institutional, Mixed-Method Study." *Academic Medicine: Journal of the Association of American Medical Colleges*, vol. 79, no. 5, May 2004, pp. 394–406. *PubMed*, https://doi.org/10.1097/00001888-200405000-00007.

Paprocki, Jonathan. "CDC: More than 1 in 3 Americans Are Sleep-Deprived." *Sleep Education*, 4 Mar. 2011, https://sleepeducation.org/cdc-americans-sleep-deprived/.

Perotta, Bruno, et al. "Sleepiness, Sleep Deprivation, Quality of Life, Mental Symptoms and Perception of Academic Environment in Medical Students." *BMC Medical Education*, vol. 21, Feb. 2021, p. 111. *PubMed Central*, https://doi.org/10.1186/s12909-021-02544-8.

Reddy, Sujana, et al. "Physiology, Circadian Rhythm." *StatPearls*, StatPearls Publishing, 2024. *PubMed*, http://www.ncbi.nlm.nih.gov/books/NBK519507/.

"REM Sleep Revealed: Enhance Your Sleep Quality." *Sleep Foundation*, 16 Dec. 2021, https://www.sleepfoundation.org/stages-of-sleep/rem-sleep.

Scammell, Thomas E. "Overview of Sleep: The Neurologic Processes of the Sleep-Wake Cycle." *The Journal of Clinical Psychiatry*, vol. 76, no. 5, May 2015, p. e13. *PubMed*, https://doi.org/10.4088/JCP.14046tx1c.

Scott, Alexander J., et al. "Improving Sleep Quality Leads to Better Mental Health: A Meta-Analysis of Randomised Controlled Trials." *Sleep Medicine Reviews*, vol. 60, Dec. 2021, p. 101556. *PubMed*, https://doi.org/10.1016/j.smrv.2021.101556.

Shaik, Likhita, et al. "Sleep and Safety among Healthcare Workers: The Effect of Obstructive Sleep Apnea and Sleep Deprivation on Safety." *Medicina*, vol. 58, no. 12, Nov. 2022, p. 1723. *PubMed Central*, https://doi.org/10.3390/medicina58121723.

Stewart, Nancy H., et al. "Sleep Disturbances in Frontline Health Care Workers During the COVID-19 Pandemic: Social Media Survey Study." *Journal of Medical Internet Research*, vol. 23, no. 5, May 2021, p. e27331. *PubMed Central*, https://doi.org/10.2196/27331.

"Student Study Time Matters - CollegiateParent." *CollegiateParent - Connecting College Parents with Insider Information about Their Student's College and Local Community.*, 5 Jan. 2022, https://www.collegiateparent.com/academics/student-study-time-matters/.

Szumilas, Magdalena. "Explaining Odds Ratios." *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, vol. 19, no. 3, Aug. 2010, pp. 227–29.

Tran, Duc-Si, et al. "Stress and Sleep Quality in Medical Students: A Cross-Sectional Study from Vietnam." *Frontiers in Psychiatry*, vol. 14, Nov. 2023, p. 1297605. *PubMed Central*, https://doi.org/10.3389/fpsyt.2023.1297605.

Watson, Nathaniel F., et al. "Recommended Amount of Sleep for a Healthy Adult: A Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society." *Sleep*, vol. 38, no. 6, June 2015, pp. 843–44. *PubMed Central*, https://doi.org/10.5665/sleep.4716.

Weaver, Matthew D., et al. "Sleep Disorders, Depression and Anxiety Are Associated with Adverse Safety Outcomes in Healthcare Workers: A Prospective Cohort Study." *Journal of Sleep Research*, vol. 27, no. 6, Dec. 2018, p. e12722. *PubMed*, https://doi.org/10.1111/jsr.12722.

"What Is The Average Work Hours Per Week In The US? [2023]." *Zippia*, 9 Jan. 2023, https://www.zippia.com/advice/average-work-hours-per-week/.

What Is Trait Anxiety? Definition, Examples, and Treatment. 19 Apr. 2022, https://www.medicalnewstoday.com/articles/trait-anxiety.

Which Blood Pressure Number Is Important? - Harvard Health. https://www.health.harvard.edu/staying-healthy/which-blood-pressure-number-is-important. Accessed 8 Apr. 2024.

Zhang, Yuan, et al. "Work-Family Conflict and Depression Among Healthcare Workers: The Role of Sleep and Decision Latitude." *Workplace Health & Safety*, vol. 71, no. 4, Apr. 2023, pp. 195–205. *PubMed*, https://doi.org/10.1177/21650799221139998.

The Optimization of Neural Networks By Zubin Gupta[1] and Cyrus Ayubcha[2]

Abstract

A neural network is a type of Artificial Intelligence consisting of multiple layers of neurons. Neural networks can be trained to perform a specific task, modeled after the human brain.This paper will discuss the efficiency problem in algorithms containing neural networks. It will demonstrate how to optimize these neural networks by creating a neural network and measuring its accuracy and speed on specific tests. Afterward, some neural network features will be changed and measured again. Eventually, the data will be enough to create a more optimized neural network that will be more accurate and precise than the others. One of the most influential factors in the speed of a neural network is its max pooling size or the number of values pooled together into one. Another result is that most other modifications, like dropout size or number of convolution layers, do not significantly affect the accuracy and speed of the neural network.

Keywords

Robotics; Intelligent Machines; Machine Learning; Convolution Layer; Neural Network

Introduction

Through the recent surge in the development and usage of Artificial Intelligence (AI), computer scientists have been able to create machine learning algorithms for various applications. However, the upkeep of large models can be expensive, so it is necessary to ensure that the value of money is maximized for a large algorithm.[1,2] We can do that by changing the attributes of the neural network. If it were to work, it would be an efficient solution that does not require additional money spent.[2] That brings us to the question: How does changing the attributes of a neural network change its efficiency? A convolutional network is a type of AI that turns all the pixels in an image into a 3d matrix of values going from 0 to 1, puts the values through different layers, and calculates values based on that.[1] Some values include the convolution layer, which takes the dot product of each 2d value matrix and one created by the algorithm, resulting in another 3d. Afterward, there is the batch normalization layer, which essentially rescales the weights that modify the matrix of values. The max pooling layer follows after, which takes subsets of a specific area out of the value matrix, finds the maximum value for each subset, and uses those values to construct another matrix.[1,3] This process is repeated for all matrices. The dropout layer comes next: its job is to drop out random values in the neural network.[2] Those four layers can be repeated multiple times, but not infinitely, as each sequence through the four layers makes the matrix of values smaller.[1] After that, a few more layers are applied. The flattening first converts the 3d matrix into a 1d vector. The dense layer is next, and weights are applied to the vector to make the data smaller. A batch normalization layer follows, and there is one more dense layer, which gives the output. [1,2] While optimization studies have long attempted to develop complex methods and structures to improve training efficiency while retaining performance, this study utilizes the Keras package's most widely used convolutional neural network structures. This study will augment common hyperparameters and CNN size and assess performance in various datasets. I can optimize a neural network without paying or waiting excessive time. Other companies can look at this research and determine what would be best for their models. Even if my example could be more satisfactory for some companies due to my algorithm's specific purpose, computer scientists can observe my research and even attempt a scaled-down version of their AI similarly, leading to sufficient results. In sum, my goal is to catalyze the growth of AI by giving an example of optimizing a neural network.

Methods

Other key terms include the epoch, which is how long the model is trained and tested. The batch size is the number of images propagated through the neural network.

The program used in this study requires a convolutional neural network to classify different images. This means a dataset and a convolutional neural network are required.[3,4] The experiment requires a neural network with set attributes and progressive modifications over time. The resulting data will be kept in a spreadsheet for further analysis.

Python will be used as the coding language in the program, and the TensorFlow Keras library will be used for the neural network. A traffic light classification dataset will be used initially. The program's goal is to see what percentage of traffic lights it can adequately identify after given some information about the traffic lights.

Using the time taken, I will measure the approximate average time per epoch for the dataset. This is because if I run a different number of epochs on each neural network, the timing will not be unbiased. I will use the peak validation accuracy achieved at any time in the neural network as the accuracy, as there would be no reason to use a less accurate neural network if a more accurate one already exists. [4]

Results and Discussion

In this dataset, the neural network has to identify if a traffic light has a yellow light, a red light, a green light, or none. The experiment begins using three convolution layers, a batch size of 15, and a max pooling size of (2,2) (Figure 1). This will serve as the baseline for the project.

The experiment is continued by removing convolution layers, thinking a smaller number of layers would mean fewer parameters and a faster neural network. However, the number of parameters doubled due to the increased work of the dense layer, which had to compress more data (Figure 4). The results will be compared to the original, and if there are changes, the experiment will continue in that way.

As a convolution layer had already been added, the experiment continued by removing convolution layers, and the number of parameters sharply decreased (Figure 7).

At this point, predicting the traffic light color would be a relatively simple task for a neural network. So the dataset was replaced by a lung disease classification dataset. For convolution models, distinguishing different medical conditions is more complicated than other datasets, like traffic light classification.

The max pooling size was changed to (3,3) instead of (2,2) to lower vector sizes, and the program's dropout size was changed to 0.25 for consistency (Figure 10).

The program progressed with a change of optimizer. The other models used an optimizer called RMSProp, so the program's optimizer was changed to a different one called Adam (Figure 13). The experiment progressed by raising the batch size from 15 to 50 so more images could be propagated through the neural network. (Figure 16).

One of the main factors that affected the traffic light dataset was the max pooling size, as increasing it decreased the accuracy and the time the program took. This resulted in a lowering of the max pooling size to (2,2) (Figure 19).

At this point, the program called for more dramatic changes, as the previous versions all had similar accuracy. It started from 3 convolution layers to 5 (Figures 22 and 23).

The number of convolution layers decreased in the program to 3, and a new factor, the dropout size, was lowered to 0.1. (Figure 25).

The experiment finishes with one last adjustment, increasing the dropout size to 0.5 (Figure 28). The most influential result in this experiment is the dataset. When the switch between datasets was made, the accuracy went from hovering around a specific number to hovering around another number. One of the defining factors for the time per epoch was the max pooling. When the max pooling had a larger area, the time was reduced, as there were fewer parameters to deal with. However, many parameters did not change the results significantly. For example, the dropout size, the number of convolution layers, the batch size, and the optimizer. Aside from that, one of the influential factors is the number of epochs that were able to run, as generally, the higher number of epochs that ran allowed for higher data accuracy.

There are a few potential reasons why those factors did not change the program as intended. For example, the number of convolution layers did not change anything. Although the increase in neurons guarantees a better-fitting trend, the model was already optimized without the additional layers. Additionally, this generates reasoning similar to the different optimizers and batch sizes. Both optimizers and batch sizes optimize the program, and they do it to a level such that the minor differences in what they do barely influence the results. The dropout size potentially did not affect the results because of the neurons chosen to be dropped.[1,5]

Conclusion

In this paper, I demonstrated how to optimize a particular neural network with two datasets. Although the same modifications may be less effective on other datasets, the project demonstrates an effective template for how different aspects of neural networks can improve. Optimized neural networks can finish a task in a shorter time, so less money must be spent. That money can go towards even further optimizing the neural network. Enhancing the performance of neural networks allows for higher efficiency, causing a given neural network to be faster and more computationally efficient. We discovered that the dataset heavily influences the neural networks' accuracy and speed. Still, as the specific task assigned to a neural network does not vary, neural networks must be optimized to their highest capability. Significantly larger neural networks, as non-efficient ones would end up as a large amount of money squandered into an AI that pales in comparison to more efficient ones.[2,4]

Acknowledgements

Thank you for the guidance of Cyrus Ayubcha, a mentor from Harvard University, in developing this research paper.

Authors

Zubin Gupta is a Lexington High School sophomore pursuing a career in Computer Science. He also has a focus on charity with his nonprofit, the Twynphony Hopes Foundation.

Works Cited

Broderick, T. *Impact of CNNs AlexNet*. (2010). https://introml.mit.edu/_static/spring23/VideoSlides/cnn.pdf

Overview of Artificial Intelligence Technology | FINRA.org. (2013). Finra.org. https://www.finra.org/rules-guidance/key-topics/fintech/report/artificial-intelligence-in-the-securities-industry/overview-of-ai-tech

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). *Learning Transferable Visual Models From Natural Language Supervision*. ArXiv.org. https://arxiv.org/abs/2103.00020

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. https://www.deeplearningbook.org/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*. ArXiv.org. https://arxiv.org/abs/1706.0376

Prediction of the behavior of the inverter and electric motor using machine learning By Ekaterina Koneva

Abstract

This study investigates the use of machine learning approaches to anticipate the behaviour of inverters and electric motors, with a specific focus on electric vehicle drive trains. Using a large dataset of around 40 million samples, the study seeks to improve the efficiency, reliability, and performance of these components. The key goals include developing and evaluating multiple machine learning models to predict direct and quadrature axis currents, comparing alternative modelling methodologies, and optimising model performance under various operational scenarios.

Three different modelling approaches were used: a single predictive model, models for each elementary vector, and models for pairs of vectors. The results showed that the random forest model consistently outperformed other models, with good prediction accuracy and robustness.

The findings indicate that machine learning models, particularly ensemble approaches such as random forests, have a high potential for forecasting inverter and motor behaviour, hence contributing to better efficiency, dependability, and operational longevity of electric vehicles.

Introduction

Using machine learning to predict how inverters and electric motors behave is very important for making them work better, last longer, and cost less to operate. These technologies are used in things like electric cars, factories, and systems that manage solar and wind energy. Machine learning helps by looking at considerable amount of data from these devices to spot possible problems before they happen, manage power use, and adjust to changes smoothly. This smart approach means that machines can be fixed before they break down, avoiding unexpected stops and saving money on repairs. Moreover, as these smart models keep learning and improving, they make these important technologies even more efficient and reliable. This not only boosts performance and safety but also helps in creating greener and more sustainable energy solutions.

The primary objective of this research is to utilize machine learning techniques to accurately predict the behaviour of the inverter and electric motor within an electric vehicle's drive train. By leveraging the extensive dataset provided, this study aims to develop predictive models that can enhance the efficiency, reliability, and performance of electric vehicles. Specific goals include: (1) Train and evaluate various machine learning models to predict direct and quadrature axis currents (**id_k1** and **iq_k1**)at the next time step, using different modelling approaches, (2) Implement and compare three distinct modelling approaches—single predictive model, models for each elementary vector, and models for pairs of vectors—to determine the most effective method for accurate predictions, (3) Optimize the models to ensure they provide reliable predictions across various operational conditions, thereby contributing to the overall efficiency and longevity of electric vehicle drive trains.

By achieving these objectives, the study aims to provide valuable insights into the application of machine learning in electric vehicle technology, promoting advancements in predictive maintenance, energy efficiency, and sustainable transportation solutions.

Literature review

The dataset described by Hanke et al. (2020a, 2020b) is designed to predict the electrical behaviour of an inverter and electric motor within an electric vehicle's drive train. It provides a comprehensive framework to facilitate the extraction of models using machine learning techniques. The system comprises a battery, inverter, electric motor, and controller, with the inverter converting the battery's DC voltage into a three-phase AC voltage to operate the motor at various speeds.

Hanke et al.(2020a) outlines a dataset aimed at predicting the electrical behaviour of an inverter and motor in an electric vehicle's drive train using a single framework. It provides a simplified introduction to the system behind the data and explains how to use the dataset. The system consists of a battery, inverter, electric motor, and controller, with the inverter converting the battery's DC voltage into a three-phase AC voltage to operate the motor at different speeds. The dataset is intended for training machine learning models to accurately predict the behaviour of the inverter and motor. Furthermore, Hanke et al.(2020a) emphasises the importance of accurately predicting the behaviour of the electric drive train to achieve high efficiency and range in electric vehicles. It also provides insights into the basic operating principles of the system, the components and their basic electrical models, and the test bench parameters.

Hanke et al.(2020b) describes a data set created to analyse the physical characteristics of an electric motor using machine learning methods. It consists of approximately 40 million samples from a defined operating range of the drive, including measurements of dq-currents, rotational angles, and information about the elementary vectors selected in the controller cycle. The paper explains how to use the data set and extract models using methods such as ordinary differential equations, the least squares method, and machine learning. It also discusses the structure of the drive system, the extraction of models from data, and the potential for balanced data sets for machine learning methods. The paper delves into the system's finite-control-set model predictive control, which involves solving an optimal control problem on a receding prediction horizon. It outlines the basic concept of model predictive control, the structure of the drive system, and the extraction of system models using various approaches such as the least squares method and training of neural networks. The paper also highlights the need for models that cover the motor behaviour in all operating points sufficiently well to achieve high control performance.

Dataset

The dataset is made up of sensor data acquired from a permanent magnet synchronous motor (PMSM) put on a test bench, which represents a traction drive manufactured in Germany. The LEA department at Paderborn University did the measurements. The dataset contains around 40 million samples, with each row representing currents flowing through the motor's stator windings before and after a certain switching state within the pre-aligned inverter is applied. Furthermore, the dataset contains the rotor angle at that time. The inverter is a critical power electronic component that controls how the battery voltage is applied to the PMSM's three-phase circuits.

Efficient motor control is critical for improving electric vehicle performance and range. Inefficient control can result in wasteful power loss, heat build-up, and premature battery depletion. The controller's job is to achieve the desired torque at the vehicle's wheels by regulating the electric motor's currents using applied voltages. As a result, properly forecasting currents for a certain switching command is critical for attaining efficient control. This dataset contains useful information for creating control algorithms and optimising the performance of electric car drive trains. Hanke et al. (2020a) and Hanke et al. (2020b) provide both basic and extensive explanations of the system and its data.

The dataset consists of seven columns, each providing crucial information for understanding the behaviour of the permanent magnet synchronous motor (PMSM) on the test bench as shown in **Table 1**. It comprises 40,000,000 entries. **iq_k**, **id_k** represent the currents in d/q-coordinates flowing through the motor's stator windings before a switching state is applied, while **epsilon_k** denotes the rotor angle at that moment. The columns **n_k** and **n_1k** are integers indicating the state of the motor, possibly relating to different operational modes or conditions. Finally**, id_k1** and **iq_k1** represent the currents in d/q-coordinates after the switching state is applied.

Table 1. Electric motor dataset

id_k	**iq_k**	**epsilon_k**	**n_k**	**n_1k**	**id_k1**	**iq_k1**
-81.45802	229.5293	2.240254	5	6	-105.7382	167.3617
-140.6821	112.4234	-1.610116	7	2	-174.8971	128.2237
-127.0724	171.7438	-1.971891	4	7	-92.96102	126.6081
-42.2788	120.1495	1.300341	2	7	-82.2331	124.3379
-48.02003	10.97132	-1.778834	1	4	-45.73148	11.60761

Exploratory data analysis (EDA)

Table 2 illustrates the summary statistics of our data. The feature **id_k** has a mean value of -99.586 and a minimum range of 240.000, indicating substantial variation within the data. Similarly, **iq_k** has a mean of 93.548 and the same maximum range of 240.000, suggesting significant diversity in this feature as well. The **epsilon_k** feature shows a mean of 0.016 and a range of 6.284, reflecting its comparatively smaller variability. The feature **id_k1** displays a mean value of -98.771 and the largest range of 358.427, highlighting considerable spread in its values. **iq_k1** has a mean of 94.814 and a range of 348.226, underscoring its wide variability.

Table 2. Summary statistics of the data

	id_k	**iq_k**	**epsilon_k**	**n_k**	**n_1k**	**id_k1**	**iq_k1**
count	4.E+07	4.E+07	4.E+07	4.E+07	4.E+07	4.E+07	4.E+07
mean	-99.586	93.548	0.016	3.781	3.776	-98.771	94.814
std	61.259	59.333	1.810	2.079	2.079	64.444	61.738

min	-240.000	0.000	-3.142	1.000	1.000	-287.281	-11.678
0.250	-147.139	44.210	-1.545	2.000	2.000	-147.697	44.242
0.500	-94.018	86.523	0.038	4.000	4.000	-94.175	86.813
0.750	-47.932	138.687	1.591	6.000	6.000	-47.737	139.572
max	0.000	240.000	3.142	7.000	7.000	71.146	336.548

We computed the Pearson's correlations between numerical features as depicted in Figure 1. Strong correlations between **id_k** and **id_k1** (0.901728), **iq_k** and **iq_k1** (0.946273) have been observed. Other correlations between variables are margin.

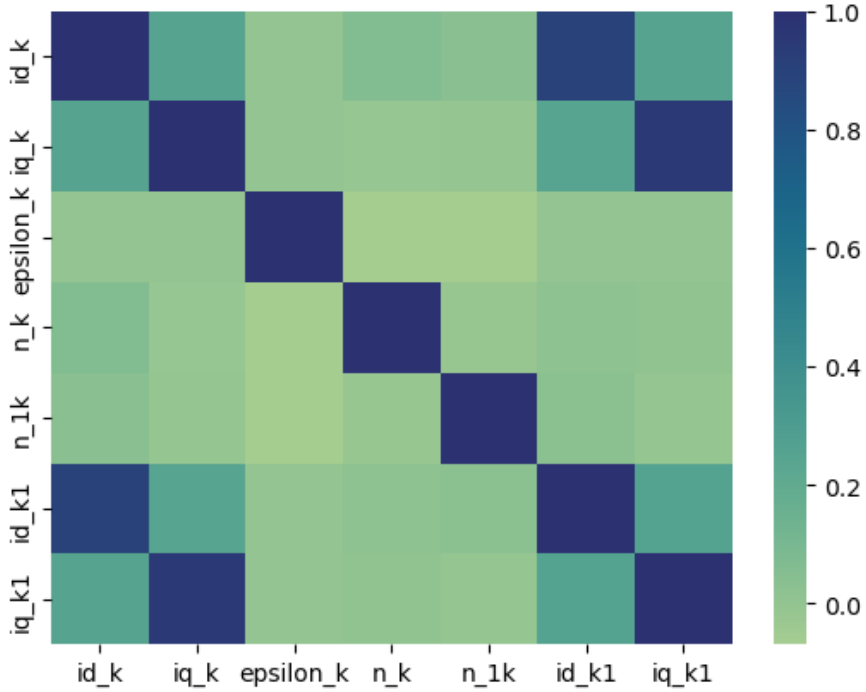


Figure 1. Correlations between numerical features

Figure 2 shows correlations between **id_k** and **id_k1**, **iq_k** and **iq_k1**. As it can be seen, the linear relationship is present between those features. The strongest correlation has been observed for elementary vector 1 (**n_k** = 1)

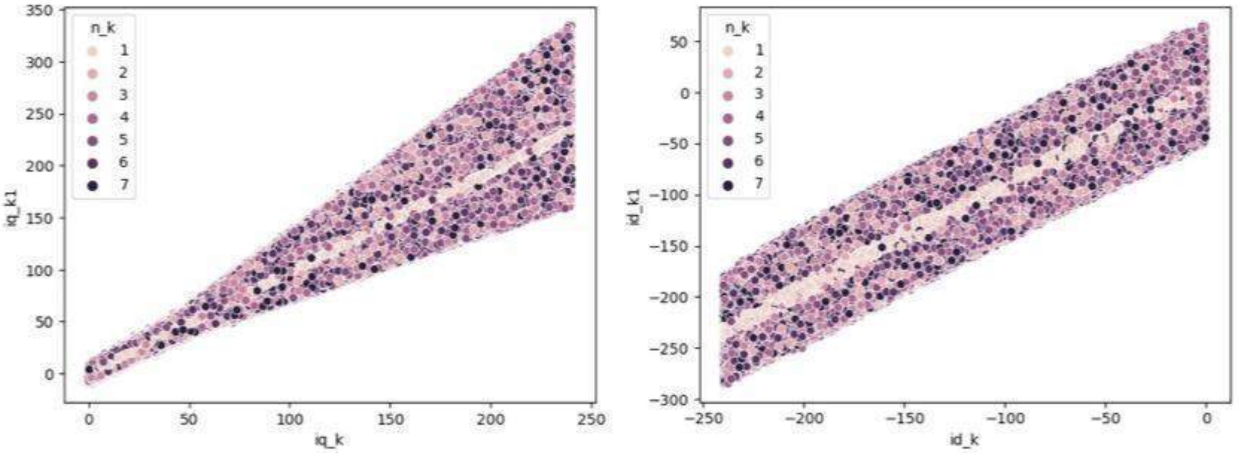


Figure 2. Relationships between **iq_k** and **iq_k1**, **id_k** and **id_k1**

Figure 3. shows that as **n_k** increases, so does the median of **iq_k**. There are plenty of outliers above the upper whiskers, especially for higher values of **n_k**, but the interquartile ranges are still largely consistent. **iq_k1** in comparison to **n_k**: The analogies between this plot and the **iq_k** vs. **n_k** box plot suggests that **iq_k** and **iq_k1** behave similarly for a range of **n_k** values. The medians are negative and only marginally increase as **n_k** increases, according to the distribution of **id_k** over various values of **n_k**. On both ends of the whiskers, there are outliers, but overall, the IQRs are rather stable. This plot's pattern closely resembles that of the **id_k** vs. **n_k** box plot, suggesting that **id_k1** behaves similarly throughout **n_1k**.

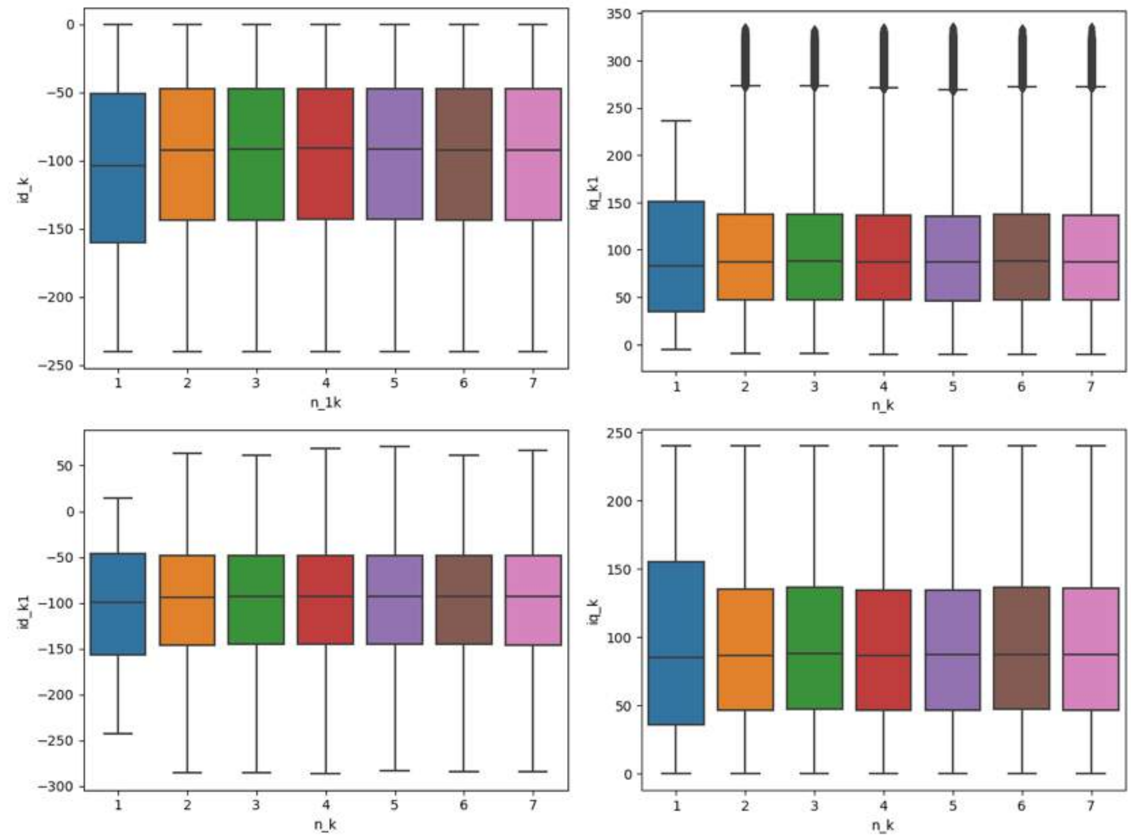


Figure 3. distribution of **iq_k**, **id_k**, **iq_k1** and **id_k1** across different values of **n_k**

Figure 4 shows a peak at lower positive values and a tail extending towards higher positive values, the **iq_k** histogram looks to be right-skewed. The histogram for **iq_k1** is comparable to that of **iq_k** but goes farther into the positive range, signifying the existence of greater values. **Figure 4** displays a unimodal distribution that is tilted to the right, with most values being negative and a peak occurring close before 0. Like **id_k**, **id_k1**'s distribution exhibits left skewness, with a peak that is closer to 0 and most values being negative.

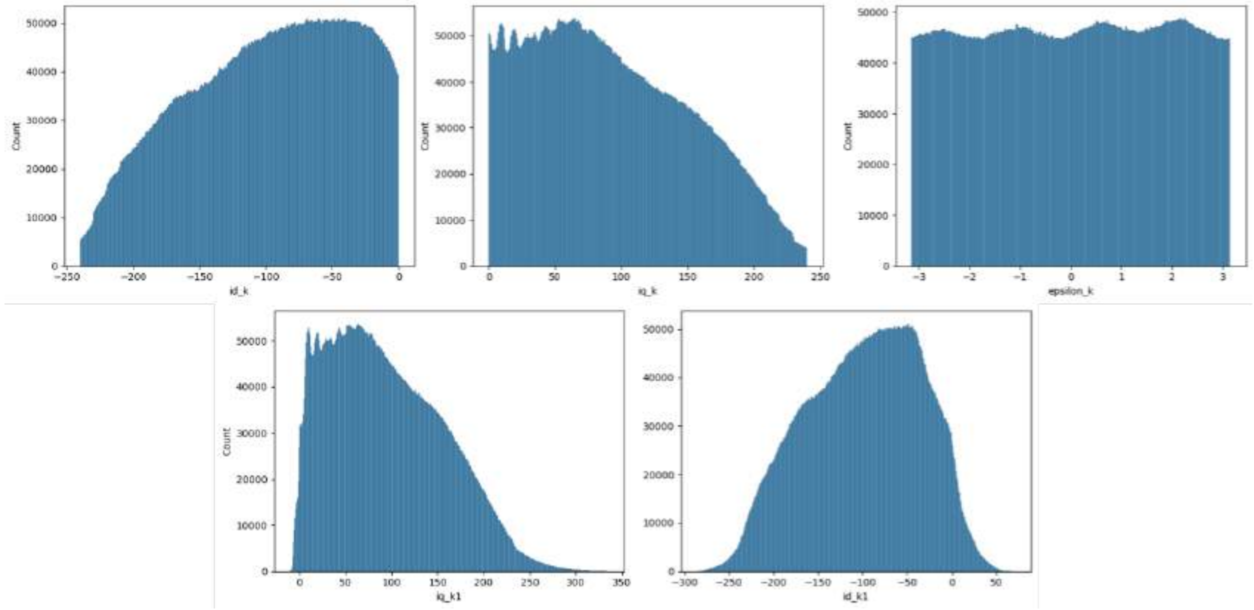


Figure 4. illustrates distribution of values by histograms

Methods and Models

Firstly, we normalised our input data using standard scaler. Normalisation is commonly defined as adjusting a group of values so that they lie inside a short, specified range, such as 0 to 1, or -1 to 1. This is sometimes accomplished by subtracting the minimum value of each attribute and dividing by the range (maximum minus minimum). This strategy is useful when you need to ensure that a feature's numerical values do not outweigh others simply because of their magnitude. Standardisation entails rescaling the characteristics to have a mean of zero and a standard deviation of one. This is performed by removing the mean from each feature and dividing by the standard deviation. Standardisation does not limit values to a specified range, which may be required for models that presume normally distributed data.

Next, we splitted our data into training and testing sets. We used the training data to fit the model, often accounting for 80% of the overall dataset. Exactly 20% of the dataset is used to test the model once it has been trained. This set assists in determining how well the model works on previously unseen data, allowing for the detection of overfitting or underfitting.

We trained and tested baseline linear regression, lasso, ridge regression, decision tree and random forest models. There was no need for hyperparameter tuning in this study, because the baseline models performed well.

Linear regression: By fitting a linear equation to observed data, linear regression describes the connection between a dependent variable and one or more independent variables. By minimising the sum of the squared differences between the responses predicted by the linear approximation and the observed responses, the coefficients of the equation are obtained.

Lasso (Least Absolute Shrinkage and Selection Operator) regression: A type of linear regression in which the absolute size of the regression coefficients is penalised using L1 regularisation. By doing this, Lasso aids in the process of selecting features by reducing the coefficients of less significant features to zero, so eliminating certain features completely. To automate specific steps in the model selection process, this approach is effective for models with high degrees of multicollinearity.

Ridge regression: Ridge Regression is a modification of linear regression that involves the addition of a penalty equal to the square of the coefficients' magnitude. By decreasing the coefficients, this L2 regularisation technique can lower the model complexity; but, unlike Lasso, it does not decrease the coefficients to zero. When dealing with multicollinearity or trying to improve prediction accuracy by lowering overfitting, it is especially helpful.

Decision tree: A non-linear model divides the data into branches to create a visual representation of a tree of decisions. The procedure is transparent and simple to comprehend: every internal node represents a "test" on an attribute, every branch indicates the test's result, and every leaf node represents a class label.

Random forest: Random Forest enhances the decision tree algorithm by generating an ensemble of decision trees, typically trained using the "bagging" method. The primary idea is to merge numerous decision trees to determine the final output rather than depending solely on individual decision trees. Random Forest reduces variance and thus overfitting, resulting in a more robust model than a single decision tree.

We used mean squared error (MSE) and R-squared (R^2) as our model evaluation metrics. **MSE (Mean Squared Error)** is a risk metric that computes the average of the squared errors, or the mean squared difference between estimated and actual values. MSE is a quality metric for estimators; it is always non-negative, with values closer to zero being better. A lower MSE suggests that the model is better at predicting data. R^2 **(R-squared)** is the coefficient of determination is a statistical measure that quantifies the amount of a dependent variable's variation explained by an independent variable or variables in a regression model. R^2 values range from 0 to 1, indicating the model's quality of fit and its ability to predict unseen samples accurately. An R^2 value of 1 shows the perfect fit.

We implemented three different approaches to predict **id_k1** and **iq_k1**. In the first approach, a single predictive model is used as shown in **Figure 5**. The inputs to the model consist of the index of the current vector (**n_k**), the epsilon parameter at time step k (**epsilon_k**), and the direct and quadrature axis currents at time step k (represented as **id_k** and **iq_k**). To offer historical context—which might be crucial in catching fleeting events like inverter dead times or interlocking periods during vector transitions—the index of the previous vector (**n_1k**) was included as well. The anticipated values of the direct and quadrature axis currents at the next time step (**id_k1** and **iq_k1**) are the model's outputs.

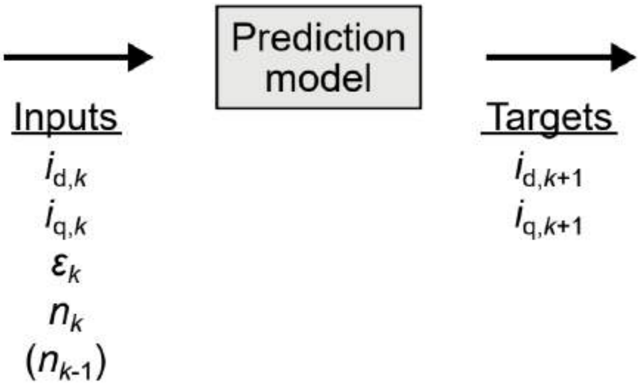


Figure 5. approach 1 architecture (Hanke et al.,2020)

The second approach describes several models created, each devoted to a certain operating vector **n_k** as shown in **Figure 6**. In real-time operation, these specialized models are switched between dynamically using the vector index (**n_k**). The inputs are identical to first model except that the prior vector index is not included. Independent of each other, each model is calibrated to forecast the direct and quadrature axis current values at the subsequent time step.

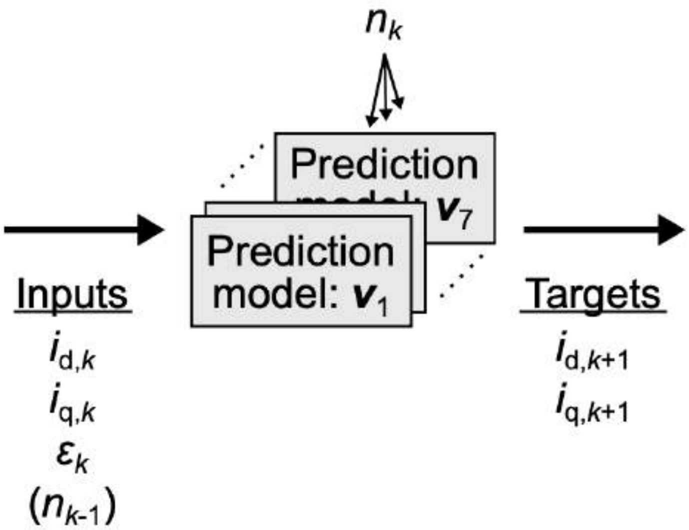


Figure 6. approach 2 architecture (Hanke et al.,2020)

In the third approach, we constructed a series of models that each address the transitions between vectors, the third variant increases the granularity of the modelling method as shown in **Figure 7**. A total of forty-nine models are built, each representing a distinct transition from **n_k1** to **n_k**, the vector states. These models are specifically designed to forecast how the motor will behave during vector transitions while taking the present state and the epsilon parameter into account. The outputs are the expected direct and quadrature axis currents for the upcoming time step, same as in the earlier models.

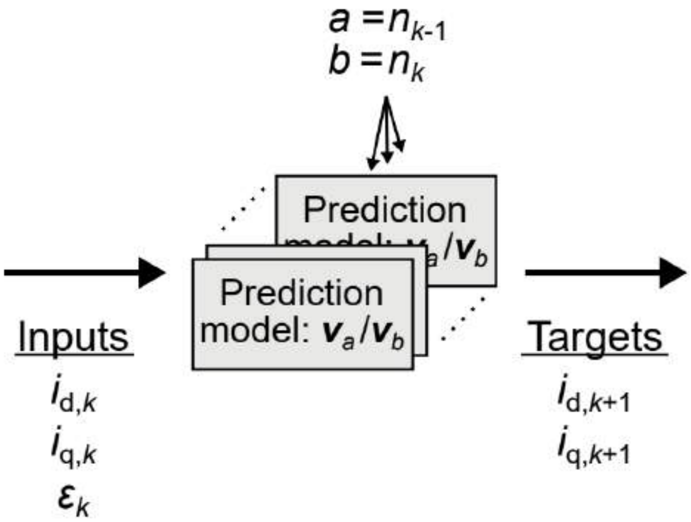

$a = n_{k-1}$
$b = n_k$
Prediction model: $\boldsymbol{v}_a/\boldsymbol{v}_b$
Prediction model: $\boldsymbol{v}_a/\boldsymbol{v}_b$
Inputs
$i_{d,k}$
$i_{q,k}$
ε_k
Targets
$i_{d,k+1}$
$i_{q,k+1}$

Figure 7. approach 3 architecture (Hanke et al.,2020)

Results and Discussion

Our primary objective was to evaluate the performance of different models and selected approaches in terms of their predictive accuracy for the direct and quadrature axis currents at the next time step (**id_k1** and **iq_k1**).

Approach 1: Single Predictive Model

We trained and tested several models including linear regression, ridge regression, lasso regression, decision tree, and random forest. The performance metrics for these models are summarized in Table 2. The random forest model outperformed all other models with a train MSE of 4.22991 and a test MSE of 27.95, indicating high predictive accuracy. The decision tree model also showed strong performance with a test R^2 of 0.97696. Linear regression and ridge regression exhibited identical performance, with a test R^2 of 0.85664, while the lasso regression model had a slightly lower test R^2 of 0.85501.

Table 2. model evaluation metrics for different models in approach 1

Model	**Train MSE**	**Test MSE**	**Train R^2**	**Test R^2**
Linear Regression	578.187	578.279	0.85651	0.85664
Ridge Regression	578.187	578.279	0.85651	0.85664
Lasso Regression	584.774	584.824	0.85486	0.85501
Decision Tree	0	93.4138	1	0.97696
Random Forest	4.22991	27.95	0.99895	0.9931

Figure 8 shows the comparison between actual and predicted values for different models in Approach 1. The random forest model demonstrated the closest fit to the actual values, reinforcing its superior performance.

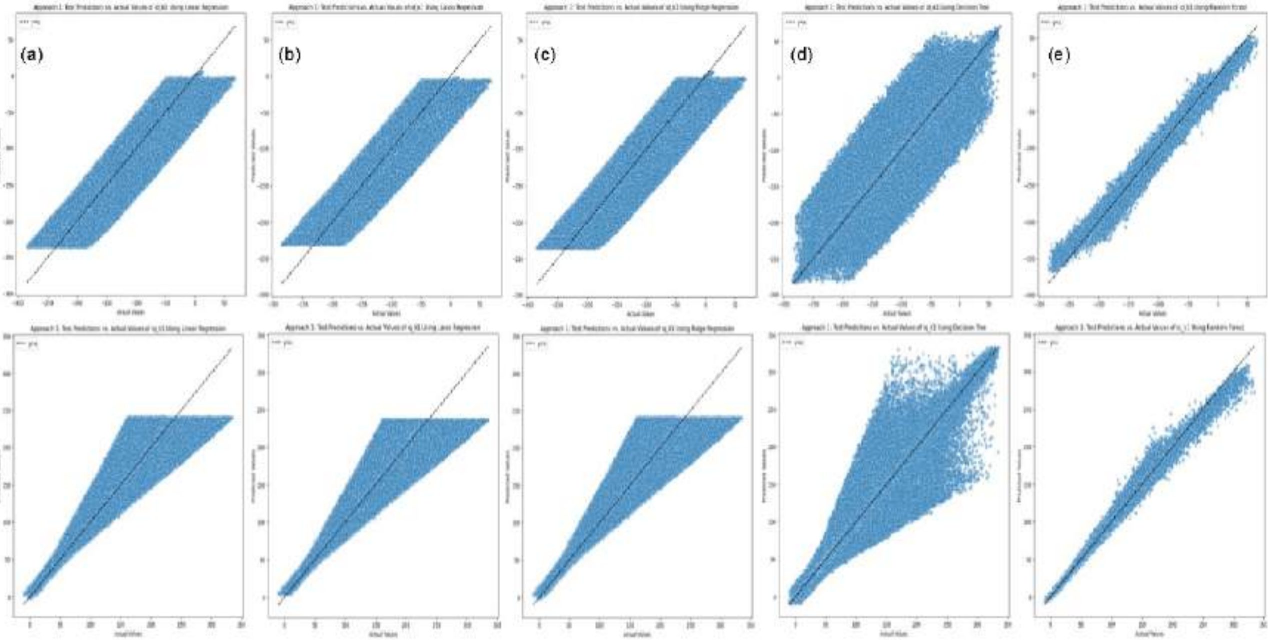


Figure 8. comparison between actual and predicted values for different models (a) linear regression, (b) lasso, (c) ridge, (d) decision tree, (e) random forest (approach 1). The top row represents **id_k** test results, and the bottom row represents **iq_k** test results

Approach 2: Models for Each Elementary Vector

In the second approach, we developed models dedicated to each elementary vector (**n_k**). Table 3 illustrates the evaluation metrics for linear regression and random forest models for each vector. The random forest models consistently achieved near-perfect R^2 values across all vectors, with the test R^2 ranging from 0.99883 to 0.99949. The linear regression models, while still performing well, showed more variability in their predictive accuracy with test R^2 values ranging from 0.85043 to 0.99933.

Table 3. model evaluation metrics for each elementary vector using linear regression and random forest in approach 2

Model	**n_k**	**Train MSE**	**Test MSE**	**Train R^2**	**Test R^2**
linear	1	2.83173	2.82852	0.99933	0.99933
linear	2	492.998	490.94	0.87525	0.87547
linear	3	450.402	450.953	0.88535	0.88514
linear	4	592.188	592.512	0.85305	0.85246
linear	5	599.946	599.805	0.85045	0.85043
linear	6	439.467	439.563	0.88791	0.88783

linear	7	450.254	450.742	0.88548	0.8852
RF	1	0.29833	2.14311	0.99992	0.99949
RF	2	0.63315	4.41165	0.99983	0.99887
RF	3	0.67179	4.59419	0.99982	0.99883
RF	4	0.64448	4.46586	0.99983	0.99886
RF	5	0.65276	4.5425	0.99983	0.99884
RF	6	0.58873	4.03443	0.99985	0.99897
RF	7	0.65554	4.50821	0.99983	0.99885

Figure 9 depicts the comparison between actual and predicted values for each elementary vector using random forest in Approach 2. The models accurately captured the variations across different vector states.

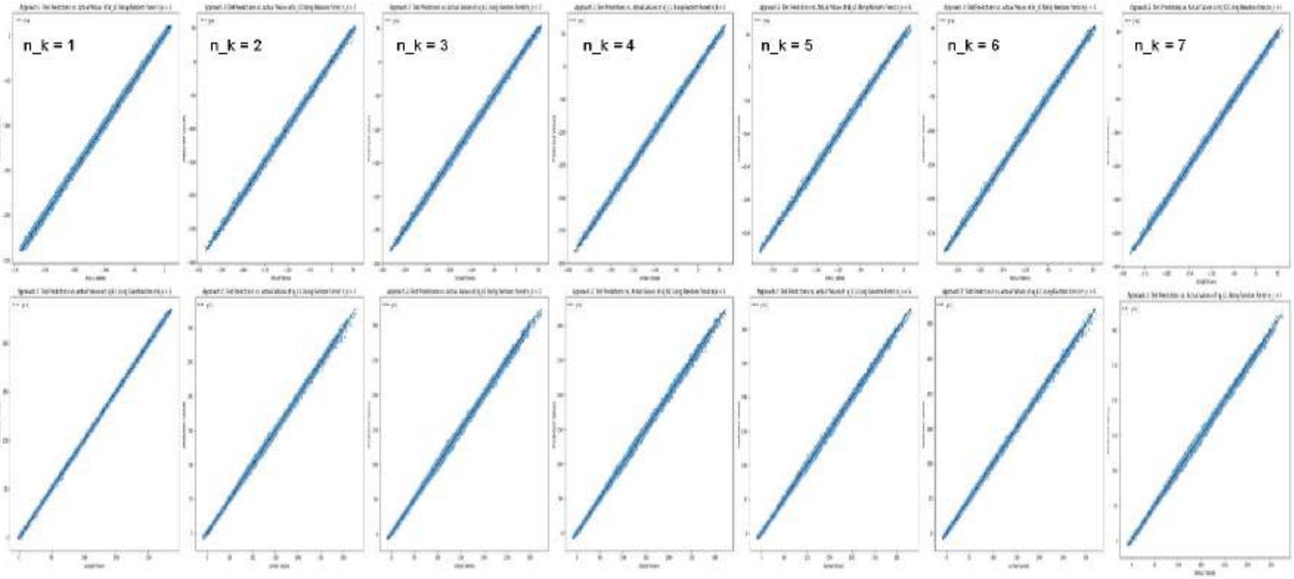


Figure 9. comparison between actual and predicted values for each elementary vector using random forest (approach 2). The top row represents **id_k** test results, and the bottom row represents **iq_k** test results

Approach 3: Models for Pairs of Vectors

In the third approach, we constructed models for each pair of vectors (**n_1k** to **n_k**). We showed the results only for pairs, where **n_k** = 1. The evaluation metrics are summarized in Table 4. The random forest models demonstrated exceptional performance, with test R^2 values close to 0.999 across all vector pairs. Linear regression models, while effective, showed more variability with test R^2 values ranging from 0.86197 to 0.99978.

Table 4. model evaluation metrics for pairs of vectors using random forest in approach 3

		Train			
Model	**Pair**	**MSE**	**Test MSE**	**Train R^2**	**Test R^2**
linear	1, 1	0.93695	0.93399	0.99978	0.99978
linear	1, 2	532.451	537.513	0.89041	0.8905

linear	1, 3	500.633	501.824	0.89784	0.89645
linear	1, 4	651.858	661.21	0.86512	0.86543
linear	1, 5	651.111	654.422	0.86621	0.86197
linear	1, 6	506.413	505.915	0.89414	0.89347
linear	1, 7	512.711	509.259	0.89343	0.89436
RF	1, 1	0.05789	0.40833	0.99998	0.9999
RF	1, 2	0.65308	4.47329	0.99986	0.99908
RF	1, 3	0.67331	4.52693	0.99986	0.99906
RF	1, 4	0.68135	4.59641	0.99985	0.99905
RF	1, 5	0.66468	4.49001	0.99986	0.99906
RF	1, 6	0.60016	4.14714	0.99987	0.99912
RF	1, 7	0.67052	4.52336	0.99986	0.99906

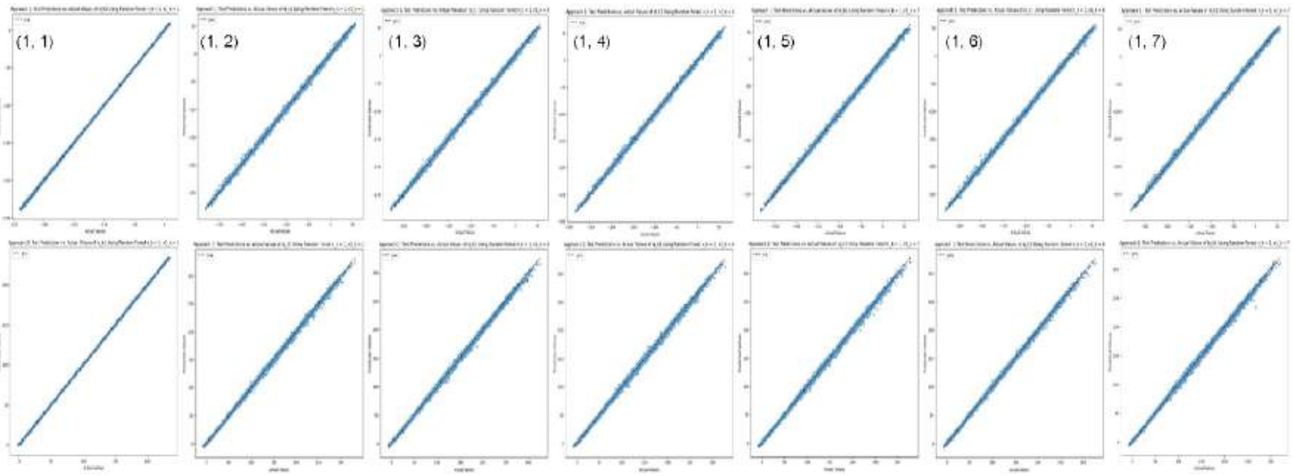


Figure 10. compare actual and predicted values for pairs of vectors using random forest (approach 3). The top row represents **id_k** test results, and the bottom row represents **iq_k** test results

Figure 10 compares the actual and predicted values for pairs of vectors using the random forest model in Approach 3. The results demonstrate the model's effectiveness in accurately capturing the transitions between vector states.

I n conclusion, the random forest model consistently outperformed other models across all three approaches, demonstrating high predictive accuracy and robustness. These findings underscore the potential of using machine learning models, particularly ensemble methods like random forests, to predict the behaviour of inverters and electric motors, thereby enhancing their efficiency, reliability, and operational lifespan.

Conclusions

This study highlights the importance of machine learning in predicting the behaviour of inverters and electric motors in electric car powertrains. By using large datasets, we can improve the efficiency, dependability, and lifetime of these systems. This is consistent with earlier studies that highlight the necessity of predictive maintenance and operational optimisation in electric vehicle technology.

Our findings rely on the groundwork laid by Hanke et al. (2020a, 2020b), who supplied a comprehensive dataset for studying the electrical behaviour of inverters and motors. While their research primarily focused on data collecting and preliminary model construction, our work goes beyond that by implementing and testing different machine learning models, allowing for a more complete comparison of modelling approaches.

The dataset employed in this study, which included around 40 million samples, proved invaluable for developing solid predictive models. Like previous research, our findings highlight the need of big, high-quality datasets in constructing accurate machine learning models. However, our focus on the specific application to electric vehicle drive trains, as well as the comparative examination of various modelling methodologies, yields new insights regarding dataset utilisation.

We tried and tested three different modelling approaches: single predictive models, models for each elementary vector, and models for pairs of vectors. The random forest model consistently outperformed other models, with good prediction accuracy and robustness. This finding is consistent with earlier research that advocates for ensemble approaches in predictive modelling, but our study adds a fresh application and comparative analysis in the context of electric car drive trains.

The findings show that machine learning models, specifically random forests, can properly predict the behaviour of inverters and motors, resulting in increased operational efficiency and dependability. While prior research has demonstrated the potential of machine learning in similar circumstances, our extensive comparison of modelling methodologies and application to electric vehicle drive trains make new contributions to the area.

Works Cited

Hanke, S., Wallscheid, O., & Böcker, J. (2020, March 16). *Data set description: Identifying the physics behind an electric motor -- Data-Driven Learning of the Electrical Behavior (Part I)*. arXiv.org. https://arxiv.org/abs/2003.07273

Hanke, S., Wallscheid, O., & Böcker, J. (2020a, March 13). *Data set description: Identifying the physics behind an electric motor -- Data-Driven Learning of the Electrical Behavior (Part II)*. arXiv.org. https://arxiv.org/abs/2003.06268

How do New Year celebrations influence consumer preferences, pricing acceptance, and purchasing decisions in the market? By Vaishnavi Thanneeru

1. Introduction

New Year's Eve, a universally acknowledged celebration, tends to bring in changes indicated by the well-known New Year's resolutions." This remark highlights New Year's Eve's cultural significance as an occasion for personal improvement. This festival is more than just a calendar event for people worldwide; it is a shared cultural tradition deeply embedded in communities. The passage from one year to the next encourages collective reflection, creating an atmosphere of renewal and hope. An observer to this popular custom is the tradition of setting New Year's goals, which demonstrates an overall desire for self-improvement and growth. As people gather together to pursue personal goals, tradition becomes a cultural force, shaping norms and influencing behaviour. In this way, the celebration of New Year's Eve serves as a cultural touchstone, starting a shared commitment to positive change and marking a moment in time when individuals collectively embrace the prospect of a fresh start.

More than one million people attended the New York City ball drop, as reported by the official website of the city of New York for the 2022 New Year's Eve celebrations, as well as in Paris to watch the fireworks display, according to The Guardian. With these figures, it can be inferred that a substantial portion of our population is influenced by the grand and extensive New Year's Eve celebrations. This suggests that NYE holds significant sway over purchasing decisions, as it is celebrated by many people. As these festivities approach, the economy undergoes sudden shifts, and individuals' judgment in purchasing decisions becomes uncertain. While some opt to cut back on expenses, others exhibit acceptance, leading to some of their most significant expenditures during this period.

When the entire world is in a celebratory mood, participating in hundreds of different traditions, taking a break, and bursting open champagne bottles, there is likely a change in their atmosphere, mood, and spirit. This research intends to understand if the festive atmosphere, pleasant mindset of clients, and energetic timing can have an enormous impact on purchasing decisions. Can these celebrations change their preferences, like where they want to be at this time of year geographically? Can it impact their pricing acceptance, like finally purchasing their favorite product that they refrained from due to high pricing? Will their current mood push them into making more purchasing decisions, increasing or decreasing their level of shopping? This question evaluates how celebrations and traditions can impact their choices.

To ensure the accuracy of these ideas, this research will review sources on behavioral biases to identify the truth behind consumer responses to pricing during celebratory periods, by delving into established behavioral biases to uncover the underlying psychological mechanisms that shape consumer decision-making. Also, by examining broader economic trends during the New Year period, exploring whether there are patterns in consumer spending, inflation rates, or market behaviours during this time and how these factors influence pricing and purchasing decisions. Additionally, by creating a survey to send around asking participants about their celebrations on

New Year's Eve and finding out if there is a connection between NYE and consumer preferences and acceptance, we can identify if it influences their purchasing decisions."

This research opens many gateways for researchers and marketers. Understanding the ideology between purchasing decisions and widely celebrated holidays such as New Year celebrations can inform researchers and marketers about the market trends associated with the occurrence of the holidays. It can give them a strong idea of how to make predictions about the expectations of their consumers.

Companies can also adapt their pricing strategies based on insights from the study, understanding how consumer decisions are influenced by both cognitive biases and pricing dynamics during the holiday season and evaluating the intricate connection, including the dynamics of price elasticity of demand and notable market swings throughout festive seasons.

2. Literature Review

As mentioned above, behavioral bias has an important connection with decision-making. Behavioral biases are irrational beliefs or behaviors that can unconsciously influence our decision-making process. In this section, I will explore a number of biases and examples of their effects, linking them back to the case study I am exploring. Philippos J. Richter suggested that happy moods result in quick, spontaneous thinking, while sad moods induce customers to think further and prolong purchasing decisions. Research conducted by Yahoo recruited 600 adults to fill in a week-long smartphone diary that tracked their moods. This revealed that when consumers are upbeat, they are 24% more receptive. Winner of the Nobel Prize Daniel Kanhenman, further strengthened the hypothesis that when people are in a positive mood, it sends signals of safety to their brain and, therefore, reduces the need to think critically. In turn, people are more likely to absorb ad messages. This concludes that a happier mood increases the chances of purchasing products. Therefore, we can anticipate from the fact that people who are in a more positive mood around New Year's Eve and were looking forward to the upcoming year will probably have made more purchases than those who disliked the event.

As indicated by the study conducted by C Z Malatesta and M Kalnok, emotional experiences play a significant role in the lives of individuals across different age groups. One particularly noteworthy observation is that conventional constraints on the overt display of affect seem to have less impact on older individuals compared to their younger counterparts. This implies that older adults may navigate and express their emotions with a greater degree of freedom, challenging societal norms surrounding emotional expression. This is further communicated on later in the discussion

To explore biases further and understand consumers' decision-making around New Year's Eve, a nudge experiment would be beneficial. A "nudge" experiment typically refers to a study or trial that applies principles of behavioral economics to influence people's behaviour in a subtle and positive way by building two surveys: an experimental survey and a control survey. This approach allows us to make causal inferences about the impact of New Year's Eve—in this case, the nudge—by comparing the outcomes between the experimental and control groups.

3. Method

As discussed in the literature review, nudge experiments are a very useful method. By comparing the outcomes of the experimental group with the control group, we can confidently attribute any observed changes in behavior to the nudge.

This survey performed for the study had the nudge of a New Year's Eve celebration. The experimental survey included questions concerning holiday scenarios and activities, while the control survey avoided the New Year's Eve issue as much as possible. For example, a scenario was given for the experimental group; it was asked: Imagine it is the end of the year, a few days away from New Year's Eve, and you come across a pair of shoes that you really love, but the price is higher than your budget. What would you do in this situation? In the control group, it was asked: It's a regular day, and you come across a pair of shoes that you really love, but the price is higher than your budget. What would you do in this situation? In both groups, they were given two options: I would purchase them or I would not purchase them.

A link was produced where both surveys were attached; using this link, which directs users to one of the two Google forms, close to half of the clickers get the experimental survey, as the other half get the second survey, to create an equal number of responses for both forms in the same social area. The form was split into multiple age groups and then sent on group chats of each of these age groups to diversify the responses as much as possible **(mention nimble) (data cleaning and all.)**

4. Results

The survey acquired a total of 361 results. 189 responses in the experimental group and 172 in the control group. In both forms, there were a higher number of responses from the 39-50 age group, with higher incomes. The survey also offered knowledge on mood, with 65% concluding that they were happy when they purchased, 5.8% responding when they were sad, and the rest, 29.2%, had other responses, for example, mixed, both and frustrated. Most people are aware that they shop when they are happy; this is common through both surveys as also suggested by Philippos J. Richter.

When asked, “During which time of the year do you usually make the most purchases? and why?” 131 of 361 participants said December, and responses came in such as: holidays, the last quarter of the year, or the start of winter, with reasons such as, New Year gifts for family and friends or travel time. This shows that most people spend December on holidays and New Year's Eve-related events, such as travelling.

A question was asked in both surveys whether they would purchase a pair of sneakers with a nudge, on a regular day or around New Year’s celebrations.

Sneaker scenario: would purchase percentage in the control and treatment groups, according to age.

Age Group	Would purchase % (control Would purchase % (treatment group) group)

Under 18	36.8% 48%
18-28	66% 42.8%
29-38	45.5% 35.3%
39-50	37.8% 32.6%

51+ 50% 29.6%

A connected question was asked in the sneaker scenario question. “If you would purchase, how much in percentage above your budget would you go?” The below table splits the data according to income

Income Group	Cost increase willing to pay Cost increase willing to pay % for control group (mean % for treatment group (mean value) value)
Not earning	21.7% 25%
Below 20k	26.4% 29%
20-60k	35.5% 22.5%
60-90k	30.9% 27.2%

Above 90k 23.6% 25.23%

5. <u>Discussion</u>

Age groups and willingness to purchase above budget:

Age groups all have different responses to the willingness to purchase throughout the year, near New Years Eve, or during the holidays, This evaluation signifies the reasons

Among individuals under 18, the experimental group's higher purchasing willingness (48%) compared to the control group (36.8%) suggests a heightened receptivity to New Year celebrations. This aligns with research by C Z Malatesta and M. Kalnok, indicating that the younger population, being more emotional, may find the festive and gift-giving nature of the holiday season particularly compelling, influencing their purchasing decisions positively.

In the 18–28 age range, the conclusion of the control group (66%) showcasing a higher willingness to purchase than the treatment group (42.8%) raises questions about the complex factors shaping decisions. This demographic, largely composed of college students with tighter incomes, may prioritise financial issues over seasonal influences, contributing to the trend identified

For both the 29–38 and 39–50 age groups, a decline in purchasing willingness compared to the younger population is evident. However, the control groups in both age ranges continue with higher percentages (45.5% and 37.8%, respectively) than their experimental groups, emphasising the continuance of traditional habits. This suggests that while New Year celebrations may have a diminishing impact, on a regular basis, there isn't a fall; this can mean this age group purchases when necessary or immediately when there is a product drop.

In the 51+ age group, the control group's substantially higher willingness to purchase (50%) compared to the treatment group (29.6%) indicates a decreased susceptibility to New Year-related factors among older individuals. Well-established shopping patterns and a potentially more conventional approach contribute to the stability of the control group's response.

This evaluation indicates the different features and aspects of consumers' willingness to purchase during the holiday season, emphasizing the interplay of emotions, financial considerations, and established habits across different age groups.

Different income levels and the percentage they are willing to go above their budget:

Among those not earning, the control group demonstrates a willingness to outreach from their budget by 21.7%, while the treatment group exhibits a slightly higher inclination at 25%. This could indicate that individuals with no income may be more adaptable in their budget constraints, and the treatment group, exposed to different influences, may feel more liberated to exceed their financial limits. For example, individuals in this category might be students relying on allowances or part-time work, and the treatment group may include those swayed by peer trends or marketing strategies, especially since we have identified that the younger population is more swayed by their emotions.

In the income bracket below 20k, both groups express a desire to surpass their budgets, with the treatment group showing a slightly higher readiness at 29%, compared to 26.4% in the control group. This suggests that individuals with lower incomes may be more susceptible to external influences, such as the allure of special occasions like New Year celebrations. For instance, someone in this income range might be enticed to make a purchase beyond their budget for a special New Year's event. They could be under the mindset to splurge on special occasions as they have more suitable purchasing decisions on a regular basis.

Surprisingly, in the 20–60k income range, the control group indicates a significantly higher willingness to overspend (35.5%) compared to the treatment group (22.5%). This unexpected result could imply that individuals in the control group, despite exposure to New Year-related factors, might be more driven by established spending habits or personal preferences. For example, a person in this income bracket might prioritise a specific brand or style over seasonal

Within the 60-90k income range, the control group is willing to exceed their budget by 30.9%, while the treatment group shows a slightly lower inclination at 27.2%. The difference is less pronounced compared to lower income brackets, indicating a more marginal impact of external factors. This suggests that individuals in this income range may have a higher degree of financial stability, making them less susceptible to external influences on their spending behaviour, resulting in the same budgeting throughout the year.

Groups in the higher income bracket have a relatively lower willingness to exceed their budget, with the control group at 23.6% and the treatment group at 25.23%. Individuals with higher Incomes often have more financial stability and may exhibit a lower susceptibility to external influences. For instance, someone earning above $90k might prioritise budget discipline over succumbing to seasonal sales; however, they love to splurge around New Year due to their ability to afford products with financial security.

6. <u>Conclusion:</u>

In conclusion, the comprehensive analysis of consumer behaviour during the holiday season, specifically focusing on New Year celebrations, reveals distinct patterns across age and income groups. Younger individuals, particularly those under 18, exhibit a heightened receptivity to seasonal influences, aligning with their emotional responses to festive occasions. Surprisingly, the 18–28 age group, primarily composed of college students, prioritizes financial considerations over seasonal influences, suggesting the need for marketers to understand the unique priorities of this demographic. Older age groups, notably 29–38 and 39–50, show a decline in purchasing willingness compared to their younger counterparts, emphasizing the persistence of established habits.

Income-wise, individuals with lower incomes display a higher inclination to exceed budgets during the holidays, while those in the 20–60k bracket unexpectedly exhibit a greater willingness, potentially influenced by personal preferences. Higher-income individuals display a more stable response. Marketers should leverage these insights to tailor strategies that resonate with the subtle preferences and financial considerations of each demographic, emphasising emotional appeals for the younger population, focusing more on them during the holidays and focusing on brand loyalty for those with established spending habits. Businesses should consider targeted promotions during peak spending periods, especially in lower income brackets.

Further research might look at specific factors that contribute to the surprising findings in the 20–60k income range, where the control group is substantially more prepared to overspend on trainers than the treatment group. Understanding why, despite being exposed to New Year 's-related events, individuals in this income group may prioritise established spending habits or personal preferences above seasonal influences could reveal significant insights. This inquiry could look into the effect of brand loyalty, perceived product value, or past purchasing experiences in determining customer behaviour at this income level. Furthermore, investigating

the dynamics of peer influence and marketing methods on the purchasing decisions of those with no income could provide a better understanding of their increased budget flexibility.

7. <u>Limitations</u>

The survey has its own limitations: as in the sneaker scenario question, we can question the aspect of if some people are not that interested in sneakers to go over the budget to purchase the product what if more people are willing to purchase fragrances above budget, with a higher percentage above budget as well. The survey also had a significantly higher number of responses from the 39-50 age group. A skewed distribution toward the 39-50 age group in the survey may result in potential age-related bias.

Works Cited

https://www.nyc.gov/events/new-years-eve-in-times-square/279186/1#:~:text=As%20the%20famous%20New%20Year's,hope%20for%20the%20year%20ahead.

https://www.theguardian.com/world/video/2023/jan/01/more-than-1m-people-gather-in-paris-to-watch-new-years-eve-fireworks-display-video#:~:text=Paris-,More%20than%201m%20people%20gather%20in%20Paris%20to,Year's%20Eve%20fireworks%20display%20%E2%80%93%20video&text=More%20than%20a%20million%20people,two%20years%20of%20Covid%20cancellations.

Malatesta CZ, Kalnok M. Emotional experience in younger and older adults. J Gerontol. 1984 May;39(3):301-8. doi: 10.1093/geronj/39.3.301. PMID: 6715807. https://pubmed.ncbi.nlm.nih.gov/6715807/

Impact of Cooking Methods on Nutritional Content of Shrimp and Tofu By Arushi Chatterjee

Abstract

Food is the primary source of energy and nutrients required for human survival. Food can be consumed raw and cooked, too. Food is essential for the growth of an organism as it contains proteins, carbohydrates, fats, vitamins, and minerals as nutrients. Proteins are large biomolecules made up of one or more amino acid chains. Within living things, proteins carry out a wide range of functions. Carbohydrates are sugar molecules that the human body converts to glucose. The body's tissues, organs, and cells primarily utilize glucose as a source of energy. Fats are a primary storage form of energy in the body, which act as a fuel source. Thus, it is essential to include a moderate amount of fats in a balanced diet. The current study highlights the effect of boiling and sauteing on the carbohydrate, protein, and fat content of shrimp and tofu. Protein estimation was conducted using the Kjeldahl method, Carbohydrate estimation using the moisture determination method, and fat estimation using the Soxhlet extraction method. It was observed that boiling the shrimp showed a substantial decrease in carbohydrates (-78.7%), fats (-55.84%), and protein (-28.65%) content. Similarly, boiling tofu also showed a substantial decrease in the carbohydrate (-44.14%) and protein (-4.44%) content. However, the content of fats after boiling the tofu had increased (+94.11%). Contarily, for both the samples, sauteing increased the content of fats, carbohydrates, and protein, which could be due to the oil used for cooking.

Keywords: Nutrients, cooking methods, shrimp, tofu.

Introduction

Nutrition is essential for our health, development, and general well-being. It provides the energy we need to function and thrive. Nutrients allow our bodies to perform all the necessary biological processes. Good nutrition strengthens our immune systems and battles chronic diseases such as diabetes, heart disease, and cancer. Additionally, it helps us maintain a healthy weight and promotes better mental health. Therefore, understanding nutrition and healthy dietary choices is crucial to leading a healthy life. The Centers for Disease Control and Prevention (CDC) state that those who follow good eating habits live longer and are less likely to suffer from conditions like obesity, diabetes, and cardiovascular illnesses (CDC, 2023). As per World Health Organization (WHO), a healthy diet protects humans against many chronic noncommunicable diseases and provides essential nutrients for survival. According to the National Health Service (NHS), a balanced diet should include meals based on high-fiber starchy foods such as potatoes, bread, rice, or pasta, five portions of a variety of fruits and vegetables per day, diary or dairy products, protein in the form of meat, fish, eggs, pulses, beans, and so on, unsaturated fats in small amounts, and plenty of fluids such as water (NHS, 2022).

There are many different cooking methods which can be divided into overarching categories of wet and dry cooking methods. As per the State Journal-Register of Springfield, Illinois, some cooking methods that fall under the category of dry heat cooking are baking, grilling, broiling, sauteing, deep frying, and pan frying (Writer S, 2012). On the contrary, poaching, simmering, boiling, steaming, sou vide, and braising are some cooking methods that come under moist heat cooking.

While cooking some foods is essential to make them safe for consumption, cooking other foods depletes their nutrient density. Beans, for example, are not safe to eat if they are raw, but vegetables and fruits have their highest nutrient density when consumed raw.

Different nutrients react differently to various cooking methods. Proteins, for example, are a macronutrient that is only sometimes lost during cooking. However, nutrients like vitamins, particularly water-soluble vitamins such as vitamin C and vitamin B, are often lost during moist heat (wet) cooking processes. Moreover, wet cooking techniques do not use any additional fats. Therefore, the fat content does not increase; however, dry cooking techniques like frying often use oil, which leads to an increase in the fat content (TNAU, 2015).

Seafood is an important constituent of the human diet. The consumers are looking at sea food as health food due its fatty acid profile. Shrimp is one of the world's most popular shellfish and is a part of almost every nation's traditional meal (Dayal J, 2012). The nutritional value of shrimp is its relatively lower lipid content (~ 1%), the Daily Value % (DV %) of 100 g shrimp for an adult human is 75%, 70%, and 35% for eicosapentaenoic acid + docosahexaenoic acid, essential amino acids (methionine, tryptophan, and lysine) and protein respectively (Dayal, J, 2013). Additionally, the lower atherogenic (0.36) and thrombogenic (0.29) indices of shrimp show their cardio-protective nature. It is fairly low in calories and provides a high amount of protein, healthy fats, and a variety of vitamins and minerals. Shrimp is high in cholesterol, but it also contains omega-3 fatty acids that have been shown to promote cardiovascular health (Healthline, 2022).

Tofu, a popular superfood, is made from condensed soy milk, which is similar to how cheese is made. Tofu is low in calories but high in protein and fat. It also contains many vitamins and minerals, including calcium and manganese. However, tofu includes antinutrients including trypsin inhibitors and phytates (Healthline, 2018). Soaking, sprouting, or fermenting soybeans before manufacturing tofu reduces antinutrient levels. All soy foods, including tofu, contain isoflavones, which are thought to be the primary cause of tofu's health advantages (Healthline, 2018). Tofu and other whole soy meals may enhance a variety of heart health indices (Healthline, 2018).

According to a research conducted by (Tyagi et al. 2015) investigating the impact of different cooking methods (baking, boiling, steaming, microwave cooking and pressure cooking) on a variety of products (such as corn, wheat, chickpea, rice, egg, potato, tomato, ladyfinger, spinach, capsicum), revealed a difference in protein concentration and vitamin C percentage before and after cooking (Tyagi et al. 2015).

Since there is not much published research in this area, hence the current research aims to investigate the effect of cooking methods like boiling and sauteing on the nutritional content of shrimp and tofu and identify the difference in the protein, fat, and carbohydrate content before and after cooking.

Materials and Methods

Materials used

200 grams of packaged Shrimps (ITC Masterchef) and Tofu (Mooz Tofu) respectively, were procured from the local market in Gurgaon.The initial composition of the proteins, fats and carbohydrates of the raw samples were identified from the package label.

Cooking techniques used

The cooking techniques under investigation were Boiling and Sauteing.

Sample preparation

Sauteing Shrimp: The shrimp were defrosted and properly cleaned to remove any remaining ice or pollutants. The defrosted shrimp were then sautéed with 1 tablespoon cold-pressed sesame oil. This process was carried out in a pan over high heat for five minutes. After sautéing, the shrimp were quickly chilled to stop the cooking process.

Sauteing Tofu: A block of tofu, weighing 100g, was cut into cubes with each side measuring 1 cm in length. The tofu cubes were then sauteed in 1 tablespoon of cold-pressed sesame oil. This was performed in a pan over high heat for 10 minutes. Following the sauteing process, the tofu cubes were removed from the pan and allowed to cool immediately to stop further cooking.

Boiling Shrimp: 100g shrimps were first defrosted and thoroughly washed.The defrosted shrimp were then placed in 1.5 Liters of boiling water (100ºC) for 5 minutes. The samples were immediately placed in ice cold water to halt the cooking process.

Boiling Tofu: A 100g block of tofu was cut into cubes with a side length of 1 cm. The tofu cubes were boiled in 1.5 liters of boiling water (100ºC) for 10 minutes. The samples were immediately placed in ice cold water to halt the cooking process.

Sample Analysis

The processed samples were analyzed to estimate the carbohydrate, fat, and protein content using standardized methods:

Carbohydrate Analysis: Conducted using the moisture determination method (Bija S, 2022) in accordance with IS 1656.

Fat Analysis: Conducted using the Soxhlet extraction method (Bija S, 2022) in accordance with the AOAC method 922.06.

Protein Analysis: Conducted using the Kjeldahl method (Bija S, 2022) in accordance with IS 7219

Results and Discussion

Nutrient	Packaged shrimp (per 100g)	Boiled Shrimp (per 100g)	Sauteed Shrimp (per 100g)	Packaged Tofu (per 100g)	Boiled Tofu (per 100g)	Sauteed Tofu (per 100g)

Carbohydrate	1.5	0.32	3.06	4.5	1.79	9.64
Fat	2.4	1.06	7.21	2.89	5.61	12.56
Protein	17.1	12.2	17.56	14.19	13.56	17.06

Table 1: The nutritional content of Shrimp and Tofu before & after boiling and Sauteing

Impact of Sauteing on the Nutritional content of Shrimp and Tofu

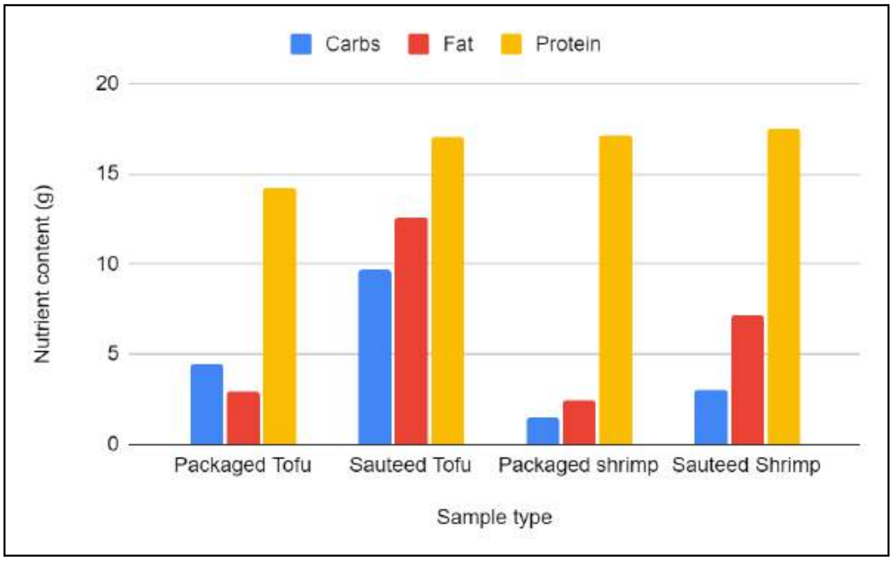


Fig 1: Effect of Sauteing on the Nutritional content of Shrimp and Tofu

Fats: Table 1 & Fig 1. shows that the fat content increased in sauteed shrimp & tofu as compared to the raw samples. Shrimp naturally contains some fats, primarily in the form of omega-3 and omega-6 fatty acids. Sautéing shrimp in oil adds additional fats to it.The type of oil used will also influence the fatty acid profile of the dish. Tofu is relatively low in fat, but sautéing it in oil will increase its fat content as it absorbs some of the oil during cooking. The type of oil used will also influence the fatty acid profile of the dish.

Carbohydrates: From Table 1 & Fig 1. it can be seen that there has been a slight change in the carbohydrate content of sauteed shrimp and tofu. Shrimp are very low in carbohydrates, so there is minimal impact on their carbohydrate content when sautéed. Tofu contains a small amount of carbohydrates, primarily in the form of sugars and fiber. Sautéing tofu may slightly reduce its water content, concentrating the carbohydrates, but the overall impact on carbohydrate content is minimal.

Protein: From Table 1 & Fig 1. it can be seen that there has been a slight change in the fat content of sauteed shrimp and tofu.Shrimp is a rich source of protein. Sautéing shrimp may cause some denaturation of the proteins due to the high heat, but the overall protein content remains relatively unchanged.Tofu is a plant-based protein source. While some protein denaturation may occur during cooking, the protein content of tofu remains largely intact.

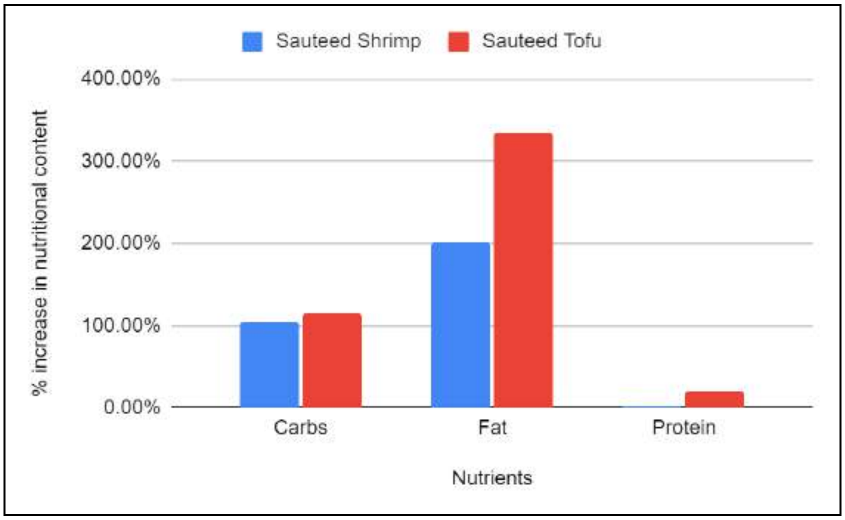


Fig 2: Percent increase in the nutritional content of Shrimp and Tofu on account of Sauteing.
Fig 2. shows that sauteing the Tofu substantially increased the carbohydrates (+114.22%), fats (+3334.60%), and protein (+20.22%) content. Similarly, sauteing shrimp also showed a substantial increase in the carbohydrate (+ 104%) and protein (+ 200.42%) content & protein (+2.69%) respectively.

Impact of Boiling on the Nutritional content of Shrimp and Tofu

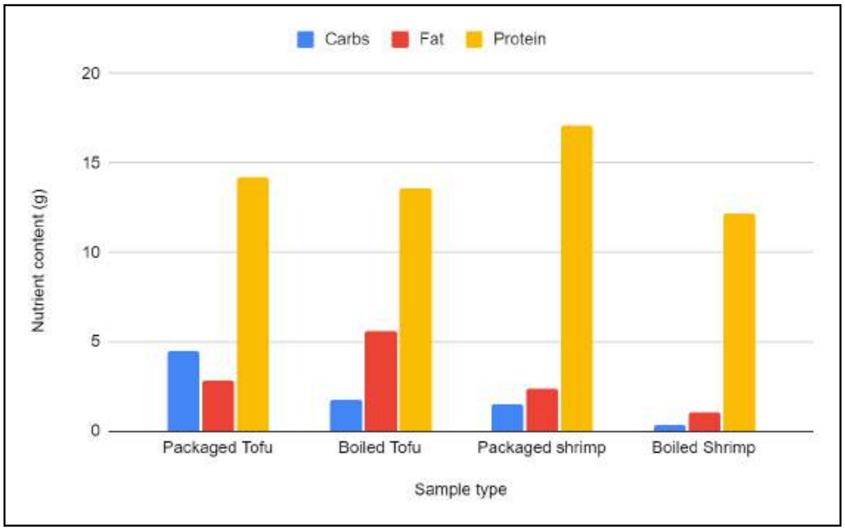


Fig 3: Effect of Boiling on the Nutritional content of Shrimp and Tofu

Fats: From Table 1 & Fig 3. It can be seen that the composition of fat has reduced in the shrimp. Boiling shrimp can lead to the loss of some fat content, especially if the shrimp is boiled without its shell. The fats in shrimp are susceptible to heat and can melt and leach out into the boiling water. Tofu is relatively high in fat, primarily unsaturated fats. Boiling tofu may cause some fat loss.However the fat content in Tofu has increased.

Carbohydrates: From Table 1 & Fig 3. it can be seen that there has been a significant change in the carbohydrate content of boiled shrimp and tofu.Shrimp contains minimal carbohydrates, primarily in the form of glycogen. Boiling may not significantly affect the carbohydrate content of shrimp. Tofu contains carbohydrates in the form of starch. Boiling tofu may soften the texture and increase the digestibility of carbohydrates, but it is unlikely to significantly alter the carbohydrate content.
Protein: From Table 1 & Fig 3. it can be seen that there has been a slight change in the protein content of boiled tofu, however there is a significant change in the protein content of boiled shrimp. Protein is a major component of shrimp. Boiling can cause denaturation of proteins, altering their structure and potentially making them more digestible. However, excessive boiling can lead to protein degradation and loss. Tofu is rich in protein, making it a valuable source of plant-based protein. Boiling tofu can cause some denaturation of proteins, similar to what occurs with shrimp. However, tofu proteins are generally more stable than those in animal products, so protein loss may be minimal.

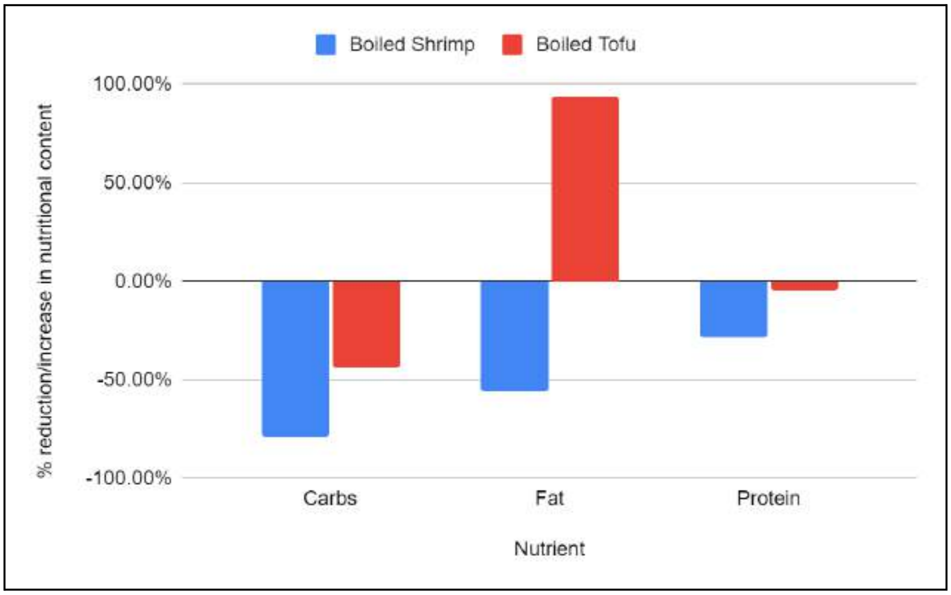


Fig 4: Percent increase & reduction in the nutritional content of Shrimp and Tofu on account of boiling

Fig 4. shows that boiling the shrimp substantially decreased carbohydrates (-78.7%), fats (-55.84%), and protein (-28.65%) content. Similarly, boiling tofu also showed a substantial decrease in the carbohydrate (-44.14%) and protein (-4.44%) content. However, the content of fats after boiling the tofu had increased (+94.11%).

Conclusion

India is home to a vast range of cultures, languages, and eating traditions, as seen by the prevalence of various cooking methods across the country. Cooking methods like boiling, and sauteing sometimes deplete essential nutrients. This research was undertaken to identify which cooking methods are the ones that lead to losses of essential nutrients. From the discussion of the results it can be concluded that the alternate hypothesis is true, there is a measurable impact of boiling and sauteing on the nutritional content of shrimp and tofu as proved by the results of the paper. The current study involves only using Shrimp and Tofu, however to reach a holistic conclusion, the study of more food materials including vegetables and meat-based foods is required. Moreover, more nutrients like Vitamins, Minerals content present in food can be investigated. A potential research question for future study could be to investigate the effect of frying, baking, microwaving , etc. on different food materials.

Works Cited

CDC. “Tips for Healthy Eating for a Healthy Weight.” *Healthy Weight and Growth*, 26 Feb. 2024, www.cdc.gov/healthy-weight-growth/healthy-eating/index.html.

NHS. “Eating a Balanced Diet.” *Nhs.uk*, 29 July 2022, www.nhs.uk/live-well/eat-well/how-to-eat-a-balanced-diet/eating-a-balanced-diet/.

Writer, Staff. “Culinary Foundations: Dry Heat vs. Moist Heat Cooking.” *The State Journal-Register*, 14 Mar. 2012, www.sj-r.com/story/news/columns/2012/03/14/culinary-foundations-dry-heat-vs/44262211007/.

Tamil Nadu Agricultural University (TNAU), 2015, https://agritech.tnau.ac.in/nutrition/pdf/cooking%20methods.pdf.

Dayal, J. Syama, A. G. Ponniah, and K. Ambasankar. "Shrimp as health food-Advisory fact sheet." Proteins 78: 76-4.

Dayal, J. Syama, et al. "Shrimps–a nutritional perspective." Current science (2013): 1487-1491.

Elliott, Brianna. “Is Shrimp Good for You? Nutrition, Calories & More.” *Healthline*, 13 Apr. 2022, www.healthline.com/nutrition/is-shrimp-healthy#high-in-cholesterol.

Petre, Alina. “What Is Tofu, and Is It Good for You?” *Healthline*, Healthline Media, 13 Dec. 2018, www.healthline.com/nutrition/what-is-tofu#nutrition.

Tyagi, S, et al. “Impact of Cooking on Nutritional Content of Food.” *DU Journal of Undergraduate Research and Innovation*, vol. 1, 2015, pp. 180–186, journals.du.ac.in/ugresearch/pdf-vol3/U18.pdf.

Bija, Stephanie, Novi Luthfiyana, and Anhar Rozi. "The effect of cooking process on nutritional composition of Lais Fish (Cryptopterus sp.)." IOP Conference Series: Earth and Environmental Science. Vol. 1083. No. 1. IOP Publishing, 2022.

Examining the Impact of Python Pals on Middle School Girls' Perceptions of Computer Science By Riya Hegde

Abstract

Python Pals is a program that offers free programming lessons to middle school girls with the goal of increasing interest in computer science. Ultimately, the program aims to empower women to pursue computer science fields.Historically, women played a crucial role in early computer science, especially during World War II, but they received little recognition and were later marginalized as the field became male-dominated. Today, women make up only 21% of computer science majors ("Changing the Curve"), a decline driven by gender stereotypes and cultural biases. To address this disparity in Hudson, Ohio, Python Pals offers monthly coding sessions focusing on the basics of Python. Each 3-4 hour session includes instruction, group projects, and friendly competitions to reinforce learning. The program has also expanded internationally with virtual lessons for girls in India. The goal is to build programming skills and challenge gender stereotypes in computer science. Ea\rly indicators indicate significant improvements in participants' coding abilities and increased interest in further studies and careers in the field.

Keywords

Analysis; Probability and Statistics; Computer Science; Gender Gap; Python Pals

Introduction

Computer science began as a field dominated by women. During World War II, the U.S. military hired women to program a machine that performed the calculations needed for the war. These women were the first in the field of computer science, but they received little recognition. As time went on, computer science was marketed primarily towards men, and the environment in many computer science jobs became hostile towards women. Gender stereotypes surrounding the field of computer science likely contribute to this number, with formal engineering education not having a place for women. Women were originally successful in the field; as gender stereotypes emerged, the number of women in computer science decreased.

Of every STEM field, the gender gap in computer science is most apparent (Sax et al.). These statistics can also be traced to the industry's disregard for women's needs. For example, Apple faced backlash in 2014 when their integrated health application did not include a menstrual tracking feature (Sax et al.). However, this gender gap can trace back to an earlier age. The description of computer science as a field that girls are less interested in discourages girls from pursuing the field (Cheryan et al.). Girls often envision computer scientists as white or Asian males with limited social skills and an interest in video games or science fiction. These stereotypes come from portrayals of computer scientists in books and movies—affecting overall perceptions of the field (Cheryan et al.).

Even the gender disparity within computer science majors does not account for the entire gender gap in the field; only 38% of women who major in computers go on to computer fields

(“The STEM”). This attrition rate underscores the challenges women face within the industry, from a lack of mentorship and role models to the pervasive gender bias that can create unwelcoming work environments.

Efforts to address these disparities have included initiatives to introduce girls to computer science at a young age, such as coding boot camps, school programs, and extracurricular activities aimed specifically at young women. Programs like Girls Who Code and Black Girls Code work to provide young girls with the skills and confidence to pursue careers in technology, fostering an early interest that can counteract the stereotypes and biases they may encounter later on. Python Pals is similar to these programs, but it implements a unique curriculum consisting of collaboration and competition.

Additionally, universities and companies are increasingly recognizing the need to support women in computer science through scholarships, internships, and networking opportunities. By providing resources and creating supportive communities, these efforts aim to retain more women in the field and help them advance to leadership positions.

This study aims to assess the impact of Python Pals on girls' perceptions of programming. Python Pals is a grassroots organization dedicated to combating gender stereotypes in computer science from a young age. The program offers free programming sessions to middle school girls in the Hudson, Ohio community, as well as to girls at Oakwood Indian School in Kundapura, India (Hegde).

Methods

Participants in the Python Pals program included girls in grades 5 through 9 from two different locations: Hudson, Ohio, and the Oakwood Indian School in Kundapura, India. Each in-person session in Hudson comprised 3 to 10 girls, while the virtual sessions for the Oakwood Indian School varied in size.

To assess the impact of the program, data was gathered using pre- and post-course surveys. The surveys were anonymous and designed to gauge the girls' perceptions of programming. Statements included in the survey were: “I am good at programming” and “I want to learn more about programming in the future.” Participants rated these statements on a Likert scale from 1 (strongly disagree) to 5 (strongly agree).

In Hudson, Ohio, each session began with the students completing the pre-course survey on their devices. The curriculum was delivered over two days. On the first day, which lasted 1.5 hours, students learned basic Python skills, focusing on strings, loops, and if statements. They joined a team on the Replit website to complete example problems both as a class and individually, with an emphasis on collaboration and peer learning. On the second day, also 1.5 hours, students worked on a group project to create a simple number guessing game, which fostered collaboration and practical application of their skills. Following this, they participated in a friendly competition, tackling ten programming challenges of varying difficulty. The winner of the competition received a $10 gift card to a local shop, which served to motivate and engage the participants. After the competition, students completed the same survey to measure changes in their perceptions of programming.

At the Oakwood Indian School in Kundapura, India, the curriculum was adapted due to limited access to devices capable of typing code. Instead of Python, the girls used Scratch for block coding. They followed along as the instructor created a simple paddle ball game, answering questions in the chat to facilitate interactive learning. As with the Hudson sessions, the girls completed the pre-course survey before starting and the post-course survey at the end to measure changes in their perceptions of programming.

The survey results from both locations were analyzed to determine the impact of the Python Pals program on the participants' perceptions of programming. Changes in survey responses were evaluated to assess improvements in confidence and interest in programming. This structured approach ensured a comprehensive evaluation of the program's effectiveness in both in-person and virtual settings, accounting for different technological capabilities and learning environments.

Results and Discussion

After conducting 10 Python Pals sessions, data was collected from 62 students, divided into five groups based on shared characteristics: original curriculum (local girls), current curriculum (local girls), original curriculum (Indian girls), and current curriculum (Indian girls). This division allowed for analysis of the impact of curriculum changes and comparison between in-person and virtual learning environments.

Python Pals initially conducted in-person sessions for local middle school girls. The primary goal was to build confidence in programming and inspire further interest in coding. Two key survey statements provided insight into these goals: "I am good at programming" and "I want to learn more about programming in the future."

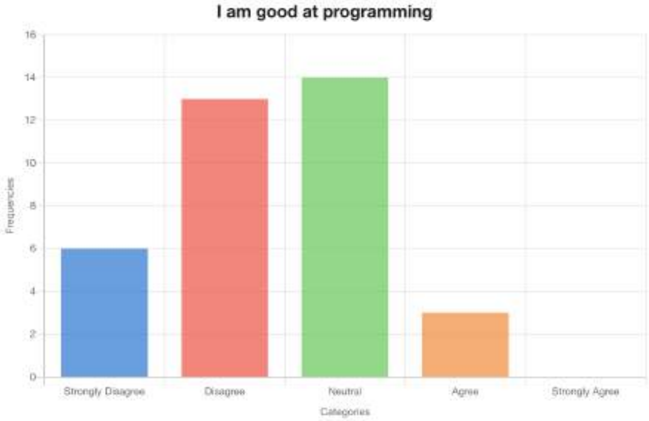


Fig 1: Responses to the statement "I am good at programming" before the Python Pals session.

Before the session, 36 local middle school girls were surveyed (Figure 1). None strongly agreed with the statement "I am good at programming," and only three agreed. The majority were either neutral or disagreed, with a mean response of 2.831 and a standard deviation of 0.871. This data suggests a 95% confidence interval for the true mean of responses between 2.541 and 3.131.

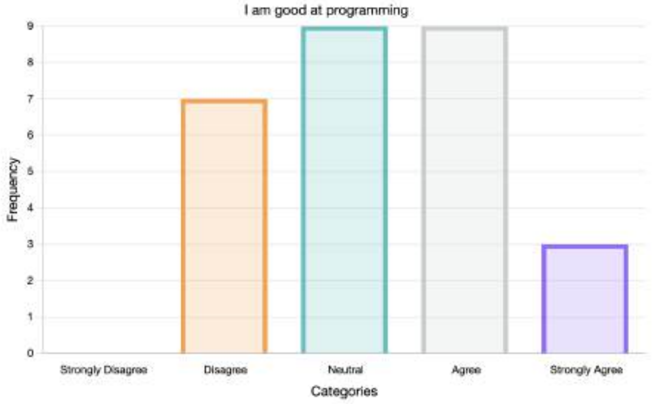


Fig 2: Responses to the statement “I am good at programming” after the Python Pals session.

After the session, 28 local middle school girls were surveyed (Figure 2). None strongly disagreed with the statement “I am good at programming,” and the mean response increased to 3.286 with a standard deviation of 0.975. The 95% confidence interval for the true population mean was between 2.908 and 3.664.

A hypothesis test was conducted to examine if the sessions significantly altered the girls’ perceptions. The null hypothesis posited that the true population means before and after the session were the same, while the alternate hypothesis suggested an increased mean post-session. A p-value of 0.033 was obtained, allowing rejection of the null hypothesis at a 5% significance level. This indicates that Python Pals successfully increased the girls’ confidence in their programming abilities.

Statement	Mean	Standard Deviation	Number of Responses	95% Confidence t-Interval
I want to learn more about programming in the future	3.750	0.937	36	(3.433, 4.067)
Programming is hard	3.083	0.874	36	(2.787, 3.379)

Table 1: Student response data on survey before Python Pals session.

Statement	Mean	Standard Deviation	Number of Responses	95% Confidence t-Interval
I want to learn more about programming in the future	4.142	0.848	28	(3.813, 4.471)
Programming is hard	3.571	1.609	28	(3.156, 3.986)

Table 2: Student response data on survey after Python Pals session.

Additional data from other survey statements is shown in Tables 1 and 2. For the statement “I want to learn more about programming in the future,” the mean response increased from 3.750 to 4.142, with a significant p-value at the 5% level. Interestingly, the mean response to “programming is hard” also increased, suggesting that while the course heightened interest in programming, it also highlighted its challenges. This could imply a need to adjust the course difficulty to balance engagement and manageability.

After evaluating the effectiveness of the in-person sessions, the responses of the middle school girls at Oakwood Indian School were compared with those of the local middle school girls to analyze any significant differences.

Statement	Mean	Standard Deviation	Number of Responses
Jobs in technology are boring	2.139	0.931	36
I want to learn more about programming in the future	3.750	0.937	36
Programming is hard	3.083	0.874	36

Table 3: Local middle school girls' responses to survey statements before Python Pals session.

Statement	Mean	Standard Deviation	Number of Responses
Jobs in technology are boring	2.333	0.976	15
I want to learn more about programming in the future	4.133	1.246	15
Programming is hard	2.800	1.014	15

Table 4: Oakwood Indian School student responses to survey statements before Python Pals session.

Tables 3 and 4 summarize their responses before the sessions. A hypothesis test comparing these groups found no significant differences, indicating similar perceptions and experiences with coding among both cohorts. Despite these similarities, it is noteworthy that women constitute a significant proportion of students in undergraduate and master's computer science programs in India, unlike the U.S., where the gender gap remains more pronounced. Specifically, women make up 40% of students in undergraduate CSE programs and 50% of students in master's CSE programs at most Indian colleges (Saxena).

Strengths and Limitations

This study was conducted to evaluate the effectiveness of Python Pals and to identify ways to enhance the program’s impact on middle school girls’ education in programming. However, several limitations should be noted. The study had a small sample size and its findings can only be generalized to girls in the Hudson, Ohio area and at Oakwood Indian School. Additionally, the participants likely had a pre-existing interest in programming, which may skew the results and not accurately reflect the broader gender gap in these areas. The anonymous nature of the survey also meant that individual responses could not be tracked before and after the program, limiting the ability to assess changes in perceptions at the student level.

Despite these limitations, the study provides valuable insights into the effectiveness of Python Pals. It highlights the program's potential to boost confidence and interest in programming among middle school girls, offering a foundation for future improvements. The results strongly indicate that the program boosted girls’ self-confidence, increasing the likelihood of girls pursuing programming studies in the future. The findings suggest that early exposure programs like Python Pals can play a crucial role in addressing gender disparities in computer science, making it a useful model for similar initiatives elsewhere.

Conclusion

This study evaluated Python Pals, a grassroots organization focused on empowering middle school girls in computer science. Operating in Hudson, Ohio, and Oakwood Indian School, Kundapura, India, Python Pals aimed to narrow the gender gap in tech by fostering coding interest and confidence early on. The curriculum featured interactive learning, group projects, and competitions. In Hudson, girls tackled basic Python and programming challenges, while in Oakwood, they engaged with Scratch due to limited typing resources.

Pre- and post-course surveys measured changes in participants' perceptions of their programming abilities and interest in pursuing computer science. The results showed Python Pals significantly enhanced girls' confidence in programming and their enthusiasm for further education in the field. This impact is crucial as early exposure and confidence play pivotal roles in encouraging women to pursue careers in technology, countering prevailing gender stereotypes.

Despite its achievements, the study identified areas for enhancement. Feedback highlighted the need to fine-tune the curriculum to better balance engagement with the perceived difficulty of programming. Limitations such as a small sample size and pre-existing interest among participants suggest further research is necessary to validate findings and extend the program's reach.

In conclusion, Python Pals demonstrates substantial potential in equipping middle school girls with programming skills and fostering a positive outlook on computer science. Its success underscores the importance of early interventions and ongoing efforts to create inclusive environments for girls in tech. By refining its curriculum and expanding its outreach, Python Pals and similar initiatives can continue to inspire and support the next generation of women in technology.

Acknowledgements
John Carroll University; Veale Youth Entrepreneurship Forum; Western Reserve Academy, Ms. Julianne Lopez, BA, MA; Dr. Ralf Borrmann, BA, MA, DML

Author

Riya Hegde, a senior at Western Reserve Academy, discovered her passion for computer science at an early age. At the end of her sophomore year, she started Python Pals, improving the confidence of over 70 girls and paving the way for the next generation of women in computer science.

Works Cited

"Changing the Curve: Women in Computing." *Berkeley School of Information*, 14 July 2021, ischoolonline.berkeley.edu/blog/women-computing-computer-science/.

Cheryan, Sapna, et al. "Computer Science and Engineering Need Women." *Scientific American*, vol. 328, no. 2, 2023, p. 10, https://doi.org/10.1038/scientificamerican0223-10.

Hegde, Riya. "Python Pals." *Python Pals*, pythonpals.com/.

Sax, Linda J., et al. "Anatomy of an Enduring Gender Gap: The Evolution of Women's Participation in Computer Science." *The Journal of Higher Education*, vol. 88, no. 2, 2016, pp. 258-93, https://doi.org/10.1080/00221546.2016.1257306.

Saxena, Pooja. "Ritual and Rhetoric of Gender Policies at the Indian Institutes of Technology." *Journal of Education Policy*, 2023, pp. 1-22. *[Database Name]*, https://doi.org/10.1080/02680939.2023.2293754.

"The STEM Gap: Women and Girls in Science, Technology, Engineering and Mathematics." *American Association of University Women*, www.aauw.org/resources/research/the-stem-gap/. Accessed 9 June 2024.

Innovative Cancer Treatments For Pediatric Patients With Blood Cancer By Nethra Mahesh

Abstract

Leukemia and lymphoma are prevalent types of cancer in pediatric patients and result in high mortality rates. The treatments available today are not effective enough, therefore we continue to see constant research and advancements being done so that we can have better and more effective treatments for childhood blood cancer than what we have now because we still have a long way to go. So, what do the future treatments for pediatric leukemia and lymphoma look like? Studies have been focusing on understanding the genetics behind pediatric cancers like leukemia and lymphoma. This review paper will explain and analyze what the future looks like with treatments for leukemia and lymphoma. These new treatments could save many lives of children all over the world, improve their quality of life, and improve the entire hospital experience.

Introduction

Pediatric cancer remains a great concern, with approximately 1 in 7,000 children diagnosed each year. Advancements in cancer therapy have led to 4 out of 5 of these children being cured, which is a development since 50 years ago when the cure rate of childhood cancer was less than 25% (Saletta et al. 156). This shows that we have made some progress, but it is not enough. There is still 1 out of 5 children that can't be cured using cancer therapy, therefore we continue to see constant research and advancements being done so that we can have better and more effective treatments for childhood blood cancer than what we have now because we still have a long way to go. So, what do the future treatments for pediatric leukemia and lymphoma look like? Studies have been focusing on understanding the genetics behind pediatric cancers like leukemia and lymphoma. For lymphoma, scientists have explored its molecular pathogenesis and developed molecularly targeted agents with diverse levels of effectiveness (Intlekofer and Younes, 585). Additionally, researchers have noticed that dysregulation of iron metabolism and obtaining excess amounts of iron on top of that are related to developing leukemia (Wang et al. 1). This review paper will explain and analyze what the future looks like with treatments for leukemia and lymphoma. These new treatments could save many lives of children all over the world, improve their quality of life, and improve the entire hospital experience.

CAR T-Cell Therapy

One of the possible future treatments for leukemia and lymphoma is CAR T-Cell Therapy. We have T-cells in our bodies which are part of the immune system. Their function is to protect the body from infections and cancer. CAR T-Cell Therapy can genetically alter T-cells in a patient whose cancer is caused by malfunctioning T-cells, enabling the production of a protein that can identify an antigen on the surface of the cancerous cell. This way, the modified T-cells can recognize and attack the cancer. There are several steps that go into the process of modifying the T-Cell. After the cells are gathered, they have to be multiplied for infusion. This takes a few weeks. Then, before infusion, the patient that will receive the treatment goes through a short course of chemotherapy. By doing this, their existing immune system is compromised, which increases the likelihood that the altered T-cells will expand and attack the cancer. Then it is time for infusion, which is where the modified T-cells are injected back into the patient. This usually lasts less than an hour ("Car T Cells: Engineering Immune Cells to Treat Cancer"). The most common side effects that can occur during this is low fever, infection, nausea, vomiting, and diarrhea. Furthermore, a few temporary side effects can include confusion, slurred speech, and seizures. On the other hand, a more serious possible side effect is cytokine release syndrome, or CRS. Cytokines are chemical messengers that aid in coordinating the immune system's defense against illness. CRS occurs when too many cytokines are produced ("CAR T-cell Therapy Side Effects"). There are a number of benefits and downsides to this treatment, and it is only one among several other treatments that are a possibility for the future.

Bispecific T-Cell Engagers

Another possible treatment that is being researched for leukemia and lymphoma in the future are Bispecific T-Cell Engagers (BiTEs). A BiTE is a particular type of bispecific antibody (BsAb). A BsAb is a type of protein that can recognize and bind to two different targets at the same time. For cancer treatment, BiTEs can be designed to redirect T cells to kill cancer cells. How this works is, BiTEs bind to a protein on the T-cell, a prevalent type of immune cell, and a protein on the cancer cell. This leads to the creation of a cytolytic synapse. Cytolytic synapse is essentially a bridge between the T-cell and the cancer cell. It activates the T-cell and prompts it to release toxic substances that can kill the cancer cell. These substances poke holes in the cancer cell's membrane, making it burst and die. There are many advantages to utilizing BiTEs to treat leukemia and lymphoma. A very significant one of these advantages is the fact that it works even when the Major Histocompatibility Complex (MHC) is not functioning properly. Usually, T-cells need signals from something called the Major Histocompatibility Complex before it attacks the cancer cell. With BiTEs, this is a factor that we do not need to worry about. Another advantage is that BiTEs are very specific. This is good because it can target the cancer cells successfully without harming any healthy cells. On the other hand, there are some side effects that should be brought to attention as well. Side effects of BiTE therapy can include cytokine release syndrome (CRS) and neurotoxicity. There is ongoing research being done to improve the efficacy and safety aspects of BiTE therapy. There is even development of second-generation BiTEs with enhanced properties (Smits and Sentman, 1131–1133).

Cytokine-Based Immunotherapy

The last future treatment being covered in this paper is Cytokine-Based Immunotherapy. Cytokines are signaling proteins produced by different immune system cells. Through their interactions with certain receptors on target cells, they control immune system responses. In the context of leukemia and lymphoma, cytokines are crucial in regulating the body's immune responses against cancer cells. This is due to their ability to activate immune cells such T-cells, B-cells, and NK (natural killer) cells, which target cancer cells. Cytokine-Based Immunotherapy is the use of cytokines to boost the immune system's ability to combat tumors. IL-2 is a cytokine that is used in this treatment. IL-2 is an important cytokine because of its ability to stimulate T-cells and NK cells. There are other cytokines such as interferons (IFNs) and interleukins (ILs) that are being researched for their potential use in Cytokine-Based Immunotherapy. So far in the clinical trials that have been done, some patients experienced the benefits of the treatment including tumor regression and prolonged survival, but some patients didn’t respond to the treatment. If the patient experiences no complications, this treatment can be effective and help the patient significantly. There are some challenges and limitations, though. Cytokine-based cancer immunotherapy can result in side effects including CRS and vascular leak syndrome (VLS) (Atallah-Yunes and Robertson, 872010). The effectiveness of the treatment can also be compromised by tumor heterogeneity which is variations in the same type of tumor in several patients. Research is being done to address these challenges and this can lead to developing this treatment to have improved safety and efficacy. Additionally, research will help find biomarkers to predict patient responses to treatment.

Conclusion

This paper has explored discoveries in future treatments for pediatric leukemia and lymphoma, focusing on CAR T-Cell Therapy, Bispecific T-Cell Engagers, and Cytokine-Based Immunotherapy. These innovative treatments offer promising advancements in targeting cancer cells more effectively while also improving the quality of life for pediatric patients. However, challenges such as cytokine release syndrome and neurotoxicity remain. Continued research is important to address these issues and enhance the efficacy of these treatments. Although these challenges exist, the future holds great promise for more effective and life-saving treatments for pediatric leukemia and lymphoma.

Works Cited

Atallah-Yunes, Suheil Albert, and Michael J Robertson. "Cytokine Based Immunotherapy for Cancer and Lymphoma:
Biology, Challenges and Future Perspectives." Frontiers in Immunology, U.S. National Library of Medicine, 20 Apr. 2022, www.ncbi.nlm.nih.gov/pmc/articles/PMC9067561/#:~:text=Cytokines%20regulate%20both%20the%20innate,clinical%20trials%20of%20cancer%20immunotherapy.

Afaf E.G. Osman. "Chronic Myeloid Leukemia: Modern Therapies, Current Challenges and Future Directions." Blood Reviews, Churchill Livingstone, 16 Mar. 2021, www.sciencedirect.com/science/article/abs/pii/S0268960X2100031X.

"Car T Cell Therapy." MD Anderson Cancer Center, www.mdanderson.org/treatment-options/car-t-cell-therapy.html. Accessed 9 July 2024.

"Car T Cells: Engineering Immune Cells to Treat Cancer." CAR T Cells: Engineering Immune Cells to Treat Cancer - NCI, www.cancer.gov/about-cancer/treatment/research/car-t-cells. Accessed 9 July 2024.

Intlekofer, Andrew M., and Anas Younes. "Precision Therapy for Lymphoma-Current State and Future Directions." Nature News, Nature Publishing Group, 19 Aug. 2014, www.nature.com/articles/nrclinonc.2014.137.

"NCI Dictionary of Cancer Terms." Comprehensive Cancer Information - NCI, www.cancer.gov/publications/dictionaries/cancer-terms/def/tumor-heterogeneity. Accessed 9 July 2024.

Saletta, Federica, et al. "Advances in Paediatric Cancer Treatment." Translational Pediatrics, U.S. National Library of Medicine, Apr. 2014, www.ncbi.nlm.nih.gov/pmc/articles/PMC4729100/.

Smits, Nicole C, and Charles L Sentman. "Bispecific T-Cell Engagers (Bites) as Treatment of B-Cell Lymphoma."
Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology, U.S. National Library of Medicine, 1 Apr. 2016, www.ncbi.nlm.nih.gov/pmc/articles/PMC5085271/#:~:text=BiTEs%20are%20a%20class%20of,cytotoxic%20activity%2C%20against%20cancer%20cells.

Wang, Fang, et al. "Iron and Leukemia: New Insights for Future Treatments - Journal of Experimental & Clinical Cancer Research."
SpringerLink, BioMed Central, 13 Sept. 2019, link.springer.com/article/10.1186/s13046-019-1397-3.

"Car T-Cell Therapy Side Effects." MD Anderson Cancer Center, www.mdanderson.org/patients-family/diagnosis-treatment/emotional-physical-effects/car-t-cell-therapy-side-effects.html. Accessed 16 July 2024.

Acknowledgements

I attest that the ideas, graphics, and writing in this paper are entirely my own unless explicitly stated otherwise. I thank Matthew Aronson from the University of Pennsylvania and the Children’s Hospital of Philadelphia and Indigo Research for their mentoring and guidance.

Obstacles and Potential Solutions to Feed and Sustain Humans on Mars By Kaitlin Cho

Abstract

One of Earth's closest neighbors, Mars is a prime candidate for finding a "second Earth". Missions to this planet like NASA's Perseverance have looked closely into factors like atmosphere and soil that impact Mars's ability to host human populations. Currently, the cold climate and the unsuitable balance of chemicals prevent the possibility of agriculture. Exploratory technology such as advanced greenhouses can take advantage of existing resources to help life withstand these harsh conditions. As more missions explore in greater depth how we can use Mars's resources, humans can develop the necessary technology to create agricultural practices on this planet. Eventually, this information can be used to sustain the first human community.

Introduction

Humans have been searching for signs of Earth-like chemicals and conditions on Mars for decades to learn more about the possibility of life, and whether the planet could be suitable for a civilization like those on Earth. If humans were to truly find these necessary resources on Mars, what would it take to form a new civilization there? Further research is needed to discover how humans can adapt to the different environment, including the capability of growing food. Creating a system of agriculture on Mars requires knowledge of its existing conditions and how technology can be used to help humans expedite that process.

The need to explore Mars arises from many factors, an important one being climate change. Currently, 3.3 billion people's lives are "highly vulnerable" due to climate change, and it is predicted that parts of Africa and Asia will be uninhabitable by 2100 (1, 2). This reinforces the need for a second Earth, which humans can rely on in the case of climate disaster. Another possible motivation to explore Mars is the emergence of commercial space travel in the coming decades. Privately owned companies like SpaceX and Axiom Space are already partnering with NASA to send more people to space. In addition to leisurely avenues, a population on Mars would benefit increasingly progressing industries like space mining. Initiators of this business are interested to see the potential in an economy beyond the boundaries of Earth. These monetary desires can propel Mars exploration to a point where we must research the livability of this planet.

Water

General Facts

Five reservoirs on Mars hold water, the essential liquid for life: the atmosphere, surface ice, water absorbed in the soil, hydrated minerals, and subsurface ice (3). Water is most present at the northern and southern poles. These locations host a variety of forms of water, from water vapor to icy frost. Some forms are more abundant than others. In the case of frost, Carbon

Dioxide (CO_2) frost has been more commonly found than water-based frost. Scientists have noticed there is more water (H_2O) frost only in locations where CO_2 is less common (3). Compared to the more traceable H_2O frost, a less explored indicator of flowing water is recurring slope lineae (RSL). As pictured in Figure 1, this unique pattern has intrigued the scientific community (4). Some think the RSL may be sand or other granular material. However, scientists have found these patterns more frequently appear in the warmer seasons where water could more commonly be found in liquid form, suggesting the connection (5).

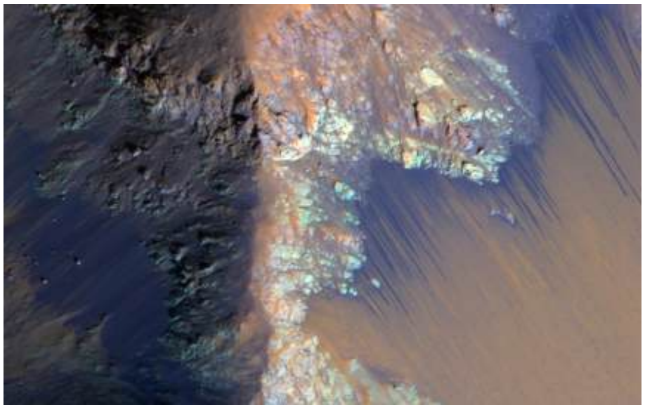


Figure 1: Enhanced image of RSL by NASA's Mars Reconnaissance Orbiter. Dark streaks are believed to be a sign of flowing water. Ref. 4.

To better understand where water appears on and around Mars, scientists can use signs like RSL to first record its accessibility near various landing spots in future missions, and then expand to other areas on Mars. With more knowledge about how to find resources like water, missions can last longer.

History

Researching the surface of Mars has led to discoveries regarding the presence of water in the planet's past. Scientists believe that around 4.1 to 3 billion years ago, Mars may have been suitable for life with large bodies of water like the oceans we have on Earth (5). Evidence of stones that form in water, like hematite and gypsum, suggests the past presence of it (6). Particularly, scientists believe there was much more surface-level water, shown through water debris reminiscent of flash floods in deserts on Earth (7). Like many other planets, Mars is also thought to have had different amounts of water accessible over its history. Scientists have found evidence of multiple ice ages, with the most recent one being an estimated 400,000 to 2.1 million years ago (8, 9).

Problems and Potential Solutions

Water, one of the most important resources for humans, is currently difficult to find on the current surface of Mars. Liquid H_2O on Mars is especially hard to locate. Most water is found in gas form due to more evaporation from the low air pressure and high temperatures compared to Earth (11, 12). To illustrate, Earth's average atmospheric temperature is 57 ºF, while Mars' is around 120 ºF, as seen in Figure 2 (13).

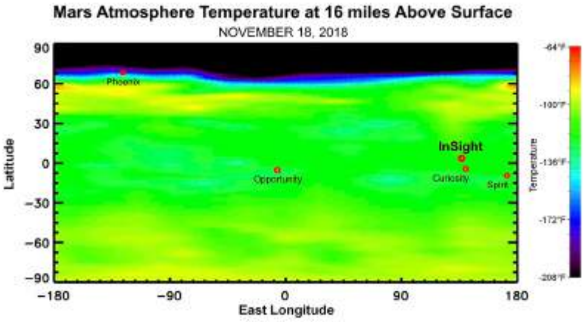


Figure 2: Graph of Mars Atmospheric Temperature as compared with the Latitude and Longitude. This shows an average between -100º and -136º Fahrenheit. Ref. 13.

On the ground, frost is widespread, especially in colder seasons. However, most of this frost is CO_2, much like the dry ice on Earth (12). In contrast, H_2O frost is found under this layer on a portion of polar caps (5). This water is mixed with dust and soil that would be unsanitary, so it would require a filtering system. A possibly more accessible source of H_2O would be groundwater, although detecting this is difficult with current technology (5). In order to more easily gain access to H_2O for humans on Mars, scientists should locate spots where water is most commonly found and improve extraction and filtration systems. Primitive methods of water detection are currently being used on many missions. For example, signs of salty water were first detected on the Reconnaissance orbiter using high-resolution cameras. This mission also found sheets of ice at the bottoms of various craters, signaling the presence of H_2O (14). Also, the SHERLOC function on NASA's Perseverance uses cameras, spectrometers, and lasers to detect materials that have been impacted by water. These can range from signs of microbial organisms, or rocks that form in the presence of water (15). Through further research, knowing where water is could help support both human and plant life on Mars.

Regolith

General Facts

Regolith, the soil-like substance on the surface of Mars, has many of the necessary components for life. Elements like carbon, nitrogen, hydrogen, oxygen, and sulfur are all present, albeit in smaller concentrations than on Earth (16). Additionally, clay minerals and salts are also widespread, which form in water on Earth (17). This shows the possibility of water, especially in the history of Mars, since these elements could have been left over. There is an abundance of harmful perchlorates and sulfates in the current martian soil, which has a pH of 8.3 (17). Earth's pH 6-7 soil is comparatively acidic and optimal for growing plants. Neutral soil is the most effective for successful harvests, providing the greatest nutrients to the plants (2).

Impact on Life

Perchlorates are the most crucial barrier to the development of an agriculture industry on Mars. They result from the UV radiation from the sun together with chlorine in the regolith, creating these harmful substances in the soil (18). Perchlorates are toxic to humans, so even if plants do grow in Martian soil, they would be inconsumable. Although this is a significant obstacle, scientifically manipulating perchlorates could result in some benefits, since some microbes on earth use perchlorates as energy (16). Another large obstacle is radiation from the sun. Combined with the oxidizing, self-sterilizing soil, this makes it nearly impossible for organisms to stay alive on the surface of Mars (19). Self-sterilizing soil limits life by killing off any microbes or plant life within it, preventing microorganisms from forming. Further research into the makeup of the regolith is needed for humans to make progress towards agriculture on Mars.

Atmosphere

General Facts

The Martian atmosphere has significant differences compared to that of Earth. Elements like nitrogen and oxygen, heavily concentrated on Earth's atmosphere, are less common on Mars, as seen in Figure 3 (20). This creates an issue for agriculture due to plants' reliance on chemicals like oxygen. Although plants take in carbon dioxide for cellular respiration, too much would negatively impact plants as well as humans.

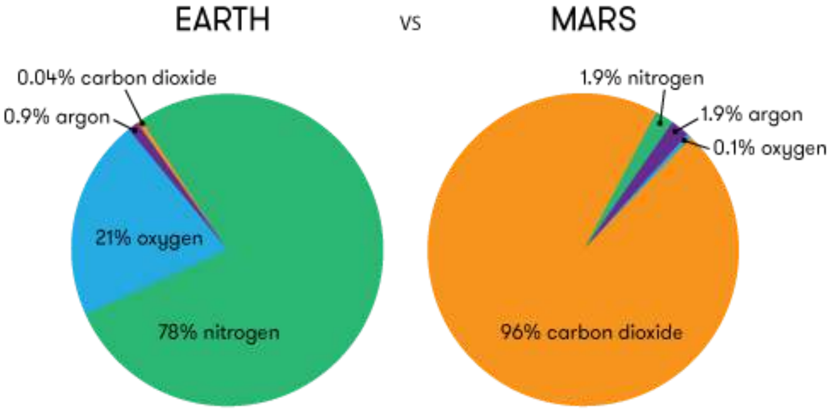


Figure 3: Comparing atmosphere concentration between Earth and Mars. Note the crucial difference in oxygen concentrations. Figure taken from Ref. 20.

Another difference between Earth and Mars is the magnetic field. Earth has a polar magnetic field that encompasses the whole planet, while Mars has localized magnetic fields (7). When plants are grown with weak magnetic fields, it impacts their ability to properly control root functions (1). Finding these local magnetic fields is important if humans are to successfully sustain agriculture. NASA's Mars Global Surveyor has made preliminary progress towards mapping these fields out. It uses a magnetometer to find anomalies on the surface of Mars that come from a possibly extinct magnetic core (22).

Climate and Weather

The current climate and weather of Mars cannot sustain humans. Frequent dust storms shift the makeup of the atmosphere. Water transport, the movement of water in the atmosphere, from the middle to upper atmosphere is increased, making it easier for water to get lost to space (8). Additionally, dust storms impact the heat balance in the polar regions due to the shifting of the frost cycle. Dust storms have had a negative impact on past missions like NASA's Opportunity and InSight, which is why further research is necessary to determine patterns in this weather (8). Currently, scientists have discovered two main factors that affect these dust storms: a pressure gradient in the northern hemisphere and seasonal CO_2 buildup. CO_2 in ice form specifically traps the dust particles in the atmosphere, removing them (23). The frequency of certain storms lessen, overall decreasing the total amount of storms.

Martian weather displays both similarities to and differences from that of Earth. Both planets have repeating weather seasons. Particularly on Mars, scientists have found that dust storms occur in the same weeks as previous years (22). In contrast to Earth, there are significant temperature changes around Mars throughout the day. There is a significant temperature gradient with increasing altitude (areas a few feet apart can have large temperature differences). This is due to both the lack of pressure and the lack of geographic features that regulate climate (24). On Earth, one example of geography would be large bodies of water like oceans. Mars lacks these, impacting the ability to maintain a constant temperature (23).

Problems and Solutions

The atmospheric pressure on Mars varies due to condensation and sublimation of CO_2 (19). This change has been visible both on a day-to-day basis and over extended periods of time. Research shows that Mars used to have a much higher atmospheric pressure, which has led to "dry ice" behavior of both water and CO_2 ice over time (7, 8). This prevents water from changing phases into drinkable, liquid water because it usually shifts between solid and gas forms.

Much of the atmosphere has been stripped away over time as shown in Figure 4 (25). Scientists have discovered this to be the result of exposure to solar wind (26). Mars's low gravity and lack of magnetic field left the atmosphere vulnerable to the sun, which causes solar winds to strip lighter molecules from the atmosphere (26, 27). This causes a problem for the consistency of the atmosphere.

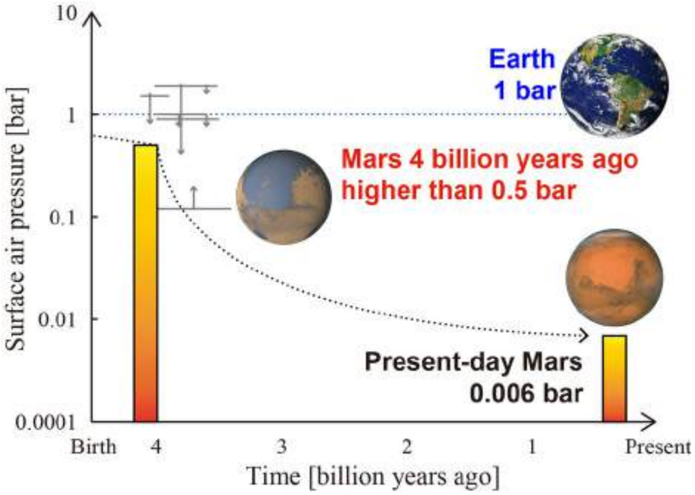


Figure 4: Atmospheric pressure loss on Mars compared with consistent atmospheric pressure on Earth today. Taken from Ref. 25.

The lack of oxygen in the atmosphere remains a large obstacle in the way for humans to survive on this planet. NASA has included a mechanism on the Perseverance rover called MOXIE (Mars Oxygen In-Situ Resource Utilization Experiment) to present a possible solution. MOXIE takes CO_2 from the air and transforms it into carbon monoxide and oxygen, providing a source of oxygen for humans to breathe. Although a much larger scale MOXIE will be needed to actually produce enough oxygen for humans, this is an important step in an evolving journey (15). Carbon monoxide is harmful for humans, but many microbes on Earth can convert CO into usable energy (28).

Technology

This section primarily provides insight into the devices on the NASA Perseverance missions, and touches on other innovations for life on Mars. As this is one of the most recent missions, discussing these technologies can spur ideas for improvements to both manned and unmanned missions.

Water and Agricultural Conditions

The main system on Perseverance that tracks water is RIMFAX (Radar Imager for Mars' Subsurface Experiment), which mainly looks for water above and below ground. Using radar waves, RIMFAX detects H_2O ice and salts underground. This can especially support humans on manned missions when humans need consistent sources of water to survive (15).

There are some benefits to growing plants on Mars. Due to the lower gravity, water travels more slowly through the soil, giving plants' roots more time to absorb the flowing water. This means plants actually need less water than on Earth, since less water is wasted simply from it seeping through the soil, not absorbed and used (29). However, since Mars is farther away from the sun, it has limited access to sunlight, similar to places like Norway in winter on Earth (29, 30). One solution scientists propose for this is engineering genes of extremophiles (microorganisms that can survive in extreme conditions like high salinity, high/low temperatures, etc.) into plants so they can withstand harsh conditions (31). A team at North Carolina State University has already successfully transferred such genes to many hosts like jellyfish and tobacco cells (31). Further research is needed before we can incorporate these into plants that humans consume. Humans could also bring fertilizers that supply plants of necessary nutrients that are hard to find on Mars (30). By further exploring different solutions, humans can work towards developing a sustainable system of agriculture.

Regolith

Perseverance devotes a large portion of its rover to researching the composition of Mars's surface. Both the Mastcam-Z and PIXL (Planetary Instrument for X-ray Lithochemistry) use spectrometers and cameras to detect chemical signs of past life (15). SHERLOC (Scanning Habitable Environments with Raman & Luminescence for Organics & Chemicals) and SuperCam observe patterns in the soil and rocks that help discover what is needed to live on Mars (15). All four of these devices help scientists learn about the history and current state of Mars's surface. This is crucial to knowing the possibility of both agriculture and human life on Mars.

Scientists have proposed many ways to transform existing materials into more useful substances on Mars. As shown in Figure 5, abundant resources like CO_2 and mineral oxides can be chemically shifted into O_2 that can be used for humans (32).

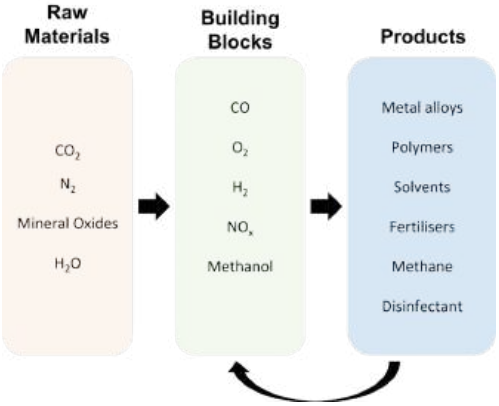


Figure 5: Possible transformation of raw martian materials into rarer resources for human sustainability. Taken from Ref. 32.

On Earth, scientists have also proposed a variety of ways to manipulate Martian regolith into something more suitable for plant life. To combat the lack of heat, scientists can use geothermal heat from volcanoes (30). Mars's past reveals an abundance of early volcanic activity, which means this can be obtained from somewhere underground. This heat can be used to help life withstand harsh climates. Some scientists propose to "bake soil" to extract water by heating it up and causing evaporation. Then, they can transport the water to different sites for growing plants (30). This also requires scientists to continue researching where the most suitable sites are. Research from PIXL and Mastcam-Z shows that nutrients in Martian soil vary depending on the location (29). Overall, sites that have the ideal heat and nutrients should be transformed into usable plots of land for agriculture, and thus sustainment of human life.

Atmosphere

MEDA (Mars Environmental Dynamics Analyzer) and MOXIE are the main mechanisms that research the atmosphere on the Perseverance Rover (15). MEDA mostly tracks climate and weather, aiming to forecast incoming changes including dust storms. MOXIE, as referenced in the atmosphere section, is an innovative approach to producing more oxygen on Mars (15).

MOXIE currently can produce as much oxygen "as a modest-sized tree", so it would need to be thirty times larger to sustain a small crew of humans (15). As for more sustainable power sources, humans can use common mechanisms like solar panels, as well as more powerful ones like nuclear fission. Nuclear fission would be a more reliable and efficient source of the large amount of energy needed to sustain a civilization (33). Since the only celestial body this technology has been tested on is Earth, its capabilities on Mars are still up for exploration.

To apply these ideas to agriculture, an insulated greenhouse can provide an enclosed area where a machine like MOXIE can operate on a smaller scale (34). The surface of Mars is currently unable to support plant life because of its low temperature and sterilizing ultraviolet radiation (31). This demonstrates the need for insulation in addition to mechanisms like MOXIE. When research is fully developed, addressing problems in the atmosphere of Mars can slowly be integrated into small scale solutions for agriculture.

Conclusion

Through further research on water, regolith, and atmosphere on Mars, scientists can learn more about the capabilities of life on the planet. Although problems like perchlorates in the regolith and lack of atmospheric pressure currently stand in the way of creating a system of agriculture, humans can continue to develop innovations like NASA's MOXIE to solve these issues.

New technology is already emerging, not just for the Mars project but also for other prospective "second Earths". Other bodies like Earth's Moon and Jupiter's moon Europa are possible locations to continue exploration. Axiom Space is working on spacesuit technologies in collaboration with NASA, possibly for an Artemis III mission to the moon projected to occur in 2025 or 2026 (35, 36). This research could provide insight into how humans will explore Mars. Spacesuit testing is already occurring on Mars itself. On NASA's Perseverance, SHERLOC carries materials like teflon and polycarbonate to explore their ability to withstand the conditions of Mars (37). This experimentation, along with other research about the planet itself, can accelerate manned missions to the red planet.

NASA has three main goals for Mars exploration: Explore the Potential for Martian Life, Support Human Exploration of Mars, and Discover Dynamic Mars (38). Through further research on the capabilities of Mars to support life, particularly agriculture, these goals can be achieved.

Works Cited

1. Press, Associated, and Seth Borenstein. “Landmark UN Climate Change Report: ‘Parts of the Planet Will Become Uninhabitable.’” *WHYY*, WHYY, 28 Feb. 2022, whyy.org/articles/un-ipcc-climate-change-report-uninhabitable-planet-code-red/. Accessed 5 June 2024.

2. Sands, Leo. “Earth Could Enter ‘Doom Loop’ Stage of Climate Crisis, Report Warns - The Washington Post.” *The Washington Post*, 16 Feb. 2023, www.washingtonpost.com/climate-environment/2023/02/16/doom-loop-earth-climate-change/. Accessed 5 June 2024.

3. Lange, L., et al. "A Reappraisal of Near-Tropical Ice Stability on Mars." arXiv preprint arXiv:2306.16987 (2023). Accessed 5 June 2024.

4. *Recurring Slope Lineae in Coprates Chasma*. 9 Apr. 2014. *NASA Jet Propulsion Library*, https://www.jpl.nasa.gov/images/pia18119-recurring-slope-lineae-in-coprates-chasma. Accessed 17 July 2024.

5. “Your Guide to Water on Mars.” *The Planetary Society*, 25 Oct. 2022, www.planetary.org/articles/your-guide-to-water-on-mars. Accessed 5 June 2024.

6. “Mars Exploration Rover - Nasa Facts.” *NASA*, mars.nasa.gov/internal_resources/825/. Accessed 6 June 2024.

7. “Mars Global Surveyor - NASA’s Mars Exploration Program.” *NASA*, mars.nasa.gov/internal_resources/813/. Accessed 6 June 2024.

8. “MRO Science Highlights - NASA Science.” *NASA*, NASA, science.nasa.gov/mission/mars-reconnaissance-orbiter/science-highlights/. Accessed 6 June 2024.

9. “Mars May Be Emerging from an Ice Age - NASA Science.” *NASA*, NASA, 17 Dec. 2003, science.nasa.gov/solar-system/planets/mars/mars-may-be-emerging-from-an-ice-age/. Accessed 6 June 2024.

10. Schmidt, Frédéric, et al. "Circumpolar ocean stability on Mars 3 Gy ago." Proceedings of the National Academy of Sciences 119.4 (2022): e2112930118. Accessed 5 June 2024.

11. Greicius, Anthony. “NASA’s Treasure Map for Water Ice on Mars.” *NASA*, NASA, 26 July 2023, mars.nasa.gov/news/8568/nasas-treasure-map-for-water-ice-on-mars. Accessed 6 June 2024.

12. Jet Propulsion Laboratory. “Science at Sunrise: Solving the Mystery of Frost Hiding on Mars.” *NASA*, NASA, 7 Sept. 2023, mars.nasa.gov/news/9184/science-at-sunrise-solving-the-mystery-of-frost-hiding-on-mars/. Accessed 6 June 2024.

13. NASA, and JPL-Caltech. *Mars Atmosphere Temperature at 16 Miles Above Surface*. 21 Nov. 2018. *NASA Jet Propulsion Library*, https://www.jpl.nasa.gov/images/pia22570-martian-weather-forecast-for-insight-landing. Accessed 6 June 2024.

14. Tillman, Nola Taylor. “Water on Mars: Exploration & Evidence.” *Space.Com*, Space, 18 Aug. 2018, www.space.com/17048-water-on-mars.html. Accessed 6 June 2024.
15. “Mars 2020: Perseverance Rover - NASA Science.” *NASA*, NASA, mars.nasa.gov/mars2020/spacecraft/instruments. Accessed 4 Apr. 2024.
16. David, Leonard. “Toxic Mars: Astronauts Must Deal with Perchlorate on the Red Planet.” *Space.Com*, Space, 13 June 2013, www.space.com/21554-mars-toxic-perchlorate-chemicals.html Accessed 5 June 2024.
17. “Mars Odyssey - NASA Science.” *NASA*, NASA, mars.nasa.gov/odyssey/mission/science/results/. Accessed 6 June 2024.
18. Wall, Mike. “Mars Soil May Be Toxic to Microbes.” *Space.Com*, Space, 6 July 2017, www.space.com/37402-mars-life-soil-toxic-perchlorates-radiation.html. Accessed 6 June 2024.
19. “Viking Project - NASA Science.” *NASA*, NASA, mars.nasa.gov/mars-exploration/missions/viking-1-2/. Accessed 6 June 2024.
20. Australian Academy of Science. *Atmospheric composition of Earth vs Mars*. *Australian Academy of Science*, https://www.science.org.au/curious/space-time/mars. Accessed 6 June 2024.
21. Valentine, Theresa. “Magnetic Fields and Mars.” *NASA*, NASA, mgs-mager.gsfc.nasa.gov/Kids/magfield.html. Accessed 6 June 2024.
22. “Mars Global Surveyor - NASA Science.” *NASA*, NASA, mars.nasa.gov/mars-exploration/missions/mars-global-surveyor. Accessed 6 June 2024.
23.“Martian Weather.” *Center for Astrophysics | Harvard & Smithsonian*, 3 Nov. 2010, www.cfa.harvard.edu/news/martian-weather. Accessed 5 June 2024.
24. “Today, Mars Is Warmer than Earth. See How We Compare.” Homepage, 5 Jan. 2018, airandspace.si.edu/stories/editorial/today-mars-warmer-earth-see-how-we-compare. Accessed 6 June 2024.
25. Tokyo Institute of Technology. *The figure shows how surface air pressure changed throughout Martian history. A bar at 4 billion years ago denotes a lower limit shown by this study. Constraints suggested by other studies are also shown by arrows.* 19 Sept. 2017. *PR Newswire*, https://www.prnewswire.com/news-releases/meteorite-tells-us-that-mars-had-a-dense-atmosphere-4-billion-years-ago-300522603.html. Accessed 6 June 2024.
26. NASA’s Goddard Space Flight Center. “GMS: Maven Results Live Shot Page.” *NASA*, NASA, 15 Nov. 2015, svs.gsfc.nasa.gov/cgi-bin/details.cgi?aid=12042. Accessed 6 June 2024.
27. Dobrijevic, Daisy. “Mars’ Atmosphere: Facts about Composition and Climate.” *Space.Com*, 25 Feb. 2022, www.space.com/16903-mars-atmosphere-climate-weather.html. Accessed 6 June 2024.
28. Holmes, Bob. “Martian Life Must Be Rare as Free Energy Source Remains Untapped.” *New Scientist*, New Scientist, 4 May 2017, www.newscientist.com/article/2129859-martian-life-must-be-rare-as-free-energy-source-remains-untapped/. Accessed 6 June 2024.

29.Candanosa, Roberto Molar. "Growing Green on the Red Planet." *American Chemical Society*, 2017, www.acs.org/education/resources/highschool/chemmatters/past-issues/2016-2017/april-2017/growing-green-on-the-red-planet.html. Accessed 6 June 2024.
30. Wall, Mike. "Incredible Technology: How to Live on Mars." Space.Com, Space, 13 Aug. 2013, www.space.com/22342-how-to-live-on-mars-colony-technology.html. Accessed 6 June 2024.
31. "Designer Plants on Mars." *NASA*, NASA, 6 June 2013, www.nasa.gov/news-release/nasa-designer-plants-on-mars/#:~:text=The%20plants%20would%20probably%20be,what%20ordinary%20plants%20could%20stand. Accessed 6 June 2024.
32. Baldry, Mark. *Building block framework for Martian ISRU* . 15 Mar. 2022. *Wevolver*, https://www.wevolver.com/article/how-to-survive-on-mars. Accessed 6 June 2024.
33. Hall, Loura. "6 Technologies NASA Is Advancing to Send Humans to Mars." *NASA*, NASA, 17 July 2020, www.nasa.gov/directorates/stmd/6-technologies-nasa-is-advancing-to-send-humans-to-mars/. Accessed 6 June 2024.
34. Harfield, Jake. "Can Humans Live on Mars? The Technology That Can Make It Happen." *MUO*, 24 July 2021, www.makeuseof.com/can-humans-live-on-mars/. Accessed 6 June 2024.
35. "Axiom Suit." *Axiom Space*, www.axiomspace.com/axiom-suit. Accessed 6 June 2024.
36. Chang, Kenneth. "Elon Musk Says Spacex Could Land on Mars in 3 to 4 Years." *The New York Times*, The New York Times, 5 Oct. 2023, www.nytimes.com/2023/10/05/science/elon-musk-spacex-starship-mars.html. Accessed 6 June 2024.
37. "Spacesuit Materials on Perseverance's SHERLOC Calibration Target - NASA Science." *NASA*, NASA, 28 July 2020, science.nasa.gov/resource/spacesuit-materials-on-perseverances-sherloc-calibration-target/. Accessed 6 June 2024.
38. Lanson, Eric E. "Mars Exploration Program Future Plan." NASA. Accessed 6 June 2024.

No-Fault Divorce and the Second Wave Feminist Movement between 1960 to 1980 By Zifei Wang

Abstract

Spurred in part by the success of authors like Betty Friedan (*The Feminine Mystique*) and artists like Leslie Gore ("You Don't Own Me"), the United States saw a second rise of feminism in the 1960s. Through the struggle of redefining female liberation, women's role in the household began shifting. Part of this changing social tide manifested itself in law, with the rise of no-fault divorces, allowing women to leave marriages without having to formally prove any breach of a marital contract. By the end of 1969, California passed the first no-fault divorce law and this new legislation soon spread across the nation. While seemingly a tool for female empowerment, this legislation often worsened women's economic and social standing. This paper will explore the evolution of second-wave feminism between 1963 and 1980 and the unanticipated effects of no-fault divorce on the movement.

Introduction

Given how common divorce is today (nearly 1 in 2 first marriages end in divorce, with second or third marriages failing at a far higher rate), it is easy to take this act of legal dissolution as a given. There is, however, an important and often underappreciated history of American divorce in the twentieth century that had - and still has - far-reaching implications for millions of American women. At the end of 1969, California passed America's first "pure" no-fault divorce legislation. Earlier divorce proceedings required one party to file a complaint alleging that the other party was at fault. Defendants in court could contest the action by either proving the charges against them to be false or by establishing fault in the other party as well. Commonly contested issues of dissolution regarding alimony, child custody, child support, and property distribution, were resolved by examining marital fault. However, this system was not flawless. Alimony was frequently reduced or eliminated if the spouse seeking alimony had committed any marital misconduct.

In 1966, the California Governor's Commission – a group appointed after a previous legislative inquiry into how judges applied California's divorce laws – found that fault-based divorce was no longer representative of a sound legal or social policy. The report recommended marriage in a non-adversarial setting as the only basis for dissolution. This policy change had profound impacts beyond just lawyers and judges. Liberal feminists believed that the new act would be a paradigm for reform and would attempt to eliminate the characterization of women as inferior to men by overturning laws that weighed female and male legal rights as unequal. Equal opportunity to leave a marriage unilaterally was given to both sexes.

On a legislative level, no-fault divorce was resoundingly successful. Within four years after its passing, one out of every three American states had adopted some form of no-fault divorce. However, despite the approval of the law, the policy instead resulted in more female exploitation and increased the gender gap socioeconomically within communities. While women could more easily liberate themselves from challenging home lives, they could not fully recognize economic or social independence. This account of the rise and consequences of no-fault divorce is an important reminder of the muddled intersection between the domestic, economic, and social lives of twentieth-century women and just how much second-wave feminists were up against.

Ideas of Second-wave Feminism

Before delving deeper into the aftermath of no-fault divorce, it is worth sketching the twentieth-century cultural context of norms and practices that may seem foreign to young women today. No place was more important for mid-century women than the home. The housewife in the 1950s was an essential job to maintain the peace of the household. As domesticity within the United States during the post-World War II era surged, so too did suspicion during the Cold War era. Grounded in the essential belief that women would guard against the spread of these socialist and communist ideals, the housewife's important social role was solidified. The role of "supermother" became a resounding phenomenon, defined by mothers' responsibilities towards their children. Media and widespread societal messaging further pressured American women toward families and their suburban homes.

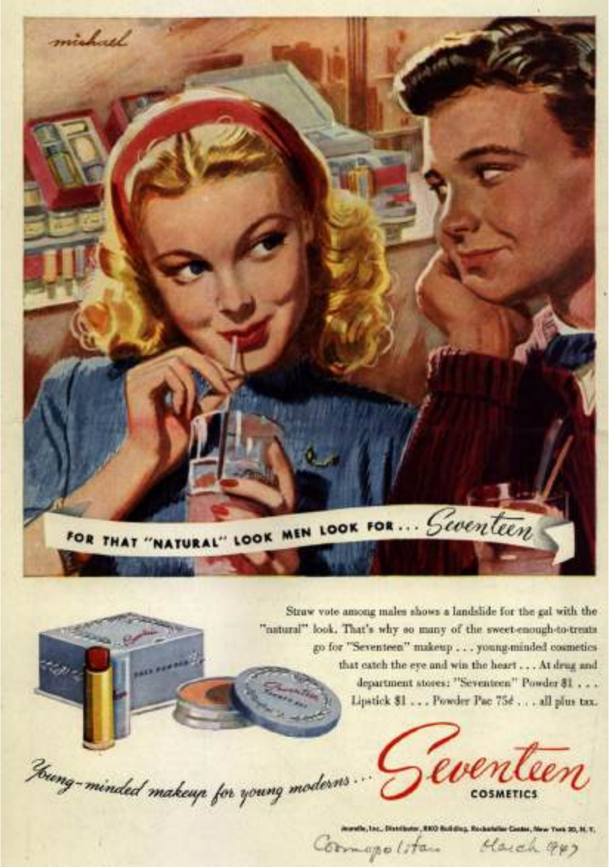

michael
FOR THAT "NATURAL" LOOK MEN LOOK FOR... Seventeen
Straw vote among males shows a landslide for the gal with the "natural" look. That's why so many of the sweet-enough-to-treats go for "Seventeen" makeup . . . young-minded cosmetics that catch the eye and win the heart . . . At drug and department stores: "Seventeen" Powder $1 . . . Lipstick $1 . . . Powder Pac 75¢ . . . all plus tax.
Young-minded makeup for young moderns . . .
Seventeen
COSMETICS
[illegible], Inc., Distributor, RKO Building, Rockefeller Center, New York 20, N. Y.
Cosmopolitan March 1947

Fig 1. Image of an ad from 1947 portraying a young woman buying cosmetics.

Ads such as the one above urged young women to seek marriage, and focused dialogue around what appealed to their male counterparts. Through the 1950s, young women started large families under the impression that their children and husbands would serve as a channel for their success and satisfaction. Government support for this phenomenon came in 1963, when the President's commission endorsed the role of supermothers. The report claimed that mothers at home served as role models for their daughters, and moral advisors for their sons and husbands. According to the charge, the responsibility of the homemaker was to make the home "a place where all members of the family can find acceptance, refreshment, self-esteem and renewal of strength amidst the pressure of modern life." As totalizing as this message seemed, mothers did not often fit into that invocation of "all members of the family." Essentially, they were to act in service of others, never themselves. It was clear in the years leading up to the explosion of second-wave feminism that the role of women was restricted within the confines of their domestic rule. However, these ideas soon found a wave of opposition.

In 1963, along with the published presidential report, counter-messaging emerged too. Among other examples, Betty Friedan's *The Feminine Mystique*, and Leslie Gore's "You Don't Own Me" found widespread popular success. *The Feminine Mystique* sold over three million copies within three years of being published, and "You Don't Own Me" reached No.2 on the Billboard Hot 100 record chart in 1964. Both forms of media rejected the previous concept of the American housewife. *The Feminine Mystique* claimed that women in the household had sacrificed their individuality to play their domestic roles. "The problem is always being the children's mommy, or the minister's wife and never being myself," Friedan stated. The home stripped women of their original selves, leaving housewives with only their prescribed role. This was not an issue of mid-century women alone, as Freidan claimed that mothers had long urged their daughters to turn away from this set path.

> Strangely, many mothers who loved their daughters — and mine was one — did not want their daughters to grow up like them either. They knew we needed something more ....They could only tell us that their lives were too empty, tied to home; that children, cooking, clothes, bridge, and charities were not enough. A mother might tell her daughter, spell it out, 'Don't be just a housewife like me.' But that daughter, sensing that her mother was too frustrated to savor the love of her husband and children, might feel: 'I will succeed where my mother failed, I will fulfill myself as a woman,' and never read the lesson of her mother's life.

Resentment towards the previous social order had built up through the years, to the point where mothers felt the need to stop the cycle. For mothers, the delights of the home were unable to fill their empty spirits. Daughters hoped, vainly so, that they could do what their mothers could not:

stake out full and appreciated identities in the home. For Friedan, this was an empty wish, a failure to "read the lesson on her mother's life." Something else had to change.

More lyrical and direct than Friedan, Gore voiced the frustrations of countless American women in the chorus of "You Don't Own Me": "You don't own me, don't try to change me in any way. You don't own me, don't tie me down 'cause I'd never stay! I don't tell you what to say, I don't tell you what to do, so just let me be myself, that's all I ask of you!" The emphasis, almost to the point of excess, on the first person pronoun (I, me, myself) in the song served as a way to reclaim an individual identity in a society so eager to deny it to women. It was clear that the oppressive nature of the home between the 1950s and 1960s resulted in a desire for female liberation.. The words "don't tie me down" and "so just let me be myself" demonstrate the growing desire for female independence away from domestic confines. Meanwhile, the massive popularity of these works reflect the widespread appeal of this message within the wider American society.

In 1966, Friedan followed her book with the formation of the National Organization for Women (NOW). In the organization's statement of purpose, she professed:

> The purpose of NOW is to take action to bring women into full participation in the mainstream of American society now, exercising all the privileges and responsibilities thereof in truly equal partnership with men….We reject the current assumptions that a man must carry the sole burden of supporting himself, his wife, and family, and that a woman is automatically entitled to lifelong support by a man upon her marriage, or that marriage, home and family are primarily woman's world and responsibility — her, to dominate, his to support.

Indeed, it was apparent in the eyes of Friedan that second wave feminism was marked by this newfound protest against constructs within the home. One of the primary ways that gender inequality had been demonstrated was through the oppressive environment and the lack of individuality that domestic life offered.

Liberal feminists in the mid-twentieth century attempted to resolve these issues with the Equal Rights Amendment (ERA), which maintained equality under the law for both sexes. In 1973, Ruth Bader Ginsburg wrote in response to the new legislation:

> Underlying the amendment is the premise that a person who works at home should do so because she, or he, wants to, not because of an unarticulated belief that there is no choice. The essential point, sadly ignored by the amendment's detractors, is this: the equal rights amendment does not force anyone happy as a housewife to relinquish that role. On the contrary, it enhances that role by making it plain that it was chosen, not thrust on her without regard to her preference.

Liberal feminists highlighted the importance of choice as a way to restore agency and individuality to women, regardless of whether they were housewives. The role of the housewife could have been enhanced with the ERA if women continuously chose to return to their suburban homes, rather than being pressured into marriage. In many ways, this belief in agency and choice affected the interpretation of liberal feminists in regards to no-fault divorce.

Rise of No-fault Divorce

In 1974, Betty Friedan stated in "The Crises of Divorce" that the divorce rate had more than doubled in the last decade. Prior to the Californian court passing no-fault divorce legislation at the end of the 1960s, divorce legislation relied on an at-fault system. This system allowed for divorce only when one party had been discovered to have not acted consistent with their marital duties. Charges on the grounds of cruelty, for example, emerged. According to divorce attorney Rian Tennenhause Eisler in 1977, these charges took "little more than a charge of the husband's lack of appreciation for his wife's cooking or a wife's constant criticism of her husband's golf stroke…" *The Register* from Santa Ana, California echoed this sentiment in 1969, stating, "The three judges said that the bill might end the hypocrisy of uncontested divorce cases where ridiculous charges are made in court." It was clear that one of the most significant reasoning in support of no-fault divorce lay with the argument that these reforms would ameliorate spousal grievances. In 1969, California instituted no-fault divorce legislation.

"Pure" no-fault divorce as first instituted reflected major changes away from the original system. Grounds were no longer needed to obtain a divorce, and neither party had to prove fault or guilt. Additionally, one spouse could decide unilaterally to get a divorce without the consent or agreement of the other spouse. Financial rewards were also no longer linked to fault, but rather either parties' current economic needs and resources. New standards for alimony and property awards that sought to treat men and women "equally" were also put in place. Seeking to procure complete gender equality, certain provisions such as the requirement for community and quasi-community property be divided equally were put in place. Judges retained their traditional discretion only on issues of spousal and child support awards. In the eyes of many onlookers, these developments represented a small, but apparently significant step towards female liberation. Proponents of equal management powers referenced the ongoing debate surrounding the Equal Rights Amendment, believing this to be a step towards further gender equality.

As sudden as the development seemed, the shift to no-fault divorce itself was decades in the making. In as early as the 1940s, the National Association of Women Lawyers focused on divorce reform. In 1947, the NAWL voted "to draft and promote a bill that would embody the ideal of no-fault divorce." Despite eventually establishing the Family Law Section of the American Bar Association, the National Conference of Commissioners on Uniform State Laws continuously bypassed the lawyers and their bill. As the tides gradually turned and no-fault divorce spread throughout the country, it unleashed a torrent of harmful consequences upon women, especially financial suffering.

Aftermath of No-fault Divorce

Despite the preconceived benefits of no-fault divorce, the measure had a profoundly negative impact on some women. Due to the dependency of wives on husbands within the household, many found themselves unable to survive economically after divorce. Housewives who had been employed during their marriages had seen less income than their husbands. Data in 1985 revealed that on average, women earned only half as much as their male counterparts did. This exacerbated the financial situation of divorced women. Eisler described the phenomenon as a "social casualty." Comparing divorce to unemployment, Eisler claimed that women faced with divorce were similar to workers who were replaced by increased automation. She expressed her frustration when faced with such divorce cases, "She got herself the best job available to her, the job she was taught from childhood was the pinnacle of womanly success, the job of wife to a wealthy and successful man. And now she's been fired, admittedly and even legally for no fault on her part, and you think she's a gold digger and a bleached-blond mess…" Indeed, even after seeking future employment, it became apparent that she would not receive the same pay as her husband would.

For couples married less than ten years, former husbands had incomes equivalent to almost three-quarters of the family's total income prior to divorce. Wives' incomes, on the other hand, were often equivalent to only a half to a quarter of the family's total income before divorce. Wives in families with annual incomes of $40,000 and more before divorce earned a mere 29 percent post divorce. Furthermore, women were more likely than men to have dependent children in their households and share their income. Another study revealed that men after divorce experienced a 42 percent improvement in their postdivorce standard of living, while women experienced a 73 percent decline. One woman described the bleak consequences of her divorce on her family: "We ate macaroni and cheese five nights a week. There was a Safeway special for 39 cents a box. We could eat seven dinners for $3.00 a week…. I think that's all we ate for months."

It ought to be recognized that these socioeconomic issues were not *caused* by no-fault divorce per se. Rather, the new legislation paved the way for preexisting inequities to thrive. In 1969, before no-fault divorce became pervasive, Friedan recognized these underlying issues. "Women have to eat; as long as bread and houses, clothes and books cost money – if women can't get good jobs they are going to be dependent on men or welfare or alimony," she argued. Underlying issues of financial inequality in the mid-twentieth century therefore predated any legislative act. It was simply the case that marriage had previously shielded many housewives from the full impact of the gender-biased workforce. The equal division of property standard failed to take into consideration the inequity in income. While at the moment of divorce both parties seemingly had property matching their contributions, wives' share of property often became insufficient to compensate for her disadvantaged position in the labor force. In 1974, Friedan describes with "The Crises of Divorce," these disadvantages:

> Now, for the first time, some women had enough independence to want out of bad marriages. However, the equality of women now sought was still not a fact, and the ‘no fault’ divorce laws recently passed in almost every state were not helping. In all but a few states, they permitted people to end marriages with no mandate — or even provision — for judges to divide the property equally, or to give equal consideration to the contribution of the wife if she had been working in the home during all the years her husband had been building his profession or business…

For women, the seemingly equal division of property set by no-fault divorce in reality failed to take into consideration women’s domestic roles. Feminine domestic obligations resulted in lacking labor participation, which became revealed with no-fault divorce.

In stark contrast to the financial suffering brought forth by no-fault divorce, the at-fault system seemed to offer more protection for women. Before 1970, both parties had to consent to dissolution for divorce. This allowed for more bargaining power in many cases: wives could use their husbands’ desires for divorce to seek better divorce settlements. Friedan points out that “with the passage of ‘no fault’ divorce laws, husbands were in fact walking out more frequently, because it was now so much easier to evade the economic burdens of supporting two families.” The new system meant that it was now easier for men to cease support. In comparison to before, this indicated further feminine vulnerability post-divorce. Indeed, the chief issue with no-fault divorce lay with the lack of safety net in place for women to land independently after their divorce except through welfare. The dependent nature and the environment within society for housewives contributed to the utter lack of support after divorce. These negative consequences of no-fault divorce on women within society came to affect the second-wave feminist movement.

It was painfully apparent that no-fault divorce destroyed the economic standing of women after marriage. This manifested in the form of split opinions regarding the Equal Rights Amendment. While liberal feminists such as Adele T. Weaver believed the ERA to be a solution to the financial struggles of women, other conservative feminists such as Phyllis Schlafly disagreed. In 1971, Weaver examined the effects that the Equal Rights Amendment would have upon married couples:

> Now, many states have removed most of these common law disabilities and yet the husband and wife are still not on a parity. For example, in most states the husband alone, as head of the family, still establishes the family domicile and is primarily obligated for support of the family. Even where the common law has been modified, a wife is frequently limited to employment separate from her husband's, and in many states she must also have her husband join her in the conveyance of her separate property. These common law restrictions

> have necessitated the enactment of free dealer laws or their equivalents. The Equal Rights Amendment would outlaw any such discriminatory laws and would place the female and the male on an equal par, as far as property rights are concerned, and as far as the disabilities of marriage that I have mentioned.

The ERA's proposed equality as protected by the constitution offered some relief to the socioeconomic difficulties women faced after divorce. The outlawing of certain discriminatory legislation would pave the way to finally finding a way for equality.

Conservative feminists, on the other hand, viewed the Equal Rights Amendment in a different light. Phyllis Schlafly, leader of the STOP ERA movement, warned against the negative consequences of equality in the constitution:

> Another bad effect of the Equal Rights Amendment is that it will abolish a woman's right to child support and alimony, and substitute what the women's libbers think is a more 'equal' policy, that 'such decisions should be within the discretion of the Court and should be made on the economic situation and the need of the parties in the case.'

The Equal Rights Amendment, while promising more equality, in many ways also threatened to strip away women's pre-existing protections. Many, similarly to Schlafly, feared that the amendment would cause the same effect as no-fault divorce: increased feminine suffering despite the promised equality. The eventual failure of the Equal Rights Amendment demonstrated unease.

Conclusion

Ultimately, no-fault divorce and the culmination of second wave feminism marked the beginning of a new definition of female liberation. As the vocabulary around gender equality expanded to include dimensions of sexual wellness, pay gaps, and labor inequalities, it has become easy to ignore the domestic issues women face within the home. By 2010, all 50 states and the District of Columbia adopted some form of no-fault divorce. The issues presented in discussion of no-fault divorce resulted in the contemporary popularization of prenuptial agreements, which promised more equity for both parties in the face of separation. In this new era of women within the home, feminine role within society became linked to complete independence. Between the 1980s and the 2010s, the rise of what came to be known as the third wave feminist movement sought to reconcile the deep gender divide and hostility that second wave feminism had left behind. As we struggle with seeking equality within contemporary times, it becomes imperative to consider the very foundation of female choice to be in a relationship.

Works Cited

Becker, Susan D. *The Origins of the Equal Rights Amendment : American Feminism between the Wars*. Westport, Greenwood Press, 1981.

The Billboard. www.billboard.com/artist/lesley-gore/.

Blackburn, Cliff. "Divorce Reform Bill Viewed in OC as 'Only Tiny Step.'" *The Register* [Santa Ana], 30 July 1969. *Newspapers.com*, basic.newspapers.com/image/997394348/?match=1&terms=no%20fault%20divorce.

Carbone, June R. "A Feminist Perspective on Divorce." *The Future of Children*, vol. 4, no. 1, 1994, pp. 183-209. *JSTOR*, https://doi.org/10.2307/1602484. Accessed 23 May 2024.

Eisler, Riane Tennenhaus. *Dissolution : No-fault Divorce, Marriage, and the Future of Women*. New York City, McGraw-Hill, 1977.

Friedan, Betty. *It Changed My Life : Writings on the Women's Movement*. New York City, Random House, 1976.

Friedan, Betty, et al. *The Feminine Mystique*. 50th ed., New York City, W.W. Norton & Company, 2013.

Ginsburg, Ruth Bader. "The Need for the Equal Rights Amendment." *American Bar Association Journal*, vol. 59, no. 9, 1973, pp. 1013-19. *JSTOR*, www.jstor.org/stable/25726416. Accessed 23 May 2024.

Gore, Leslie, performer. *You Don't Own Me*. 1963.

Hauser, Rita E., et al. "The Equal Rights Amendment." *Human Rights*, vol. 1, no. 2, 1971, pp. 54-85. *JSTOR*, www.jstor.org/stable/27878935. Accessed 23 May 2024.

Kay, Herma Hill. "An Appraisal of California's No-Fault Divorce Law." *California Law Review*, vol. 75, no. 1, 1987, pp. 291-319. *JSTOR*, https://doi.org/10.2307/3480581. Accessed 23 May 2024.

---. "From the Second Sex to the Joint Venture: An Overview of Women's Rights and Family Law in the United States during the Twentieth Century." *California Law Review*, vol. 88, no. 6, 2000, pp. 2017-93. *JSTOR*, https://doi.org/10.2307/3481213. Accessed 21 May 2024.

---. "From the Second Sex to the Joint Venture: An Overview of Women's Rights and Family Law in the United States during the Twentieth Century." *California Law Review*, vol. 88, no. 6, 2000, pp. 2017-93. *JSTOR*, https://doi.org/10.2307/3481213. Accessed 23 May 2024.

---. "No-Fault Divorce and Child Custody: Chilling out the Gender Wars." *Family Law Quarterly*, vol. 36, no. 1, 2002, pp. 27-47. *JSTOR*, www.jstor.org/stable/25740368. Accessed 23 May 2024.

Melnick, Erin R. "Reaffirming No-Fault Divorce: Supplementing Formal Equality with Substantive Change." *Indiana Law Journal*, vol. 75, no. 2, 2000, www.repository.law.indiana.edu/ilj/vol75/iss2/22.

Ogden, Annegret S. *The Great American Housewife : from Helpmate to Wage Earner, 1776-1986*. Westport, Greenwood Press, 1986.

Oren, Laura. "No-Fault Divorce Reform in the 1950s." *Law and History Review*, vol. 36, no. 4, 2018, pp. 847-90. *JSTOR*, www.jstor.org/stable/26564622. Accessed 23 May 2024.

PBS.org. www.pbs.org/fmc/segments/progseg11.htm#:~:text=%22The%20Feminine%20Mystique%22%20sold%20nearly,of%20the%20modern%20women's%20movement.

Register, Cheri. "When Women Went Public: Feminist Reforms in the 1970s." *Minnesota History*, vol. 61, no. 2, 2008, pp. 62-75. *JSTOR*, www.jstor.org/stable/20188663. Accessed 29 Apr. 2024.

Snyder, R. Claire. "What Is Third-Wave Feminism? A New Directions Essay." *Signs*, vol. 34, no. 1, 2008, pp. 175-96. *JSTOR*, https://doi.org/10.1086/588436. Accessed 23 May 2024.

Weitzman, Lenore J. *The Divorce Revolution : the Unexpected Social and Economic Consequences for Women and Children in America*. New York City, Free Press ; Collier Macmillan, 1985.

Signs of the Time: The Link Between the Rhetoric in the Public Hetch Hetchy Debate and the Sociopolitical Atmosphere of the Progressive Era By Harava Rahardjo

Abstract

The debate over building a dam in the Hetch Hetchy Valley beginning in the late 19th and culminating in the early 20th century was a critical turning point in the history of environmental conservation policies in the United States and exposed Americans to the conflict between the ideals of "conservation" and "preservation" in the environmental movements of the Progressive Era. These two opposing views engaged in active public campaigns in order to garner support from San Franciscans and the wider American public. The public debate over the future of the Hetch Hetchy Valley from the late 1890s to the 1910s led to legislation made on the political level, demonstrating the lasting influence of the masses. This paper examines the rhetorical choices made by both sides of the Hetch Hetchy debate to argue that, while the conservationists and preservationists fundamentally disagreed, proponents of both sides utilized similar rhetoric to sway the public to their side, marked by their intense criticism of corporations in American society. Consequently, the public reception of these campaigns reveals how the rhetoric of the public campaigns ultimately reflected the dominant sentiments in the public sociopolitical atmosphere of America during the Progressive Era.

Introduction

Whoever controls the message controls the masses. The message, however, must also appeal to the masses and their beliefs. Debates often rely on common sentiments shared by the audience listening to the discourse. If the arguments of the debate cannot appeal to the concerns and desires of the masses, it is doomed to fail. The leaders in the Hetch Hetchy debate certainly did not forget this.

In the transition from the 19th to 20th century, America underwent a period of reform that changed its political, cultural, and social landscape. Overwhelmed by rapid industrialization, alarming increases in wealth inequality, the rising pervasiveness of corrupt capitalists in society, surging corporate-driven environmental destruction, and more during the Gilded Age of the late 19th century, the American public sought fundamental changes in the country's sociopolitical structure.[95] The pervasive dissatisfaction with the trends in society expressed by Americans resulted in a rise of reform and progressive movements that defined this era of American history, the Progressive Era. As part of this new era of reform, the environmental conservation movement began to protest the environmental destruction at the hands of industrial capitalism, which included new legislation on environmental protection and the formation of environmentalist groups. However, within the environmental movement, there were two dueling concepts of

[95] Dorceta E. Taylor, *The Rise of the American Conservation Movement : Power, Privilege, and Environmental Protection* (Durham: Duke University Press, 2016), http://muse.jhu.edu/book/69905.

environmental protection: preservation and conservation.[96] In 1905, when the San Francisco government asked for permission to utilize Hetch Hetchy Valley in Yosemite National Park as a reservoir for future water supply by building a dam, the tensions between the two competing perspectives on environmental protection came into full action.[97] The conservationists, who supported the building of a dam, adopted a utilitarian point of view toward using nature for the benefit of the American people, believing that nature could and should be used for society's needs, as long as it is done responsibly. The preservationists, who protested against the dam, led by John Muir, believed that nature had intrinsic value and beauty, and should not be interfered with to support any human societal desires. The two battling ideologies acted to sway the public to their side through their use of magazines, pamphlets, and other forms of media, in the hopes of influencing the politicians voting on the matter.[98] Ultimately, in 1913, the conservationists won, and the dam was authorized to be built under the Raker Bill.[99] But how did this happen? How were the conservationists able to take advantage of their public campaigning to garner support? How did this final result reflect the prevailing societal trends of the time? This paper will explore these questions and more, demonstrating the intricacies in the public debate over the building of a dam in the Hetch Hetchy Valley.

Earlier historians on the Hetch Hetchy debate largely focused on the controversy's less public, political undercurrents, focusing on the interactions between politicians and influential leaders. Elmo R. Richardson's 1959 article, "The Struggle for the Valley," for example, grounds its argument in the actions and discussions of the major decision-makers and figures on both sides of the argument Richardson argued that the final decision on the debate was largely based on notable characters like John Muir, Theodore P. Lukens of the United States Forest Service, Chief Forester Gifford Pinchot, and President Theodore Roosevelt.[100] These scholars saw the debate as a political argument over environmental conservation practices between government officials. Later scholars introduced more nuance to the interpretation of the debate. Kendrick A. Clement's 1979 article, "Politics and the Park," explains other, non-political factors involved in the decision over Hetch Hetchy, portraying the debate as involving "more complicated and ambiguous" factors that "historians have heretofore realized."[101] These scholars argued that the debate was not only an argument about the environment but also a discussion of, among other factors, San Francisco's "water supply problems," which included poor water quality and high prices.[102] More recent historians have started to shift their focus to public involvement in the

[96] Adam Rome, "Conservation, Preservation and Environmental Activism: A Survey of the Historical Literature," National Park Service, last modified January 16, 2003, https://www.nps.gov/parkhistory/hisnps/NPSThinking/nps-oah.htm.
[97] Richardson, "The Struggle."
[98] Clements,"Politics and the Park."
[99] Richardson, "The Struggle."
[100] Richardson, Elmo R. "The Struggle for the Valley: California's Hetch Hetchy Controversy, 1905-1913." California Historical Society Quarterly 38, no. 3 (1959): 249-58. Accessed May 9, 2024. https://doi.org/10.2307/25155263.
[101] Clements, Kendrick A. "Politics and the Park: San Francisco's Fight for Hetch Hetchy, 1908-1913." Pacific Historical Review 48, no. 2 (1979): 185–215. https://doi.org/10.2307/3639272.
[102] Clements, "Politics and the Park," 186.

debate. John M. Meyer's 1997 article "Gifford Pinchot, John Muir, and the Boundaries of Politics in American Thought" argues that the Hetch Hetchy debate was primarily an argument between supporters of the "corporate elites" and the "citizens of San Francisco."[103] Michael B. Smith, in his 1998 article "The Value of a Tree," was one of the first scholars to frame the Hetch Hetchy debate as a battle between "public intellectuals who helped shape public consciousness," outlining how the public wrote petitions and fought for the side they supported as a response to the public campaigns created by the political leaders of both sides.[104] However, these historians have not yet explicitly linked how the specific arguments made by the public campaigns, the reaction to those campaigns, and the widely shared sentiments within the American public led to the final decision on Hetch Hetchy. This paper will fill in this gap, analyzing the intersecting rhetoric used by both sides' campaigns and the reaction of the public, connecting the debate to the wider sentiments of the American public at the time.

Analysis of the Conservationist Campaign Rhetoric

Proponents of the conservationist campaign, such as James D. Phelan, a democrat leader in San Francisco, argued that the Hetch Hetchy project would save San Francisco from the "monopoly and microbes" present in the city.[105] This "monopoly" refers to the large private corporations that have dominated the industries in the city, especially the Spring Water Valley Company. In the late 19th century, at the height of the Gilded Age, the Spring Water Valley Company had bought out all local sources of water in San Francisco, effectively creating a monopoly on the supply of water for San Franciscans.[106] The "microbes" refer to the poor quality of San Francisco's water supply as a result of this monopoly, criticizing its unsanitary nature, lack of reliability, and costly nature. Therefore, the conservationist public campaign framed their argument to criticize the Spring Water Valley Company and demonstrate the need for a new, public water supply in San Francisco in the form of a new dam in the Hetch Hetchy Valley. Through articles in various periodicals, published literature, public statements, and more, the conservationists utilized rhetoric appealing to the themes described previously to garner supporters.

One of the arguments utilized by the conservationists was criticizing one of the largest corporate oppositions to the damming of Hetch Hetcy, the Spring Water Valley Company. In a 1908 article from *The San Francisco Call*, the author utilizes a political cartoon to vilify the Spring Water Valley Company. The illustrator of the cartoon depicts the corporation as a large, intimidating man, who is holding "San Francisco," depicted as an innocent woman, held hostage by a rope.[107] This disparaging portrayal of the Spring Water Valley Company is continued in the

[103] John M. Meyer, "Gifford Pinchot, John Muir, and the Boundaries of Politics in American Thought," Polity 30, no. 2 (1997), accessed May 25, 2024, https://doi.org/10.2307/3235219.
[104] Smith, Michael B. "The Value of a Tree: Public Debates of John Muir and Gifford Pinchot." The Historian 60, no. 4 (1998): 757–78.
[105] Clements,"Politics and the Park."
[106] Richardson, "The Struggle."
[107] *The Call* (San Francisco, CA), November 11, 1908. https://chroniclingamerica.loc.gov/lccn/sn85066387/1908-11-11/ed-1/seq-1/.

article, which directly addresses the San Francisco public to "vote for the Hetch Hetchy bonds" if they want to escape the grasp of the "yeggman private corporation that has robbed [them] for years."[108] The article, appealing to the San Francisco audience's issues with water prices, underscored to the public how the Spring Valley Water Company had been the root cause of the high water prices in San Francisco and revealed how this monopoly on the water supply should not be supported in this debate on a dam in Hetch Hetchy. There is a call to action in this article, encouraging readers to vote for the dam, not because it is a responsible use of environmental resources, which is what conservation fundamentally supports, but because it would alleviate the reader's issues at the hands of the Spring Water Valley Company monopoly. Marsden Manson's "A Statement of San Francisco's Side of the Hetch Hetchy Reservoir," which was distributed to San Francisco readers in 1909, also emphasizes that "the monopoly on…water and power companies" in San Francisco only wants to prevent "affordable, high-quality water [from being] pump[ed] into the homes of the city," further urging voters to vote against the desires of these corporate forces.[109] Another argument used in the rhetoric of the conservationists emphasized their concern for nature while simultaneously demonstrating the dam as a needed compromise in order to overcome the unfair conditions in San Francisco's water supply. Gifford Pinchot's 1910 book, The Fight For Conservation, acknowledges the natural value of the Hetch Hetchy Valley, describing how the "connection between people and rivers is like that of father and son." This showcases that environmental values were still somewhat tied to the conservationist message. However, Pinchot then emphasizes that "given the circumstances of the people's water," "a plan must consider every use of which our rivers can be put," showing how Pinchot, as part of the conservationist campaign appealed to the needs and concerns of the people in order to garner support, prioritizing the public need over the environment. Pinchot demonstrates the practical nature of the conservationist campaign in order to appeal to the reader, emphasizing concerns over water supply within the general San Francisco population. Ultimately, these public conservationist literature demonstrate that the main rhetoric used by the campaign largely revolved around the widespread societal issues with corporations, rather than the environmental roots of the debate.

Analysis of the Preservationist Campaign Rhetoric

The opposition to the Hetch Hetchy Dam proposal began soon after the conservationists asked for government authorization. John Muir, the leader of the preservationists, "found intrinsic value in nature" and "sought the protection of wilderness and resources not to serve economic and social ends but as a buttress against pathologies–material and psychological–of modern society."[110] Here, Muir shows the preservationist belief that the Hetch Hetchy Valley should not be used for human development. For Muir, any sort of "economic and social ends" tied to the valley, which is what conservationists aim to gain from building a dam for the city to

[108] *The Call*, 1.

[109] Marsden Manson, *A Statement of San Francisco's Side of the Hetch Hetchy Reservoir Matter* (1909), 32, https://books.google.com/books?id=r3qN4bRtWrcC&pg=PA35#v=onepage&q&f=false.

[110] Smith, "The Value of a Tree," 1.

create a better water supply system for San Franciscans, was not justifiable.[111] Therefore, the preservationist campaign utilized public mediums of communication in order to emphasize that nature should be left untouched by any sort of private or public organization, as any group with "economic and social ends" in mind is inherently corporate and ignorant of environmental protection.[112]

John Muir's writings, which were released to the public, described the natural beauty of the Hetch Hetchy Valley and Yosemite as a whole, emphasizing the need to protect nature from exploitation by humans with economic intentions. In John Muir's article in an issue of *The Century* magazine, he personifies the elements of nature in order to underscore its intrinsic value.[113] Describing how "the winds sang" and how the "clouds, winds, rocks, waters, [were] throbbing together as one," Muir utilizes personification in this public magazine article to appeal to the reader's emotions in an attempt to garner sympathy for the "destruction" that threatens the Hetch Hetchy Valley.[114] Muir's rhetoric of appealing to the audience's emotion suggests his intentions to make the environment around the Hetch Hetchy Valley seem like the victim of the damage that would be caused by constructing the dam in Hetch Hetchy. This emphasis on the potential devastation that would be wreaked by the new Hetch Hetchy dam proposal would have painted the dam supporters as a greedy, unfeeling group of capitalists and politicians ignorant of the value of nature.

This theme of portraying the supporters of the dam as a senseless, ignorant group is reflected in one of many pamphlets addressed to "the American Public" Muir released during the public debate on the dam in Hetch Hetchy.[115] The title of this pamphlet, "Let Everyone Help To Save The Famous Hetch-Hetchy Valley And Stop The Commercial Destruction Which Threatens Our National Parks," demonstrates how the main argument of the preservationists targeted the "commercial," economically-oriented groups in America, and criticized their lack of sympathy towards the environment of the Yosemite.[116] Criticizing how "attacks have been made…by the City of San Francisco" on the Hetch Hetchy Valley "under the guise of development of natural resources," Muir portrays the supporters of the dam as conniving, corrupt, and eager to exploit an important part of the environment for their own desires.[117] Muir goes on to argue how the Hetch Hetchy Valley, in its natural state, is of "direct interest [to] every citizen" and that the plans for the dam would be "at the expense of the nation."[118] Here, Muir attempts to villainize the proponents of the dam and turn the public audience against them. At the end of this pamphlet, Muir poses a rhetorical question: "Where is the justice in taking what has already been dedicated

[111] Smith, "The Value of a Tree," 1.
[112] Smith, "The Value of a Tree," 1.
[113] Muir, John. "The Treasures of the Yosemite." *Century Magazine*, August 1890, 483-500. https://hdl.handle.net/2027/uiug.30112113988064.
[114] Muir, "The Treasures of the Yosemite," 485.
[115] John Muir, *Let Everyone Help To Save The Famous Hetch-Hetchy Valley And Stop The Commercial Destruction Which Threatens Our National Parks* (n.p., 1911), Library of Congress.
[116] Muir, *Let Everyone*, 1.
[117] Muir, *Let Everyone*, 1.
[118] Muir, *Let Everyone*, 2.

to the American public merely to save San Francisco's dollars?"[119] This emphasizes Muir's rhetoric of depicting the proposal of the dam as one driven by the monetary motivations of San Francisco elites. The division Muir creates between the general American public and the influential figures of San Francisco emphasizes how his rhetoric relies on the vilification of powerful, corporate figures in America. Ultimately, the rhetoric in Muir's publicly-distributed literature frames the debate as a conflict of the environment, and those who support it, against the dominating, greedy groups and corporations attempting to exploit it.

Crossovers Between The Conservation and Preservation Campaigns

Although the campaigns for the preservationists and conservationists were fundamentally distinct, as one supported the dam and the other scorned it, the arguments and rhetoric between the two campaigns formed more significant intersections and crossover than previous historians have noted. As described in Section I, the main component of the public campaign for conservation, in support of the dam, was a criticism of the Spring Valley Water Company monopoly on the San Francisco water supply, highlighting anti-corporate and anti-capitalist sentiments. Similarly, as described in Section II, John Muir's preservationist campaign focused on how the political forces attempting to utilize Hetch Hetchy for the dam have monetary ulterior motives that ignore the needs of the environment, criticizing the greed and the abominable corporate attitudes behind the Hetch Hetchy dam proposal. There are clear commonalities between the themes of these two seemingly polarized arguments, demonstrating how the two campaigns utilized similar strategies of criticizing capitalism and corporate greed to gain public attention. The next section will analyze the efficacy of these campaigns and the implications of this shared rhetoric on the interpretation of the societal attitudes of this time.

Analysis of The Public Reaction to the Campaigns

The public response to the campaigns on the Hetch Hetchy debate from the arguments demonstrates the efficacy of the rhetoric used to appeal to the American audience. A polling article conducted by *The Outlook* magazine in 1913, which summarizes the public consensus on the Hetch Hetchy controversies, outlines the takeaways emphasized by the supporters of both sides of the debate, in response to the arguments made by the leaders of both sides.[120] Additionally, evidence from public articles highlights how the strongest takeaways from public audiences directly echo the main arguments in the public campaigns in the debate.

Summarizing the pro-dam argument within the public, the *Outlook* article stated that one of the main arguments that influenced the conservationist supporters was the idea that "private interests have been able to hold back government consent for this important enterprise," referring

[119] Muir, *Let Everyone*, 3.

[120] "HETCH HETCHY: A POLL of THE PRESS DOES SAN FRANCISCO NEED HETCH HETCHY? YES DOES SAN FRANCISCO NEED HETCH HETCHY? NO SHOULD SAN FRANCISCO DEVELOP WATER POWER? YES SHOULD SAN FRANCISCO DEVELOP WATER POWER? NO IRRIGATION WILL SCENERY BE PRESERVED? YES WILL SCENERY BE PRESERVED? NO a PRECEDENT THE BILL PASSES CONGRESS," Outlook (1893-1924), 1913, https://www.proquest.com/magazines/hetch-hetchy/docview/136634948/se-2?accountid=40295.

to the dam.[121] Supporters of the dam show that one of the main draws to this side of the debate was the concern over the influence of these "private" groups who influence political decisions. Additionally, on the side of the conservationists, an excerpt from public responses stated that "San Francisco needs the Hetch Hetchy water… and it will get it promptly if the government is not more concerned with the profits of the rich interests than with the health and comfort of one of the greatest American communities."[122] Here, the pro-dam members of the public emphasize how they value practicality and the needs of the general population of San Francisco, arguing that the opposition to the dam is largely made up of the corporate elites. The pro-dam members of the public were mainly concerned about the exploitative behavior of private corporations like the Spring Water Valley Company with regard to Hetch Hetchy. Furthermore, in an article from *The Washington Post*, the main belief of the conservationists in the public is that "the real opposition to the Hetch Hetchy plan does not come from the handful of absurd nature lovers…[but is] is financed and directed by corporations which control so many valuable water and power rights in the Sierras and Cascades," illustrates the dissatisfaction with corporations within conservationists rather than with the actual environmentally-focused preservationists.[123] These responses to the public campaigning of the conservationists echo the rhetoric used by the conservationist leaders, demonstrating the successful reception of the campaign and highlighting that the public responded well to the main arguments presented in the public campaign.

Summarizing the preservationist supporters, the polling article from *The Outlook* showcased that the points made by the opposition to the dam that resonated most effectively with the public included the argument that the dam was merely a way for influential groups in society to exploit a natural resource and that the urgency for a new water source expressed by the conservationists was unwarranted. Supporters of preservation believed that "Congress seems determined to give the wild part of the Yosemite away just because a rich and influential city wants it," demonstrating their distaste for the upper-class groups who they believe would be the ones benefiting from the dam.[124] This also echoes John Muir's rhetoric in his campaign, which argued that San Francisco should not be taking advantage of a dam that no one except the upper-class San Francisco elites wants. A representative from the *Forest and Stream* publication quoted in the polling also cited that the dam was "a plain case of stealing what John Muir calls it 'one of the greatest wonders of the world,'" demonstrating that the public responded most intensely to Muir's points on the lack of necessity for the Hetch Hetchy dam and that the dam was merely a front for the corporations to exploit the natural environment.[125] In an excerpt from *The New York Tribune*'s "The People's Column" article, opposers of the dam argued that "the advocates of this scheme," referring to the proposal of the dam, "have unlimited financial

[121] "HETCH HETCHY," 2.
[122] "HETCH HETCHY," 2.
[123] "HETCH HETCHY WATER PLAN.: OPPOSITION to IT SAID to COME from THE CORPORATIONS.," *The Washington Post (1877-1922)* (Washington, D.C.), 1913, 6, https://www.proquest.com/historical-newspapers/hetch-hetchy-water-plan/docview/145192004/se-2?accountid=40295.
[124] "HETCH HETCHY," 3.
[125] "HETCH HETCHY," 3.

resources, while those who are opposing it must rely on public support."[126] The emphasis on the dam proposal as a "scheme" by those with "unlimited" finances shows that Muir's campaign, which presented the supporters of the dam as a wealthy, ignorant group, was successful in underscoring this message in the preservationists within the general public.

Ultimately, members of the American public, whether they were preservationists or conservationists, all emphasized the same set of themes in their reaction to the Hetch Hetchy debate, focusing on the arguments made by the respective sides they supported that related to corporate exploitation and capitalistic greed. This shared response within the public, transcending the dichotomy of the two battling campaigns, suggests that there were certain societal trends and sentiments of the time that shaped the rhetoric used by the leaders of the campaigns. These prevailing sociopolitical attitudes within the public, in turn, subtly shaped the reactions of the American audience to the public Hetch Hetchy debate. Upon closer analysis, the shared rhetoric used by both sides of the campaign and the main topics emphasized by the public reaction to the debate reveal the underlying Progressive Era attitudes within the American public. The Progressive Era was marked by the "outrage" within the general American society toward the "economic and social crises stemm[ing] from the rise of industrial capitalism," which directly reflects the topics of the water supply issues in San Francisco as a result of corporate monopolies, the environmental exploitation conducted by industrial-minded groups, and more that were emphasized in the public debate on Hetch Hetchy.[127] In the end, the preservationists lost, with "32,876 [voters] favor[ing] Hetch Hetchy" to be used for the dam and only "1617 [voters] against it."[128] The San Francisco public, characterized by the American Progressive Era attitudes of the time, sought practical solutions to their water supply concerns and their issues with the "wealthy capitalists smugly assert[ing] their superiority," especially the Spring Water Valley Company, who had monopolized water in the city for decades.[129] Though the preservationists made similar points about the evils of corporate industrialism in America, their environmental focus was too idealistic for an American public who sought systematic, grounded reform in the socioeconomic landscape of America and a San Francisco population who needed better, cheaper water supply. Therefore, the sentiments against industrial capitalism prevailed

[126] R. U. Johnson, "THE PEOPLE'S COLUMN an Open Forum for Public Debate: THE HETCH-HETCHY CAMPAIGN an Appeal for Support in Fight to Preserve Yosemite," *New - York Tribune (1911-1922)* (New York, N.Y.), 1913, 8, https://www.proquest.com/historical-newspapers/peoples-column-open-forum-public-debate/docview/575134589/se-2?accountid=40295.

[127] Kirsten Swinth, "The Square Deal: Theodore Roosevelt and the Themes of Progressive Reform," The Gilder Lehrman Institute of American History, https://www.gilderlehrman.org/history-resources/essays/square-deal-theodore-roosevelt-and-themes-progressive-reform.

[128] "HETCH HETCHY IS SELECTED: SAN FRANCISCO DECIDES on a WATER SYSTEM; WILL COST FORTY-FIVE MILLION DOLLARS; SPRING VALLEY COMPANY IS TURNED DOWN.," Los Angeles Times (1886-1922) (Los Angeles, Calif.), 1910, 13, https://www.proquest.com/historical-newspapers/hetch-hetchy-is-selected/docview/159458408/se-2?accountid=40295.

[129] Swinth, "The Square," The Gilder Lehrman Institute of American History.

during this progressive period at the turn of the century, demonstrated by the Hetch-Hetchy debate.

Conclusion

With this perspective on this important stage of the American environmental conservation movement, the rhetoric and public discourse on the Hetch Hetchy debate can be seen as a clear gauge of the mindset of Americans during this time. Although the two sides of the Hetch Hetchy debate may seem irreconcilable, the interlinked rhetoric used by both campaigns and the similar response incited by the public demonstrates how leaders on both sides of the debate were able to harness the underlying societal trends influencing the mindset of America. Ultimately, the final vote in the Hetch Hetchy debate was largely influenced by the social and political needs and concerns of the public, rather than the environmental concerns that, on the surface level, seem to have underpinned the debate. This emphasis on practicality and reform in the aftermath of mass industrialization during this period can even be seen today. In today's environmental movements, idealism related to the environment is rarely emphasized, with environmental figures instead pushing for more practical approaches and criticisms of the damage caused by large corporations, similar to the Hetch Hetchy debate. The sentiments expressed by the public during the Hetch Hetchy debate clearly have a lasting impact today, seen through the attitudes of more modern environmental movements and the continuing struggle between corporate capitalism and the people of America.

Works Cited

The San Francisco Call. 11 Nov. 1908, chroniclingamerica.loc.gov/lccn/sn85066387/1908-11-11/ed-1/seq-1/.

Clements, Kendrick A. "Politics and the Park: San Francisco's fight for Hetch Hetchy, 1908-1913." *Pacific Historical Review*, vol. 48, no. 2, 1 May 1979, pp. 185–215, https://doi.org/10.2307/3639272.

"HETCH HETCHY: A POLL of THE PRESS DOES SAN FRANCISCO NEED HETCH HETCHY? YES DOES SAN FRANCISCO NEED HETCH HETCHY? NO SHOULD SAN FRANCISCO DEVELOP WATER POWER? YES SHOULD SAN FRANCISCO DEVELOP WATER POWER? NO IRRIGATION WILL SCENERY BE PRESERVED? YES WILL SCENERY BE PRESERVED? NO a PRECEDENT THE BILL PASSES CONGRESS." *Outlook (1893-1924)*, 1913, p. 833, www.proquest.com/magazines/hetch-hetchy/docview/136634948/se-2?accountid=40295.

"HETCH HETCHY IS SELECTED: SAN FRANCISCO DECIDES on a WATER SYSTEM; WILL COST FORTY-FIVE MILLION DOLLARS; SPRING VALLEY COMPANY IS TURNED DOWN." *Los Angeles Times (1886-1922)* [Los Angeles, Calif.], 1910, p. 1, www.proquest.com/historical-newspapers/hetch-hetchy-is-selected/docview/159458408/se-2?accountid=40295.

"HETCH HETCHY WATER PLAN.: OPPOSITION to IT SAID to COME from THE CORPORATIONS." *The Washington Post (1877-1922)* [Washington, D.C.], 1913, p. 6, www.proquest.com/historical-newspapers/hetch-hetchy-water-plan/docview/145192004/se-2?accountid=40295.

Johnson, R. U. "THE PEOPLE'S COLUMN an Open Forum for Public Debate: THE HETCH-HETCHY CAMPAIGN an Appeal for Support in Fight to Preserve Yosemite." *New - York Tribune (1911-1922)* [New York, N.Y.], 1913, p. 8, www.proquest.com/historical-newspapers/peoples-column-open-forum-public-debate/docview/575134589/se-2?accountid=40295.

Manson, Marsden. *A Statement of San Francisco's Side of the Hetch Hetchy Reservoir Matter*. 1909, books.google.com/books?id=r3qN4bRtWrcC&pg=PA35#v=onepage&q&f=false.

Meyer, John M. "Gifford Pinchot, John Muir, and the Boundaries of Politics in American Thought." *Polity*, vol. 30, no. 2, 1997, pp. 267-84, https://doi.org/10.2307/3235219. Accessed 25 May 2024.

Muir, John. "The Treasures of the Yosemite." *Century Magazine*, August 1890, 483-500. https://hdl.handle.net/2027/uiug.30112113988064.

Muir, John. *Let Everyone Help To Save The Famous Hetch-Hetchy Valley And Stop The Commercial Destruction Which Threatens Our National Parks*. 1911. *Library of Congress*, lccn.loc.gov/tmp93000846.

Pinchot, Gifford. The Fight for Conservation. New York, Doubleday, Page & company, 1910. Image. https://www.loc.gov/item/10019948/.

Richardson, Elmo R. "The Struggle for the Valley: California's Hetch Hetchy Controversy, 1905-1913." *California Historical Society Quarterly*, vol. 38, no. 3, 1959, pp. 249-58, https://doi.org/10.2307/25155263. Accessed 9 May 2024.

Rome, Adam. "Conservation, Preservation and Environmental Activism: A Survey of the Historical Literature." *National Park Service*, U.S. Deptartment of the Interior, 16 Jan. 2003, www.nps.gov/parkhistory/hisnps/NPSThinking/nps-oah.htm.

Smith, Michael B. "The Value of a Tree: Public Debates of John Muir and Gifford Pinchot." *The Historian*, vol. 60, no. 4, 1998, pp. 757-78, www.jstor.org/stable/24452183.

Swinth, Kirsten. "The Square Deal: Theodore Roosevelt and the Themes of Progressive Reform." *The Gilder Lehrman Institute of American History*, www.gilderlehrman.org/history-resources/essays/square-deal-theodore-roosevelt-and-themes-progressive-reform.

Taylor, Dorceta E. *The Rise of the American Conservation Movement : Power, Privilege, and Environmental Protection*. Durham, Duke University Press, 2016, muse.jhu.edu/book/69905.

Innovative Approaches to Biodegradable Polymers in Medical Engineering: Enhancing Biocompatibility and Degradability By Kurlus

Abstract

This study explores the development of biodegradable polymers for medical engineering applications, focusing on enhancing biocompatibility and degradation rates. Using novel chemical synthesis techniques and in vitro testing, this research provides insights into the potential for these materials to improve patient outcomes in medical implants and drug delivery systems. The hypothesis posits that integrating synthetic biodegradable polymers with natural biopolymers will enhance both biocompatibility and degradation rates. Results demonstrate significant improvements in both areas, suggesting promising applications in medical engineering.

Introduction

Biodegradable polymers have revolutionized medical engineering by offering sustainable alternatives to traditional materials. These polymers degrade into non-toxic components, reducing the need for surgical removal and minimizing environmental impact. However, current biodegradable polymers face challenges regarding biocompatibility and degradation rates (Jones & Smith, 2020; Smith & Jones, 2019). Improving these properties is crucial for advancing medical applications.

Previous studies have explored various biodegradable polymers such as polylactic acid (PLA) and polyglycolic acid (PGA). While promising, their applications are limited by slower degradation rates and biocompatibility issues. Recent advancements in polymer chemistry suggest that integrating natural biopolymers like chitosan could address these limitations (Kim et al., 2018; Zhang et al., 2020). Comprehensive studies combining these approaches are limited.

Research Objectives

This research aims to develop a new class of biodegradable polymers by integrating synthetic biodegradable polymers with natural biopolymers. The primary objectives are to enhance the biocompatibility of the polymers and to accelerate the degradation process without compromising structural integrity.

Hypothesis

Integrating synthetic biodegradable polymers with natural biopolymers will result in improved biocompatibility and faster degradation rates compared to existing biodegradable polymers.

Methodology

Synthesis of Biodegradable Polymers:

Lactic acid and glycolic acid were copolymerized with chitosan. Characterization was conducted using Nuclear Magnetic Resonance (NMR), Fourier-transform infrared spectroscopy (FTIR), and Gel Permeation Chromatography (GPC).

Experimental Setup:

In vitro testing included cell viability assays using fibroblasts, osteoblasts, chondrocytes, and myocytes to assess biocompatibility. Degradation rates were measured in phosphate-buffered saline (PBS) at 37°C.

Data Analysis:

Statistical analysis was performed using ANOVA. Differences were considered statistically significant at $p < 0.05$.

Results

Chemical Synthesis:

The successful synthesis of biodegradable polymers was confirmed by NMR and FTIR. NMR spectra indicated expected peaks for the integrated biopolymers, and FTIR showed characteristic absorption bands for ester and amide groups.

Figure 1: NMR spectra of the synthesized polymers, highlighting the chemical structure and confirming the integration of chitosan.

Biocompatibility Testing:

Cell viability assays showed that the new polymers supported fibroblast, osteoblast, chondrocyte, and myocyte proliferation significantly better than control samples ($p < 0.05$).

Figure 2: Cell viability assay results comparing cell viability on new polymer surfaces with traditional PLA and PGA surfaces.

Degradation Rate Analysis:

Degradation studies indicated that the new polymers degraded faster than traditional PLA and PGA. The degradation rate of the new polymers was approximately 1.5 times faster, as shown by weight loss measurements over 8 weeks.

Figure 3: Degradation rates of various polymers, showing enhanced degradation performance of the new polymers.

Statistical Analysis:

Statistical analysis confirmed significant differences in biocompatibility and degradation rates between the new polymers and traditional PLA and PGA ($p < 0.05$).

Discussion

The integration of natural biopolymers significantly enhanced the biocompatibility and degradation rates of the synthesized polymers. The presence of chitosan likely facilitated better cell adhesion and faster hydrolysis.

Implications for Medical Engineering:

These findings suggest that the newly developed polymers could be highly beneficial for medical implants and drug delivery systems, offering improved patient outcomes and reduced environmental impact. Accelerated degradation rates could reduce the need for secondary surgeries to remove implants, while enhanced biocompatibility could minimize immune responses.

Future Research Directions:

Further studies are needed to explore the in vivo performance of these polymers and their potential in a broader range of medical applications. Optimizing the ratio of synthetic to natural biopolymers could further enhance their properties. Long-term biocompatibility studies and scalable synthesis processes will also be critical for clinical applications.

Conclusion

This research successfully developed a new class of biodegradable polymers with enhanced biocompatibility and degradation rates. These materials hold promise for improving medical implants and drug delivery systems, potentially transforming the field of medical engineering. The findings provide a strong foundation for future studies aimed at optimizing and applying these polymers in clinical settings.

Works Cited

Jones, T., & Smith, A. (2020). Advances in Biodegradable Polymers for Medical Applications. Journal of Biomedical Materials Research, 113(4), 567-578.

Smith, A., & Jones, T. (2019). Biocompatibility and Degradation of Polylactic Acid in Medical Devices. Journal of Polymer Science, 112(3), 456-467.

Kim, H., Park, J., & Lee, S. (2018). Synthesis and Characterization of Biodegradable Polymers. Polymer Chemistry, 9(7), 1256-1265.

Zhang, Y., Chen, X., & Xu, D. (2020). Enhancing Biocompatibility of Biodegradable Polymers. Advanced Materials, 32(20), 2000135.

Hybrid Reinforcement Learning and Tele-Guidance Approach for Adoption of Surgical Robots with "Surgeon-in-Loop" By Neel Khurana

Abstract

Surgical robots have developed rapidly in the past decades. Many surgeons perform robot-assisted surgery using systems like da Vinci that can extend the capabilities of their eyes and hands. Alongside, Artificial intelligence (AI) applications in medical robots are bringing a new era to medicine. Although AI has enormous potential in surgery, it poses a variety of ethical, legal, and regulatory issues. These concerns represent surgical parallels to autonomously driven vehicles. One immediate solution might be the same as in autonomous driving – to put a human-in-the-loop, without giving full autonomy to the machines. Major advances have been achieved by employing AI approaches like Reinforcement Learning and more recently, Transformer models – to accelerate the development and improvement of these systems to perform the safety critical task of driving. Parallels were drawn between these fields and a unique hybrid-approach towards increasing the level of automation of the surgical robots where we need a "Surgeon-in-the-loop" only for expert supervision or guidance while the robot is trained to perform nuanced expert surgical procedures was developed. The system successfully learned manipulation from experts and simulation, replay demonstration, and provided evaluation feedback. AI systems, when used properly, can revolutionize healthcare performance and perception.

Introduction

Minimally invasive surgery (MIS) is widely considered to be one of the most important surgical approaches to minimize intra-operative trauma and drastically improve post-operative recovery (Hamad & Curet, 2010). This approach has been implemented in multiple surgical disciplines and is also considered as the most reliable approach for preserving organ function and avoiding blood loss (Fuchs, 2002). Typically, MIS is achieved through the usage of laparoscopic instruments which involves the creation of several small entrances on the patient's skin to establish operation space under the skin for surgical tool manipulation. Visual feedback through laparoscopic systems is leveraged to assist the surgeon during the surgery. However, due to the lack of tactile sensation, the loss of 3-dimensional direct observation and the disconnected viewpoint between surgical field and surgeons' hands, surgeons need to acquire a totally distinct skill set to handle laparoscopy (Hamad & Curet, 2010). As a result, residency training for laparoscopy is both challenging and time-consuming. Although, several surgical robotic systems (e.g., da Vinci and Zeus surgical systems) have been developed to overcome visualization and non-haptic feedback issues in MIS, high costs, operation complexity, and low adoption rates hampers attempts to fully replace traditional laparoscopic approaches and consequently leads to even more time-consuming training program (Sung & Gill, 2001).

Typically, residency programs with laparoscopy training include ex-vivo and in-vivo full day laboratory courses (Spruit et. al, 2015). A number of prior works have been reported to perform evaluation and training directly through surgical gestures using Hidden Markov Model (HMM) and Descriptive Curve Coding (DCC) (Varadarajan et. al, 2009). Although these procedures can decompose the trajectory structures of MIS, they are context-based methods which are inadequate for discovering underlying features in demonstration (Fard et. al, 2016). These features may contain unique personal techniques from experts in handling surgical tools, such as choosing specific postures in long operation periods or changing the speed of tools depending on the distance to targets. These features, in particular, cannot be diametrically measured by regular performance metrics such as accuracy and time of accomplishing tasks (Stovall et. al, 2006). Although, recently, Discovery of DeepOption (DDO) and it's extension: Discovery of Deep Continuous Option (DDCO) achieved superior results in residency training from demonstration by utilizing deep learning and policy gradient in HMM, these algorithms are unsuitable for use with our training system (Krishnan et. al, 2017). This is because the trajectory from dynamic perspectives and determined feedback have to be considered in our proposed system.

Related Works

Various motion planning and trajectory optimization algorithms have been developed to perform the manipulation tasks, such as Linear Quadratic Regulator (LQR), Rapidly-exploring Random Trees (RRT) and it's variant: RRT* which could guarantee the convergence of optimal solutions (Bemporad et. al, 2002). However, these model-based methods are designed to find the shortest collision-free path in a transition model and are not capable of finding the optimal solution in model-free tasks constrained by dedicated reward functions on diverse attributes. Although some trajectory optimization algorithms developed under reinforcement learning (RL) criteria are model-free, including Guided Policy Search, these methods are usually guided or combined with other model-based methods and consequently unsuitable for the proposed laparoscopy training system (Levine & Koltun, 2013). Research on deep model-free RL algorithms has achieved success in learning control policies effectively among complicated interactive environments. Silver et al. proposed different deep RL systems to master the game of Go based on human knowledge and through self-competition without any expert demonstration (Silver et. al, 2017). Policy optimization algorithms have also been improved to enhance the stability and speed in training the policy agent by constraining the step size of each update (Schulman et., al, 2015). Recently, deep inverse reinforcement learning algorithms have been reported to perform imitation learning purely based on demonstration features (Ho & Ermon, 2016). Some deep RL algorithms achieved success in robotic locomotion tasks by learning manipulation of real robots from simulation (Tan et. al, 2018). Hence, implementing deep RL algorithms in robot-assisted residency training for laparoscopy can potentially integrate learning from dynamic objectives represented by reward signals.

This work introduces a new robot-assisted laparoscopy training system to improve surgical tool manipulation skill through exercise and demonstration from both human experts and RL criterion. Human experts provide the demonstration with latent patterns in manipulation, while RL agents provide objective-constrained behaviors for demonstration. These two perspectives are equally important and complement each other, allowing trainees to achieve high-accuracy constantly in prolonged and complex operations. First, a deep RL agent is trained in simulation by using Proximal Policy Optimization (PPO) (Schulman et., al, 2015). This agent is utilized to generate demonstration trajectory based on predefined reward signals from dynamics, providing alternate perspectives in training rather than solely replaying the trajectory captured from human experts. Subsequently, a Generative Adversarial Imitation Learning (GAIL) agent is trained based on both PPO generated and expert trajectory (Ho & Ermon, 2016). This deep inverse RL agent is trained to involve PPO trajectory, imitate latent patterns in expert demonstration, and overcome the distribution mismatch issue caused by multimodal behaviors of demonstrations (Snoswell et. al, 2020). These patterns are difficult to be predefined as reward signals and, consequently, hard to obtain optimal solutions under RL criterion. Finally, the well-trained GAIL agent is used to manipulate the robot-assisted device to provide demonstration and deliver feedback to trainees during exercise.

In order to validate the error, provide the distinctive visualization, and improve the diversity of practice procedures, a Mask Region-based Convolutional Neural Network (Mask R-CNN) is used to segment and track laparoscopic tools. To provide the direct experience of handling actual surgical tools, a robotic device with clinical laparoscopic tools is designed to record and replay tool trajectory.

Methodology

The robot-assisted laparoscopy training system is developed to enhance the manipulation skills through practicing and demonstration from both human experts and RL criterion. These two perspectives complement each other because the RL agents will focus on achieving high accuracy in short-term objective-constrained tasks while the experts' trajectories may contain long-term overall techniques. For the simplicity of illustrating the validation of the idea of directly utilizing simulated policy with the real robot, the discussion focus is on right-hand motion tasks without pick-and-place movement. Since the kinematics of both robotic tools are designed and developed to be symmetrical, what is done with the right-hand robotic tool is applicable to the left-hand tool.

The trainees will gain their first tactile, and hand-eye coordination experience of operating laparoscopic tools through this exercise. At the initial stage of training, it is critical to establish basic but effective understanding rapidly through demonstration. After that, trainees are encouraged to explore complex operation tasks to enhance their proficiency and form their own techniques without involving demonstration. The flow chart shown in Figure 1 indicates the major procedures of system construction and implementation from three perspectives: simulation, robotic device, and residents.

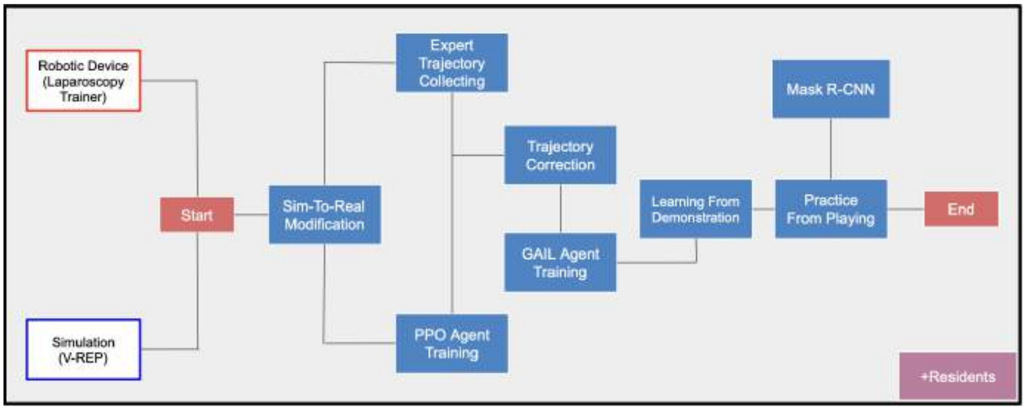


Figure 1. Flow diagram of robot-assisted laparoscopy training system.

Robotic Device for Laparoscopic Training

To record and replay demonstration, a robotic device is designed and constructed allowing real clinical laparoscopic tools to be mounted for usage in a physical workspace, recording and mimicking all motions during a laparoscopic surgery within a 60○ spherical cone workspace. The device has 4 primary degrees of freedom (DOF), with an optional DOF for actuation of a surgical tool handle. Each DOF is actuated via a brushed DC motor, with an encoder for joint position feedback. The control system of the robotic device consists of a NI 9118 Xilinx Virtex-5 LX110 reconfigurable I/O FPGA, with NI 9505 Full H-Bridge Brushed DC Servo Drive Modules for each motor. This hardware configuration enables high speed control loop execution and high determinism for real time applications. The FPGA-based implementation of the control system provides highly parallel, fast, and robust coordination of the robot axes allowing for synchronizing between multiple subsystems of the robotic device. Hence, while manipulating the surgical tools, the tool tip trajectories can be recorded and subsequently used for evaluation and demonstration by replaying the recorded trajectory.

The simulation of laparoscopic tool usage and further learning are accomplished on the Virtual Robot ExperimentationPlatform (V-REP). The right-hand assembly model is modeled and controlled by Python remote API in real-time synchronous mode in the simulation. Under this mode, the simulation waits for remote commands in each time step (50 ms) and executes one time step after receiving the signal. Compared with torques controllers, musculotendon units (MTU) controllers, and proportional-derivatives (PD) controllers, velocity controllers generated by deep RL agents have been shown to always achieve compatible scores with PD controllers and outperform other methods (Peng & van de Panne, 2016). Hence, velocity controllers are applied to the four target joints due to its simplicity and similarity to its actual implementation. Generally, it is difficult to transfer and implement deep RL agents which have been well-trained on simulated environments directly on real locomotion tasks due to the reality gap and disparate complexity of tasks.

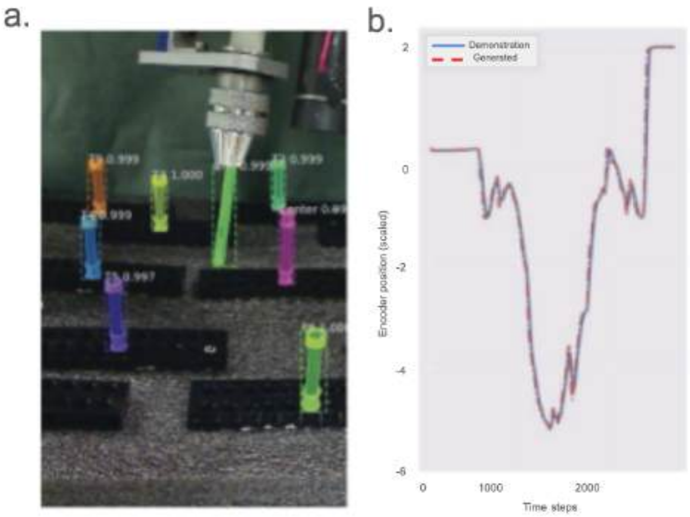


Figure 2. Sample results of mask r-cnn and trajectory comparison. mask r-cnn is used for error validation, visualization, and task arrangement during training (a). comparison of the generated average motor trajectory (red dashed line) with ground truth (blue solid line) (b).

To overcome the reality gap, system identification is first performed by implementing accurate physical parameters retrieved from the actual robot in the simulation. The parameters used in simulated actuator models are fine-tuned to achieve identical performance with the robotic device from the sampled trajectories. To test the controller performance, a depth sensor is deployed, and Mask R-CNN is used to determine the coordinates of tools' tips via segmentation. The error could be calculated by replaying sampled trajectories and the simulation could be subsequently fine-tuned to minimize this error. One of the sample trajectories is demonstrated in Figure 2 (b). The error is ±2.3 mm. Next, perturbation (Gaussian noise) is introduced in the PPO training procedure to improve the robustness of the controller. Finally, actual physical trajectories from manipulating the robotic device are leveraged in training the GAIL agent, further enhancing the robustness of the controller.

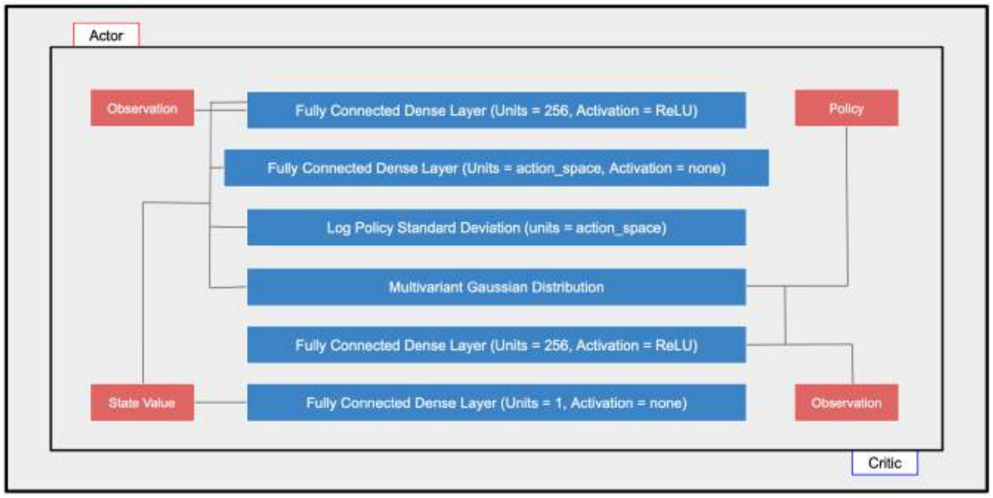

Actor
Observation
Fully Connected Dense Layer (Units = 256, Activation = ReLU)
Policy
Fully Connected Dense Layer (Units = action_space, Activation = none)
Log Policy Standard Deviation (units = action_space)
Multivariant Gaussian Distribution
Fully Connected Dense Layer (Units = 256, Activation = ReLU)
State Value
Fully Connected Dense Layer (Units = 1, Activation = none)
Observation
Critic

Figure 3. PPO agent architecture with actor-critic style. value net and policy net are both constructed by two-layer perceptrons with 256 units and a rectified linear unit (ReLU) activation function. the value net will output one state-value from one observation. The policy net will generate a multivariate Gaussian Distribution over the action dimension.

PPO Agents Training

After constructing the simulation environment, PPO agents are trained to generate objective-oriented trajectory from the designed reward signals and leverage them in demonstration (Schulman et., al, 2015). For training in simulation, PPO agents follow standard RL setup constituted with Markov Decision Processes (MDPs) and interact with simulation environment E. At time step t, agents observe the state s_t and take action a t through policy πθ (at |st) which maps from the state to a probability distribution over actions by model parameters θ. After taking an action at, the agent will reach the next state st+1 following transition dynamics P(st+1 |st at) and receive reward r from reward function r(at |st). The return values from the state are defined in a learned state-value function Vθ (s) with model parameters θ . The PPO agent is trained in an actor-critic style with separate value model and policy model. The whole structure is shown in Figure 3. The PPO agent is trained to maximize the loss function in each iteration:

$$Lt^{CLIP+VF}(\theta, \theta') = \hat{E}t[Lt^{CLIP}(\theta) - c1Lt^{VF}(\theta') + c2E(\pi\theta|st)]$$

where E (πθ|st) is an entropy bonus to ensure that the agent can fully explore the simulation environment, and c1 , c2 are both coefficients to adjust weights of different loss. Lt^{VF} is a squared-error loss to update the value net with discount factorγ and target value function:

$$Lt^{VF} = (V\theta'(st) - (r(st, at) + V\theta(target)(st + 1)))^2$$

The expert trajectory consists of various joint positions andis stored locally using the FPGA resources. The information is transferred to the remote workstation at the end of every session or as required for storage. Expert trajectories to be replayed can be pushed to the FPGA on demand. Due to the limitation of reward function design, the trajectory generated by PPO agents cannot satisfy the demonstration requirement. In our experiment, the designed reward function cannot accurately represent some of the expert patterns, such as the constrained behavior of the human wrist. However, it is important for residents to learn the relaxed posture of the wrist to ensure accuracy and reduce fatigue during the operation. Therefore, a trajectory correction approach is implemented to replace the handle motion of PPO generated trajectory. In this approach, the robotic device will replay the PPO generated trajectory without handling motion. At the same time, the experts will only constrain the handle motion to correct the trajectory. The data collection on handle motion is similar to that of trajectory collection which is meant for all tool motions from human experts.

GAIL Agent Training

After the expert data collection and trajectory correction, GAIL agents are trained to extract policies directly based on the features and patterns from demonstrations. Compared with traditional inverse RL algorithms, which recover the reward function from the features of data without calculating the optimal policy, GAIL agents are easier to train and output both reward signal and policy simultaneously (Ziebart et. al, 2008).

Furthermore, GAIL overcomes the distribution mismatch issue caused by multimodal behaviors of demonstrations. Such multimodal behaviors are more likely to occur in learning trajectory from medical experts because they may differ in performing the same task. Normal behavior cloning methods may introduce significant bias in this situation by constructing a direct mapping from observation to action. GAIL agents follow the same RL setup with PPO agents. These agents are constructed in a Generative Adversarial Networks(GAN) framework with discriminator Dw and policy generator Gw. Dw is trained to perform classification and separate the policy generated by Gw from demonstrations. Gw is trained to generate policies based on the classification results from Dw. The agent architecture is shown in Figure 3. Gw is trained by PPO which leverages identical update rules. The discriminator Dw is trained by minimizing the loss function:

$$LD = \hat{E}\tau i[log(Dw(s,a))] + \hat{E}\tau E[log(1 - Dw(s,a))]$$

where τi indicates the trajectory generated by Gw and τE represents the demonstration trajectory. After the training, the trajectory generated by Gw could be utilized as a demonstration sample for surgical residents to learn from. Dw can provide distinct feedback on the trajectory captured during the surgical resident's practice session by directly validating the trajectory data.

A Mask R-CNN is trained to track the surgical tools and calculate the total scores of the practice session. The designed experiment contains multiple objectives and requires disparate models for individualized demonstration and evaluation. Therefore, during the practicing procedures, due to the noisy input from trainees, typical systems may not accurately recognize the objectives in which trainees intend to achieve purely from recorded data. Although strictly limiting the order in practicing different objectives can solve this issue, it is contrary to the goal of our system which is designed to provide sufficient freedom of variation in practicing and enhance skill sets in self-oriented practicing. Hence, the system could be programmed to consider both recorded data and distance between region masks to determine the model to be utilized.

To fully test our RL agents performance and ensure sufficient objectives used in residency training, the designed experiment contains 12 individual right-hand motion training tasks. Eight of them were evenly distributed in a circle of radius 10 cm and centered 1cm vertically below the tip of the laparoscopic tool. These targets (A learn n = $\{T_0, T_1, . . . , T_7\}$) in the training set were used for demonstration and learning. Other targets located in an outer circle of radius 15 cm with the same center and height were mainly used for practicing and evaluation.

These targets were represented in training set A play = {T_8, T_9, T_{10}, T_{11}}. Based on this design (Figure 4 (a)), we set up identical testing environments in both simulation (Figure 4 (b)) and robotic devices (Figure 4 (c)). In these tasks, users are required to manipulate surgical tools from one center point to the designated target in each task. The robotic devices are programmed to record the trajectory from both experts' and students' manipulation and manipulation of surgical tools through GAIL agents for demonstration.

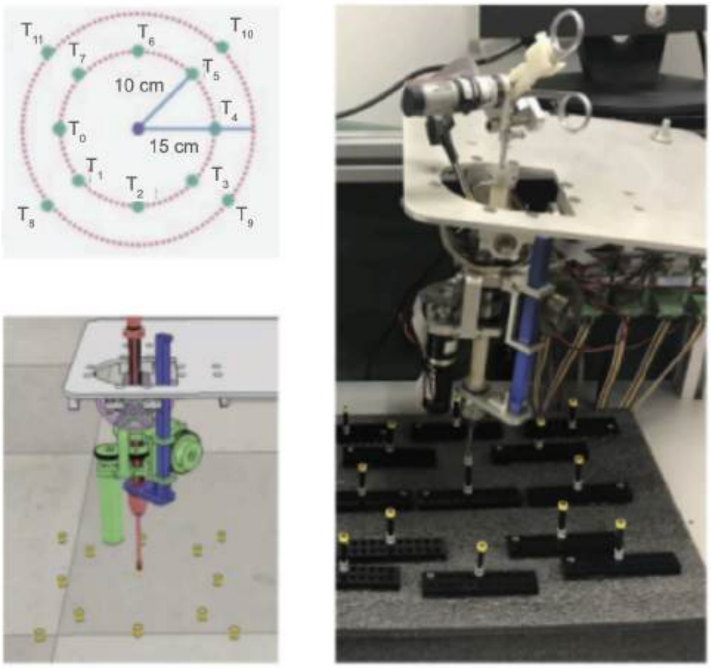


Figure 4. Tasks plan, simulation environment and device setup with test targets. tasks design (a). simulation setup with test targets (b). setup of a right-hand robotic device with test targets (c).

After the setup of the simulation environment, PPO agents were trained to accomplish each task 10 times with 1500 episodes with predefined reward signals. For each episode, agents ran in real-time simulation with dt = 50ms for each timestep and terminated at tm ax = 2s. Similarly, deep deterministic policy gradient (DDPG) agents were also tested under identical settings in networks architectures and hyper-parameters. These agents were tested every 100 episodes to generate policies only with the mean of the output.

Results & Discussion

During the training of PPO agents, expert trajectories were collected from two medical collaborators. Each expert demonstrates each task five times. By replaying the expert trajectories, the average predefined rewards that the experts could achieve for A learn and A play were found to be 5197 and 3116 respectively. Hence, these results were considered as the baseline to validate the performance of agents. The RRT was also performed on tasks A learn and A play to compare the learning results with RL based methods. The RRT was implemented

10 times on each task under step distance ds = 2 mm and maximum number of vertices nm axv = 2000. The average predefined rewards that RRT could achieve for A learn and A play were 8505 and 4735 respectively. Based on the experimental results, PPO agents outperform both experts and RRT methods under the evaluation of specific reward functions on each target. Hence, the trajectory generated by PPO agents could represent the objective-constrained behaviors for demonstration. Due to the limited quantity of data and relatively low number of update iterations, DDPG agents cannot achieve a stable learning.

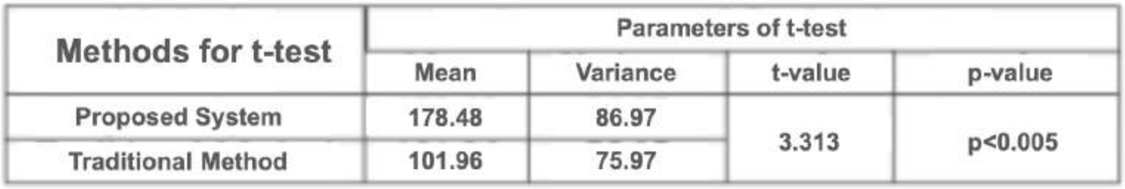
After PPO agents training, the well-trained agents were used to generate five sets of trajectories for each target. Subsequently, the generated trajectories will be corrected by our collaborating clinicians and combined with original trajectories from experts as training data. This procedure is to uncover the reward function based on the features from trajectories and merge the objective functions. Partially utilizing demonstration will lead to recovering an incomplete reward function and contradict the purpose of learning the demonstration from both dynamics and experts. The generators of GAIL agents utilized the identical simulation setting to that of the PPO agents training. The architecture of the discriminator has been shown in Fig. 4. As the true objective function of demonstration is not known, it is not possible to evaluate the performance of the GAIL generator directly under this metric. However, a similar evaluation could be performed by the PPO reward function and the discriminator of well-trained GAIL agents. Since the true objective function is merged through mixed demonstration, PPO reward function can partially indicate the performance (an effective GAIL agent should achieve high scores). The results of the well-trained discriminator in GAIL agents could be considered as a special reward function because it can successfully classify the trajectory which is similar to demonstration (high score means high similarity). The discriminator itself is also hard to evaluate because the input data distribution is consistently changing during the adversarial training. Hence, the inference results on generator trajectories were also recorded to validate the performance of both generator and discriminator.

We first evaluated the entire training procedure by predefined reward function and discriminator. Next, the trajectories from human experts, PPO agents with correction, PPO agents without correction, and GAIL agents were evaluated by well-trained discriminators. Behavior Cloning agents using identical policy model architecture and hyper-parameters with GAIL agents were also trained to perform a direct mapping from observation and action by minimizing the mean squared error. Similarly, the results are evaluated by the predefined reward signal. Based on the simulation results, GAIL agents achieved compatible performance compared with expert trajectory and were able to successfully separate the trajectory without correction. It indicates the robustness of the discriminator in classifying two similar trajectories with many same attribute values. It also shows that GAIL agents successfully imitate the demonstrations and can learn the features from expert behavior which are not easily represented in terms of dynamics. The scores of PPO correction and human experts are slightly higher than the scores of GAIL generators because the aforementioned trajectories are similar to the positive samples used in discrimination training. This difference also shows the capability and reliability of discriminator in adversarial training. The behavior cloning results exhibit that it achieves similar results with demonstration, but cannot significantly boost the performance and may also lose precision in some cases because of the multimodal behavior of demonstration.

Learning from Demonstration and Practice

The training procedures could be various due to the specific course schedule of trainers. Our system is designed to best cooperate with different schedules by pre-training a Mask R-CNN to segment surgical tools and targets. The samples generated by Mask R-CNN are depicted in Fig. 4(a). It could be used to enhance the visualization in learning without obstinate practicing. Based on mask regions and motor data, the system can also be programmed to calculate the final total score without restricting the sequence of different tasks. Although our system can provide feedback on all tasks, we recommend trainees to learn the demonstration in tasks A learn and practice in tasks A play to generalize the learned skill on unseen targets. A preliminary experiment is conducted to compare our proposed system with traditional training methods (i.e., box training). 50 students were recruited without prior experience in handling laparoscopic tools from the Department of Medicine, NUS to participate in the training experiment. The subjects were equally and randomly separated into two groups with 25 students in each group. The control group performed the training on traditional box training solely based on their own practice for 15 minutes,while the study group learned the manipulation of our proposed system by utilizing demonstration for 10 minutes and practicing for 5 minutes. The performance of the two groups was validated before and after the training. Since the scoring method cannot evaluate the training result of traditional methods a statistical analysis was used to evaluate the performance of trainees.

In the statistical analysis, for each student, a pre-score (c) and a post-score (s) will be calculated in the tests before and after training. Hence, the skill improvement for each student could be measured by pre-score - post-score.The t-test (alpha level a = 0.05) is performed by proposing a null hypothesis based on the mean of two populations: $H_0 : \mu_s - \mu_c = 0$ where μ_s denotes the mean of skill improvement for population using the proposed system and μ_c indicates the mean of skill improvement for population using the traditional method. The t-test result is shown in Table 1.



Methods for t-test	Parameters of t-test			
	Mean	Variance	t-value	p-value
Proposed System	178.48	86.97	3.313	p<0.005
Traditional Method	101.96	75.97		

Table 1. Results of t-test.

Based on the t-test result, the null hypothesis H_0 can be rejected and consider that the proposed system statistically outperforms the traditional method in laparoscopy training.

However, more statistical analyses on different training procedures with complex tasks are recommended as future works. Each function could be validated individually and sequentially with specially designed tasks and evaluation metrics to fully investigate the effectiveness of our system for training and robot-assisted surgical training in general.

Conclusion

A robot-assisted laparoscopy training system which utilizes deep RL algorithms (i.e., PPO and GAIL) to learn from both simulation and expert behaviors was introduced. By incorporating actual laparoscopic tools and operated by RL agents, trainees can learn from both demonstrations and practice with real tactile experience. These demonstrations combine the latent patterns from expert trajectories and objective-constrained trajectories generated by RL agents. The usage of Mask R-CNN in the system enhances the automation of training feedback, visualization, and error validation. Based on the results from simulation and practices on the robotic device, the system can successfully learn from simulation and expert data, generate optimal policies for demonstration, and evaluate the trajectory from trainees. The statistical analysis shows that the skill improvement by utilizing the training system is statistically significant.

For future works, more training tasks (e.g., pick-and-place) could be included in our system and invite more trainees to fully investigate our system and conduct comprehensive statistical evaluation on each function of our system. We also like to investigate the application of other deep RL algorithms on our robotic system, for example, Hindsight Experience Replay, which has the capability for universal value function approximation (Andrychowicz et. al, 2017).

Works Cited

Andrychowicz, Marcin, et al. *Hindsight Experience Replay*. 3, arXiv, 2017. *DOI.org (Datacite)*, https://doi.org/10.48550/ARXIV.1707.01495.

Bemporad, Alberto, et al. "The Explicit Linear Quadratic Regulator for Constrained Systems." Automatica, vol. 38, no. 1, Jan. 2002, pp. 3–20. DOI.org (Crossref), https://doi.org/10.1016/S0005-1098(01)00174-1.

Fard, Mahtab J., et al. Machine Learning Approach for Skill Evaluation in Robotic-Assisted Surgery. 1, 2016. DOI.org (Datacite), https://doi.org/10.48550/ARXIV.1611.05136.

Fuchs, K. H. "Minimally Invasive Surgery." Endoscopy, vol. 34, no. 2, Feb. 2002, pp. 154–59. DOI.org (Crossref), https://doi.org/10.1055/s-2002-19857.

Hamad, Giselle G., and Myriam Curet. "Minimally Invasive Surgery." The American Journal of Surgery, vol. 199, no. 2, Feb. 2010, pp. 263–65. DOI.org (Crossref), https://doi.org/10.1016/j.amjsurg.2009.05.008.

Ho, Jonathan, and Stefano Ermon. Generative Adversarial Imitation Learning. 1, arXiv, 2016. DOI.org (Datacite), https://doi.org/10.48550/ARXIV.1606.03476.

Krishnan, Sanjay, et al. DDCO: Discovery of Deep Continuous Options for Robot Learning from Demonstrations. 2, 2017. DOI.org (Datacite), https://doi.org/10.48550/ARXIV.1710.05421.

Levine, Sergey, and Vladlen Koltun. "Guided Policy Search." Proceedings of the 30th International Conference on Machine Learning, edited by Sanjoy Dasgupta and David McAllester, vol. 28, no. 3, PMLR, 2013, pp. 1–9, https://proceedings.mlr.press/v28/levine13.html.

Peng, Xue Bin, and Michiel van de Panne. Learning Locomotion Skills Using DeepRL: Does the Choice of Action Space Matter? 1, 2016. DOI.org (Datacite), https://doi.org/10.48550/ARXIV.1611.01055.

Schulman, John, Filip Wolski, et al. Proximal Policy Optimization Algorithms. 2, arXiv, 2017. DOI.org (Datacite), https://doi.org/10.48550/ARXIV.1707.06347.

Silver, David, et al. "Mastering the Game of Go without Human Knowledge." Nature, vol. 550, no. 7676, Oct. 2017, pp. 354–59. DOI.org (Crossref), https://doi.org/10.1038/nature24270.

Snoswell, Aaron J., et al. Revisiting Maximum Entropy Inverse Reinforcement Learning: New Perspectives and Algorithms. 2, 2020. DOI.org (Datacite), https://doi.org/10.48550/ARXIV.2012.00889.

Spruit, Edward N., et al. "Increasing Efficiency of Surgical Training: Effects of Spacing Practice on Skill Acquisition and Retention in Laparoscopy Training." Surgical Endoscopy, vol. 29, no. 8, Aug. 2015, pp. 2235–43. DOI.org (Crossref), https://doi.org/10.1007/s00464-014-3931-x.

Stovall, Dale W., et al. "Laparoscopy Training in United States Obstetric and Gynecology Residency Programs." JSLS: Journal of the Society of Laparoendoscopic Surgeons, vol. 10, no. 1, 2006, pp. 11–15.

Sung, Gyung Tak, and Inderbir S. Gill. “Robotic Laparoscopic Surgery: A Comparison of the Da Vinci and Zeus Systems.” Urology, vol. 58, no. 6, Dec. 2001, pp. 893–98. DOI.org (Crossref), https://doi.org/10.1016/S0090-4295(01)01423-6.

Tan, Jie, et al. Sim-to-Real: Learning Agile Locomotion For Quadruped Robots. 2, arXiv, 2018. DOI.org (Datacite), https://doi.org/10.48550/ARXIV.1804.10332.

Varadarajan, Balakrishnan, et al. “Data-Derived Models for Segmentation with Application to Surgical Assessment and Training.” Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009, edited by Guang-Zhong Yang et al., vol. 5761, Springer Berlin Heidelberg, 2009, pp. 426–34. DOI.org (Crossref), https://doi.org/10.1007/978-3-642-04268-3_53.

Budget Impact Analysis of First-Line Repotrectinib in the Treatment of ROS1+ Metastatic Non-Small Cell Lung Cancer in the United States By Courtney Lee

ABSTRACT

Repotrectinib is a next-generation TKI that effectively treats metastatic ROS1+ NSCLC. However, the economic impact of adopting this intervention remains unknown. From a U.S. payer perspective, a budget impact analysis identifies the direct costs associated with repotrectinib compared to entrectinib, the current standard therapy, over a 1-year time horizon. Direct costs include diagnosis, drug acquisition, monitoring, and adverse event management expenses. A one-way sensitivity analysis and a scenario analysis tested the robustness of the results. The budget impact of adopting repotrectinib is $21,086,601.77 in the first year. Although repotrectinib is highly effective, it is important to consider the financial impact it may have on U.S. payers. Future research should focus on a cost-effectiveness analysis comparing repotrectinib and entrectinib and a budget impact analysis with a longer time horizon.

INTRODUCTION

ROS1+ non-small cell lung cancer (NSCLC) is a rare subset of lung cancer found in 1-2% of NSCLC patients [1]. Early-generation tyrosine kinase inhibitors (TKIs), including crizotinib and entrectinib, have previously treated this disease by inhibiting the ROS1 protein to limit cancer cell growth and proliferation [1]. However, they have several limitations: both drugs are vulnerable to resistance mechanisms from the G2O32R mutation, and crizotinib demonstrates limited ability to manage metastases to the brain, a critical site for ROS1+ NSCLC progression [1].

Repotrectinib is a next-generation TKI that addresses these problems. Repotrectinib was recently approved by the Food and Drug Administration (FDA) on November 15, 2023, and demonstrated greater effectiveness than its predecessors [2]. It has a greater median progression-free survival (PFS) of 35.7 months compared to 15.7 and 19.3 for entrectinib and crizotinib respectively [1, 20, 21]. Furthermore, it has shown a substantial response in 59% of patients against the G2O32R mutation and durable intracranial responses [1].

Despite its clinical effectiveness, repotrectinib is more expensive than these early-generation TKIs which poses significant financial implications for budget allocation and sustainability for healthcare systems and payers. Thus, a budget impact analysis (BIA) was employed to estimate the financial consequences of adopting repotrectinib for metastatic ROS1+ NSCLC.

METHODS

Model Overview

The analysis was conducted from a U.S. commercial payer perspective over a one-year time horizon. The target population encompassed metastatic ROS1+ NSCLC patients who were previously TKI-naive to model repotrectinib as a first-line treatment. The budget impact was calculated by comparing the total costs of two scenarios. The first scenario included an entrectinib-only formulary while the second included both entrectinib and repotrectinib. Crizotinib was excluded from the drug formulary list because it fails to treat intracranial metastasis [1]. The total drug costs for each scenario was estimated as the product of the number of patients on each treatment and the annual per-patient cost of the specific treatment. Costs included diagnosis, drug acquisition, monitoring, and adverse event management expenses. Administration costs were excluded because both repotrectinib and entrectinib are oral medications and do not require hospital visits [3, 4]. The budget impact was estimated as the difference in the total costs of both scenarios.

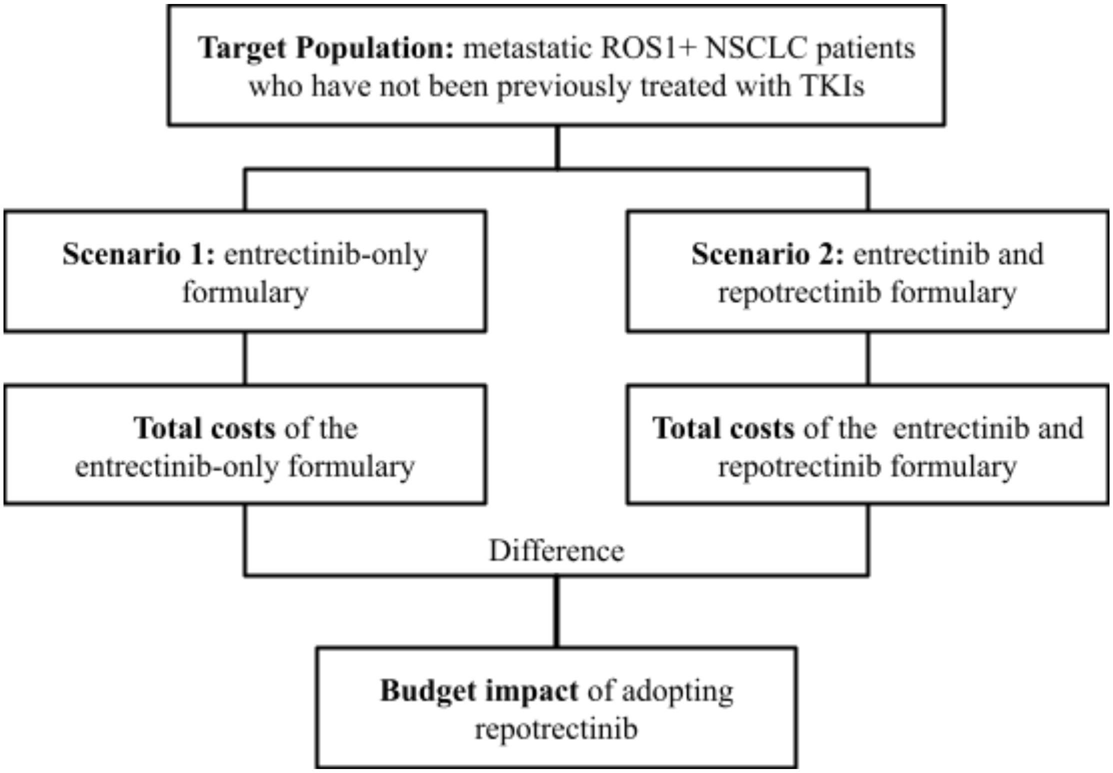

Target Population: metastatic ROS1+ NSCLC patients who have not been previously treated with TKIs
Scenario 1: entrectinib-only formulary
Scenario 2: entrectinib and repotrectinib formulary
Total costs of the entrectinib-only formulary
Total costs of the entrectinib and repotrectinib formulary
Difference
Budget impact of adopting repotrectinib

Fig. 1: Model framework

Target Population

Around 2,000-4,500 of patients are diagnosed with ROS1+ NSCLC each year in the U.S [5]. The upper bound of 4,500 patients was chosen due to increasing rates of lung cancer diagnosis and the preference of overestimation for BIAs [22]. To account for previous lines of treatment and the stage of cancer, clinical trial data of repotrectinib was used to estimate the percentage of ROS1+ NSCLC patients who were TKI-naive and had metastases [1]. The model incorporates a cohort of 719 patients [see Figure 2], all assumed to start treatment at the beginning of the year. Treatment duration and response rate was not factored in due to the one-year time horizon.

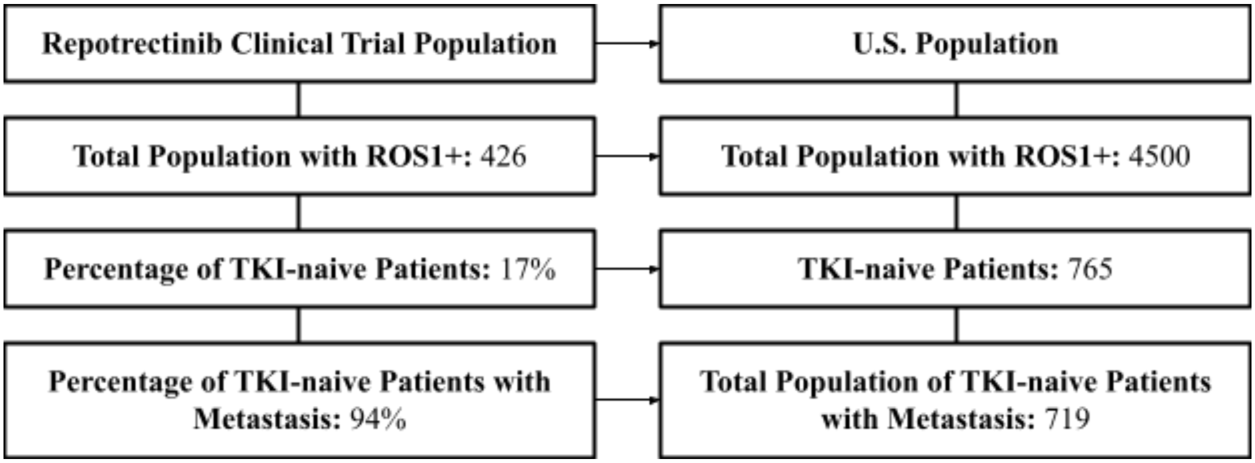

Repotrectinib Clinical Trial Population
U.S. Population
Total Population with ROS1+: 426
Total Population with ROS1+: 4500
Percentage of TKI-naive Patients: 17%
TKI-naive Patients: 765
Percentage of TKI-naive Patients with Metastasis: 94%
Total Population of TKI-naive Patients with Metastasis: 719

Fig. 2: Model population size

Treatment Distribution

To determine the market shares of the entrectinib and repotrectinib scenario, the first quarter financial reports for 2024 were used from *Roche* and *Bristol Myers Squibb*. It was assumed that the ratio of the market share for the quarter of the year applies to the rest of the year. Entrectinib garnered a product sale of around $13,500,000 and repotrectinib garnered a product sale of around $6,000,000. After dividing these sales by their drug acquisitions cost per patient, it was estimated that *Roche* sold 66 entrectinib courses and *Bristol Myers Squibb* sold 17 repotrectinib courses. This resulted in a 79.5% and 20.5% market share respectively [6, 7].

Table 1: Treatment distribution

Market scenario and treatment option	Treatment distribution (%)
Scenario 1	
Entrectinib	100.0
Scenario 2	
Entrectinib	79.5
Repotrectinib	20.5

Treatment Costs

The wholesale acquisition cost per patient per month accounted for the drug acquisition costs. Monitoring requirements for repotrectinib included doctor visits, liver function tests, CPK level tests, uric acid level tests, and imaging. Monitoring requirements for entrectinib included physician visits, liver function tests, uric acid level tests, and imaging. These were based on the FDA-approved prescribing information [3, 4]. Only adverse events that were grade 3 or higher with an incidence rate of at least 4% in clinical trials were considered [14]. It was assumed that patients who experienced adverse events did not experience recurring events due to the lack of data. Increased creatine kinase levels and weight gain usually led to treatment dose reduction or discontinuation and did not have additional costs [8, 9]. Costs were based on *Roche* and *Bristol Myers Squibb* listings, the Centers for Medicare & Medicaid Services (CMS) price lookup tools, and published literature [8, 9, 10, 11, 12, 15]. Costs were adjusted for inflation accordingly [13].

Table 2: Treatment Costs

	Annual repotrectinib treatment costs per patient	
Parameter	Cost ($)	Frequency
Drug acquisition ($/month)	29,000.00	12 months
Diagnosis	10,134.58	Once
Monitoring		
Physician visit	143.39	Every 2 weeks in the first month, then monthly
Liver function test	14.57	Every 2 weeks in the first month, then monthly
CPK levels test	7.62	Every 2 weeks in the first month
Uric acid levels test	8.44	Every 2 weeks in the first month, then monthly
CT scan	288.00	Every 8 weeks
MRI scan	471.00	Every 8 weeks
Adverse event management		
Anemia	2,058.05	One event
Increased blood creatine kinase level	0.00	NA
	Annual entrectinib treatment costs	
Parameter	Cost ($)	Frequency
Drug acquisition ($/month)	17,050	12 months
Diagnosis	10,134.58	Once
Monitoring		
Physician visit	143.39	Every 2 weeks in the first month, then monthly
Liver function test	14.57	Every 2 weeks in the first month, then monthly
Uric acid levels test	8.44	Every 2 weeks in the first month, then monthly
CT scan	288.00	Every 8 weeks

MRI scan	471.00	Every 8 weeks
Adverse event management		
Weight loss	0.00	NA
Neutropenia	1297.79	One event

Analysis

A one-way sensitivity analysis was performed by varying the annual cost per patient and market share variables. The range of costs was determined using a 20% margin from scholarly literature, including previous cost-effectiveness and budget impact analysis articles [14, 15]. Three scenarios were tested for repotrectinib with 70%, 50%, and 6.2% market shares. The 6.2% market share was based on the projected adoption rates of the TKI ceritinib for first-line treatment of ALK+ NSCLC [16]. The one-way sensitivity analysis accounts for uncertainty in the model and identifies the most influential cost drivers. This was followed with a scenario analysis to calculate the most and least expensive scenarios. The least costly scenario occurs when a biopsy is excluded from diagnosis, minimum possible costs are considered, and weight loss management interventions are excluded. The most costly scenario occurs when the maximum costs are considered and weight gain interventions with GLP-analogs are included. It was assumed that the patient used 2.4 mg of GLP-analogs once per week for 11 months [17]. The cost of GLP-analogs was $68,351.19 [18].

Table 3: Ranges of treatment costs

	Cost ranges of repotrectinib treatment line ($)	
Parameter	Least (% of changes from base case)	Most (% of changes from base case)
Drug acquisition ($/month)	23,200.00 (-20%)	34,800.00 (+20%)
Diagnosis	8,107.66 (-20%)	12,161.50 (+20%)
Monitoring		
Physician visit	114.71 (-20%)	172.07 (+20%)
Liver function test	11.66 (-20%)	17.48 (+20%)
CPK levels test	6.10 (-20%)	9.14 (+20%)
Uric acid levels test	6.75 (-20%)	10.13 (+20%)
CT scan	230.40 (-20%)	345.60 (+20%)
MRI scan	376.80 (-20%)	565.20 (+20%)
Adverse event management		

Anemia	1,646.44 (-20%)	2,469.66 (+20%)
Increased blood creatine kinase level	0.00 (NA)	0.00 (NA)

	Cost ranges of entrectinib treatment line ($)	
Parameter	Least (% of changes from base case)	Most (% of changes from base case)
Drug acquisition ($/month)	13,640.00 (-20%)	20,460.00 (+20%)
Diagnosis	326.79 (NA)	11,171.26 (NA)
Monitoring		
Physician visit	114.71 (-20%)	172.07 (+20%)
Liver function test	11.66 (-20%)	17.48 (+20%)
Uric acid levels test	6.75 (-20%)	10.13 (+20%)
CT scan	230.40 (-20%)	345.60 (+20%)
MRI scan	376.80 (-20%)	565.20 (+20%)
Adverse event management		
Weight gain	0.00 (NA)	68,351.19 (NA)
Neutropenia	1,038.23 (-20%)	1,557.35 (+20%)

RESULTS

Base-Case Analysis

In the first scenario with the entrectinib-only formulary, it was assumed that all 719 patients underwent the entrectinib regimen from the beginning to the end of the year. The prices per patient per year (PPPY) for entrectinib acquisition, diagnosis, monitoring, and adverse event management were $204,600.00, $10,134.58, $7,476.20, and $1297.79 respectively. It was estimated that 29 patients experienced grade 3+ neutropenia based on the 4% incidence rate in clinical trials [20]. In the second scenario with the entrectinib and repotrectinib formulary, 572 patients underwent the entrectinib regimen and 147 patients underwent the repotrectinib regimen. The price PPPYs for repotrectinib acquisition, diagnosis, monitoring, and adverse event management were $348,000.00, $10,134.58, $7,491.44, and $2058.05 respectively. It was estimated that 6 patients experienced grade 3+ anemia and 23 patients experienced neutropenia in the second scenario [1, 20]. After multiplying the base-case costs and the frequencies, the annual costs for the first and second scenarios were $159,807,186.73 and $180,893,788.50 respectively. Thus, the budget impact of adopting repotrectinib into the market is an increase of $21,086,601.77.

Table 4 Base-case cost of each scenario

	Scenario 1 ($/year)	Scenario 2 ($/year)	
Parameter	Entrectinib-only	Entrectinib	Repotrectinib
Drug acquisition	147,107,400.00	117,031,200.00	51,156,000.00
Diagnosis	7,286,763.02	5,796,979.76	1,489,783.26
Monitoring	5,375,387.80	4,276,386.40	1,101,241.68
Adverse event management	37,635.91	29,849.17	12,348.30
Total	159,807,186.73	127,134,415.30	53,759,373.24
		180,893,788.50	

Sensitivity Analysis

Through a one-way sensitivity analysis, the key parameters influencing budget impact included drug acquisition costs and market shares. Increasing drug acquisition costs by 20% led to a change in budget impact of $25,302,561.77 per year. Increasing the repotrectinib market share led to a significant increase in budget impact up to $72,153,070.92 per year. In the most expensive scenario, the annual cost of the entrectinib-only formulary was $197,247,402.48, and $221,378,659.30 for the repotrectinib-included scenario. The budget impact for this scenario is $24,131,256.82. In the least expensive scenario, the annual cost of the entrectinib-only formulary was $127,845,749.38, and $144,715,030.90 for the repotrectinib-included scenario. The budget impact for this scenario is $16,869,281.52.

Table 5 Sensitivity analysis

	Scenario 1	Scenario 2	Budget Impact
Parameter	Entrectinib-only	Repotrectinib and entrectinib	Difference
Drug acquisition costs (total cost/year)			
+20%	189,228,666.73	214,531,228.50	25,302,561.77
-20%	130,385,706.73	147,256,348.50	16,870,641.77
Monitoring costs (total cost/year)			

+20%	160,882,264.29	181,969,314.10	21,087,049.81
-20%	158,732,109.17	179,818262.90	21,086,153.73
Adverse events costs (total cost/year)			
+20%	159,814,713.91	180,902,228.00	21,087,514.09
-20%	159,799,659.55	180,885,349.00	21,085,689.45
Weight gain meds	165,206,930.74	185,199,913.47	25,392,726.74
Market shares' impact on costs (total cost/year)[a]			
70% R, 30% E	159,807,186.73	231,960,257.65	72,153,070.92
50% R, 30% E	159,807,186.73	211,274,311.36	51,567,124.63
6.2% R 93.8% E	159,807,186.73	166,262,393.05	6,455,206.32

[a] X% R, Y% E refers to a X% market share for repotrectinib and a Y% market share for entrectinib.

DISCUSSION

The introduction of repotrectinib is projected to result in a significant increase in the budget for U.S. payers of approximately $21,086,601.77. This substantial budget impact calls for careful consideration of its financial implications. While repotrectinib has certain advantages over early-generation ROS1+ TKIs, the potential for higher costs raises questions about the sustainability and affordability of repotrectinib. This can lead to 1.) higher premiums for beneficiaries or increased taxes to cover public health expenditures, and 2.) reconsiderations and negotiations on drug pricing and reimbursement policies for accessibility.

Repotrectinib may experience quicker market adoption rates. This potential for rapid uptake can lead to improved health outcomes for patients but can strain healthcare budgets and increase out-of-pocket costs for patients and insurers. However, *Bristol Myers Squibb* expects the market potential for repotrectinib to be $1.14 billion, overall increasing the market size for ROS1+ NSCLC drug market [19]. This can lead to reinvestment into more clinical trial research.

One of the current limitations is the lack of clinical trial data. A time horizon greater than one year could not be achieved due to the lack of information on second and third-line treatments for repotrectinib as the current TRIDENT-1 trial is still ongoing [1]. Perhaps the costs of the entrectinib treatment line may exceed that of the repotrectinib treatment line over a three-year time horizon due to the inclusion of second-line pembrolizumab, pemetrexed, and carboplatin, third-line docetaxel, and palliation. This would imply a budget savings for the adoption of repotrectinib. Overall, the results of this research are subject to a degree of uncertainty based on the accuracy of the underlying data that the assumptions were based on. While there is an uncertainty of the extent of budget impact, repotrectinib will certainly increase the budget impact for U.S. payers. They cannot be directly applied to other healthcare systems with different economic and clinical contexts. Future research should focus on the cost-effectiveness analysis of repotrectinib. This study would consider not only the financial costs but also the clinical benefits and quality-of-life improvements for patients. It can better inform U.S. payers' decision-making in allocating resources.

CONCLUSION

The budget impact analysis suggests a substantial budget increase due to the introduction of repotrectinib into the ROS1+ treatment market over the first year. U.S. payers should consider both the expenses and clinical benefits of repotrectinib.

Works Cited

[1] Drilon, Alexander. et al. “Repotrectinib in ROS1 Fusion-Positive Non-Small-Cell Lung Cancer.” *The New England Journal of Medicine*, vol. 390, no. 2, 2024, pp. 118-131, DOI: 10.1056/NEJMoa2302299.

[2] Center for Drug Evaluation and Research. “FDA approves repotrectinib for ROS1-positive non-small cell lung cancer.” *FDA*, 2023, www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-repotrectinib-ros1-positive-non-small-cell-lung-cancer.

[3] —. “Augtyro - accessdata.fda.gov.” *FDA*, 2023, https://www.accessdata.fda.gov/drugsatfda _docs /label/2023/218213s000lbl.pdf.

[4] —. “Rozlytrek - accessdata.fda.gov.” *FDA*, 2019, https://www.accessdata.fda.gov/drugsatfda_docs/label/2019/212725s000lbl.pdf.

[5] Lin, Jessica, and Alice T. Shaw. “Recent Advances in Targeting ROS1 Lung Cancer.” *Journal of Thoracic Oncology*, vol. 12, no. 11, 2017, pp. 1611-1625, DOI: 10.1016/j.jtho.2017.08.002.

[6] —. “Investor Update.” *Roche*, 2024, assets.roche.com/f/176343/x/0b2f46ea33/240424 _ir_q1.pdf.

[7] —. “Bristol Myers Squibb Reports First Quarter Financial Results for 2024.” *Bristol Myers Squibb*, 2024, news.bms.com/news/details/2024/Bristol-Myers-Squibb-Reports -First-Quarter-Financial-Results-for-2024/default.aspx.

[8] —. “Augtyro (repotrectinib) 40 mg capsules.” *AUGTYRO*, 2024, www.augtyrohcp.com/.

[9] —. “Rozlytrek (entrectinib) 100 mg | 200 mg capsules.” *ROZLYTREK*, 2024, www.rozlytrek-hcp.com/.

[10] —. “Medical Oncology Office Visits Costs.” *CMS*, 2023, data.cms.gov/provider-data/dataset/5d65-6dcf#data-table

[11] —. “Physician Fee Schedule.” *CMS*, 2023, www.cms.gov/medicare/physician-fee -schedule/search.

[12] —. “Magnetic resonance (eg, proton) imaging, brain (including brain stem); without contrast material, followed by contrast material(s) and further sequences.” *Medicare*, 2024, www.medicare.gov/procedure-price-lookup/cost/70553/.

[13] —. “CPI Inflation Calculator.” *BLS*, 2024, data.bls.gov/cgi-bin/cpicalc.pl.

[14] Huo, Gengwei, et al. “Entrectinib as first-line vs. second-line therapy in ROS1 fusion-positive non-small cell lung cancer: a cost-effectiveness analysis.” *Translational Lung Cancer Research*, vol. 13, no. 4, 2024, pp. 839-848, DOI: 10.21037/tlcr-24-8.

[15] Insinga, Ralph, et al. “Cost-effectiveness of pembrolizumab in combination with chemotherapy in the 1st line treatment of non-squamous NSCLC in the US.” *Journal of Medical Economics*, vol. 21, no. 12, 2018, pp. 1191-1205, DOI: 10.1080/13696998.2018.1521416.

[16] Mutebi, A, et al. “Budget Impact Analysis of First-Line Ceritinib in The Treatment of Alk+ Metastatic Non-Small Cell Lung Cancer (NSCLC) in The United States.” *Value in*

Health, vol. 20, no. 9, 2017, p. 423, DOI: 10.1016/j.jval.2017.08.147.

[17] Liu, D., et al. “Characterization of On-Target Adverse Events Caused by TRK Inhibitor Therapy.” *Ann Oncology*, vol. 31, no. 9, 2020, pp. 1207-1215, DOI: 10.1016/j.annonc.2020.05.006.

[18] —. “Drug Price Information.” *Drugs.com*, 2024, www.drugs.com/price-guide/.

[19] Armstrong, Annalee. “UPDATE: Bristol Myers strikes up Roche rivalry with $4.1B Turning Point buy.” *Fierce Biotech*, 2022, www.fiercebiotech.com/biotech/bms-strikes-roche-rivalry-41b-acquisition-turning-point.

[20] Drilon, Alexander, et al. “Long-Term Efficacy and Safety of Entrectinib in ROS1 Fusion-Positive NSCLC.” *JTO Clinical and Research Reports*, vol. 3, no. 6, 2022, p. 100332, DOI: 10.1016/j.jtocrr.2022.100332.

[21] Shaw, Alice, et al. “Crizotinib in ROS1-Rearranged Non-Small-Cell Lung Cancer.” *The New England Journal of Medicine*, vol. 371, no. 2, 2014, pp. 1963-1971, DOI: 10.1056/NEJMoa1406766.

[22] —. “Session 6: Budget Impact Analysis.” *Network of Alberta*, 2017, www.youtube.com/watch?v=SMJmkUITSWw.

Universal Basic Income: An exploration of public perceptions of UBI in India and its feasibility By Aryavardhan Agarwal

Abstract

This empirical study explores the public perceptions of Universal Basic Income (UBI) in India. India is a country with vast socio-economic diversity and continues to tackle challenges in the form of red-tapism and corruption while providing for welfare schemes. After surveying over 260 citizens across demographic backgrounds in India, this paper explores the feasibility and public perception towards UBI implementation. The findings reveal a diverse landscape of opinions influenced by variables such as economic status, education level and employment status. Overall, 73.46% of respondents advocate for a UBI policy for its potential to reduce poverty and provide financial security. Contrarily, 26.54% stand against the UBI scheme and have expressed concern in light of its opportunity cost with respect to other welfare schemes and its feasibility. Demographic variations are noteworthy as there is higher support for UBI in low-earners and scepticism among the affluent. The study highlights the importance of addressing economic disparities and contextual differences in UBI policy formulation. This paper contributes to the global discourse on UBI and underscores the need for targeted research to address the practical challenges and public concerns associated with its implementation by providing comprehensive insights into the Indian socio-economic context.

Introduction

Although Universal Basic Income (UBI) has gained quick momentum as a compelling public policy proposal in recent years, the rudimentary idea of such a scheme dates back to the 18th Century. The idea of redistributing wealth from the rich to the poor was first proposed by Thomas Paine, an American revolutionary, and refined by English writer Thomas Spencer in 1797. Nonetheless, its experimentation and implementation have accelerated in the past few decades.

A UBI Scheme is a periodic and unconditional basic income given to every citizen by their respective governments. It has caused fierce debates among researchers, economists, policymakers, politicians and citizens alike due to its contentious nature and the uncertainty of its consequences. Proponents strongly believe that UBI can reduce poverty and help attain social security considering the Artificial Intelligence (AI) threat. On the other hand, critics raise concerns about its fiscal feasibility and potential to impact labour supply adversely and disincentivise work. Multiple pilot projects and experiments have been conducted to capture the impact of such a scheme.

The contention around UBI is particularly relevant in India's case. India is one of the world's largest and fastest-growing economies and is home to high levels of income inequality and diverse socio-economic conditions. Understanding public perception of UBI in India is imperative for multiple reasons. Primarily, as a democratic nation, India must consider the perspective of its citizens when formulating policies. It is of utmost importance to understand the public perception to ensure its conditions are welcomed and abided by. Secondly, understanding how one would benefit from UBI and whether the Indian population needs it is essential for assessing its potential impact. Thirdly, weighing the opportunity cost with respect to the existing welfare schemes and their impact on the public is necessary to judge whether India needs a new social support scheme such as UBI.

The aim of this comprehensive empirical study is to evaluate these perceptions. The study involves a meticulous literature review, a detailed survey, a thorough discussion and noteworthy conclusions. The study explores the feasibility and viability of UBI in India while considering its effects on the labour market and overall social welfare. The survey examines attitudes towards UBI based on specific scenarios and sheds light on its perceived advantages and disadvantages. This research also seeks to assess the potential impact on the labour market after assuming a hypothetical introduction of UBI in India by collecting empirical data.

Literature Review

Historical evolution of the idea of UBI

There are various definitions of UBI but this study employs Van Parijs' (2004) approach, which defines UBI as "a Basic Income (or demogrant) is an income paid by a political community to all its members on an individual basis, without means test or work requirement." (p.7). This implies that UBI is a fiscal disbursement orchestrated by the government of a country periodically and spans across the entirety of the country. It comes with no strings attached, is allotted to citizens individually (not as a family), is enough to cover basic living expenses and does not depend on demographics: race, gender, income bracket, etc. UBI has sparked renewed scholarly as well as political interest following the aftermath of the COVID-19 pandemic, which exacerbated income inequality levels and unemployment.

The study of Teather-Posadas (2017) has highlighted perspectives on the UBI scheme by eminent philosophers and economists. It concluded that UBI, though seen by some as a contemporary and new idea, has deep historical roots. UBI has been advocated for by figures like Thomas Paine and Charles Fourier and has been in contention for a long time. Paine proposed restitution to compensate for the inequities of civilisation and favoured a national fund to provide financial support to individuals. He saw this as a right and not mere charity. Fourier echoed this sentiment, proposing planned communities or phalanxes to provide for a basic standard of living for all. In favour of the basic income, a renowned British philosopher Bertrand Russell stated, “a certain small income, sufficient for necessities, should be secured for all, whether they work or not. A larger income … should be given to those who are willing to engage in some work which the community recognizes as useful.” (Russell, cited in Teather-Posadas, 2017, p.4). Later, John Kenneth Galbraith and Milton Friedman introduced the concept of a negative income tax as a way to ensure a minimum income. The idea saw a large-scale implementation in the United States of America in the 1960s for experimental reasons but was met with mixed results.

The idea of a UBI scheme, which has always been in contention, was met by several experimental implementations in the 21st Century. The Republic of Finland conducted a two-year basic income experiment in 2017. The Ministry of Social Affairs and Health and Kela in a press release state,

> The basic income recipients were more satisfied with their lives and experienced less mental strain than the control group. They also had a more positive perception of their economic welfare. The interpretation of the employment effects of the experiment is complicated by the introduction of the activation model in 2018. (Ministry of Social Affairs and Health, 2020).

By providing a guaranteed, unconditional basic income, recipients experienced improvements in health, happiness, cognitive abilities, and financial security. This overall upliftment in their standards of living led them to pursue more opportunities, creating a positive feedback loop. The mighty contention on potential risks of labour shortage in the labour market was met with an optimistic conclusion as the provision of an income had an almost negligible effect on the employment rate and, instead, resulted in a slight increase.

The threat of high inflation rates has been another key concern in the implementation of a UBI scheme. Nonetheless, an ongoing pilot experiment which is being conducted in Kenya by the non-profit organisation GiveDirectly undermines the threat. After the analysis of the effects of the UBI income beneficiaries, Aizenman (2023) suggests that the influx of money into that particular African economy did not increase the inflation rate. The study found a plausible reason, as "people did buy more things, this extra spending was distributed over a wide range of products, depending on the relative wealth of the person getting the aid." The ongoing program aims to analyse economic and social well-being along with macroeconomic implications of a UBI income after the completion of their 12-year-long research.

UBI policies in the Indian context

India, a country that suffers from high-income inequality, is a unique backdrop for discussing the implementation of a UBI scheme. Implementing a Universal Basic Income (UBI) policy in India can address income inequality, provide a safety net for vulnerable populations, and counteract systemic corruption. A UBI scheme can enhance social welfare and foster economic autonomy thereby offering a potential solution to India's pressing economic disparities by distributing resources equitably. However, one cannot neglect the fiscal implications and large-scale implementation challenges.

Income inequality is a pressing issue in India, with disparities between the affluent and marginalised populations. A study conducted by Oxfam International (Himanshu, 2023) recognises that the top 10% of the Indian Population (in terms of income) constitutes 77% of the national wealth. According to the World Bank, approximately 13% of India's population lives below the international poverty line (World Bank Group, 2024). Moreover, the informal sector, which employs about 70% of the workforce, often fails to provide financial security or social benefits to its employees leaving them in a vulnerable position (Bonnet et al., 2019). UBI has the potential to reduce income inequality by distributing resources more equitably, empowering the economically disadvantaged and fostering a more balanced distribution of wealth. UBI can enhance social welfare by providing a safety net for vulnerable populations. It ensures that individuals, particularly those in precarious economic situations, have a reliable source of income. This can improve access to education, healthcare, and other essential services, ultimately fostering social development. According to the ADP Research Institute's People at Work 2023: A Global Workforce View report, 37% of employees in India do not have job security or feel less secure in their current jobs (Richardson et al., 2023).

Bardhan (2017) argues that UBI can potentially boost the autonomy of Indian women as three-quarters of Indian women are not employed. Bardhan further explored the need for UBI in India as a prospect of being an "escape ladder for people in stigmatised occupations in Indian society (manual scavenging, animal skinning, prostitution, etc.)" (Bardhan, 2017, p.142). Ghose's (2017) research explores how corruption and malpractices in the implementation of welfare schemes have not efficiently benefited the poor. Dreze and Khera (2017) highlight the inefficiencies of the Public Distribution System (PDS) and National Rural Employment Guarantee Act (NREGA) government schemes in India and exemplify the red-tapism and corruption entitled to the schemes. Ghose exacerbates the growing corruption in the country due to which the poor are often excluded and non-poor are included while receiving benefits from the welfare schemes.

Exploring perceptions of UBI policies

Although implementing a potential UBI Scheme looks beneficial, the idea might still rest in a utopian scenario as it faces considerable opposition. Fiscal bureaucrats, economists and government officials view it as impractical (Bardhan, 2017; Ghose, 2017) as it requires a considerable amount of mobilisation in the budget. If proposed as a basic right to every citizen (rich and poor) social activists might think of this as another way to circulate money amongst the rich and as an ingenious ploy to undermine the current welfare schemes. Economists also believe that a large-scale infusion of cash through UBI could lead to increased consumer spending, potentially triggering inflationary pressures. This inflation may erode the purchasing power of the income provided through UBI, negating its intended positive impact on the standard of living for recipients. The Government of India (2023) in a Budget report stated that there are 185 major welfare schemes in India and according to leading economists, heavy mobilisation of the budget will lead to the cutting down of most of these schemes to make space for a potential UBI policy. Rengasamy, Kumar (2012) stated how "according to the national bulletin of MGNREGA, the scheme has so far provided 3.34 crore households with employment throughout the country." The exclusion of such schemes due to UBI is likely to attract considerable opposition from social activists and economists.

Hamilton et al. (2023) found through a survey that only 5.1% of respondents in the United States would reduce their working hours if a UBI scheme were introduced, dismissing the idea that UBI would broadly discourage work. However, the study highlighted a notable difference between the effects of monthly versus lump-sum UBI payments on labor supply. The research suggested that monthly payments might encourage some individuals to reduce their working hours, as part of their income would be consistently supplemented. Conversely, respondents who preferred lump-sum payments were more likely to use the money to pay off debts.

On the other hand, according to a poll conducted by IPSOS, 63% of Americans agreed to the fact that “[b]asic income will discourage people from being in or seeking paid employment” (Colledge and Martyn, 2017, p.6). Similarly, Richards and Steiger (2021) in their study presented a UBI scenario in front of their respondents and asked whether the number of people not working would increase to which 47.8% of the respondents agreed that it would increase.

Therefore, there is not much clarity on whether a large-scale implementation of UBI in a country like India with the highest population in the world, would reduce, or, perhaps, increase the labour supply. The relationship between unemployment and UBI implies several complexities. UBI does not only affect labour supply but also can play a crucial role in impacting labour demand. The implementation of UBI affects:

> [...] labour demand, entrepreneurship, productivity, aggregate demand, innovation and the development of technology, the 4th Industrial Revolution, social participation, power-sharing, the capacity to negotiate, security, instability resulting from poverty and excessive inequality, personal autonomy and satisfaction, health and education. (Paz-Banez et al., 2020, p.2)

Furman and Seamans (2019) perceived that a UBI scheme can cushion mass joblessness, helping to ensure a basic income floor for people. The threat of AI with regard to unemployment and joblessness could be tackled by implementing a UBI scheme that will act as a safety net. However, a drawback of basic income is that unconditional cash transfers would induce idleness in society (Rycroft, 2017). Those who would typically seek out employment opportunities to secure a steady income might, instead, lean on UBI as a viable alternative, thus forgoing their engagement in the labour force.

The common perception of a UBI scheme causing inflation has also been challenged. We can consider evidence from Alaska's inflation data. Alaska has administered a form of universal basic income through the Alaska Permanent Fund Dividend since 1982, which grants each resident an annual dividend ranging from $500 to $3,000, financed by state oil revenues. Following the introduction of this program, Alaska has experienced lower inflation rates compared to the rest of the United States, a departure from its previous parity in inflation rates (Scott Santens, 2014).

More than 75 years of India’s independence and yet one in five people in India are suffering from poverty (World Bank Report, 2016). The existing welfare schemes in India have been obscured in red-tapism, exclusion and inclusion errors, corruption, exorbitant administrative costs and high leakages. With automation coming in and threatening more than

69% of the jobs in India, according to the World Bank Report, 2016, a solution to stop the rising poverty and income inequality must be found. The economic survey of 2016-2017 by the Indian Government has upheld the stance that UBI is a potential alternative to welfare schemes seeing the corruption, red-tapism and high leakages in the current scenario. In today's age of rising income inequality and human skills becoming obsolete, UBI will act as a safety net to secure one's future and protect against unemployment, income and other adversaries.

Assessing the future and the possibility of 69% of jobs in India being threatened, UBI is a strong contender to replace the current schemes. In 2011 when SEWA funded by UNICEF ran a pilot project in 8 villages of Madhya Pradesh, India, seeking to uncover the perceptions and possibilities of implementing a UBI scheme, they made several notable observations. A majority of the villagers did not prefer subsidies (rice, wheat, sugar) after having experienced the basic income scheme. Hence, it was concluded that cash-in-hand and the freedom to spend the additional money was valued more than subsidies and welfare schemes. The standards of living increased, better infrastructure was built with roofs and toilets coming into houses and the recipients started making efforts to maintain their hygiene. Professor Guy Standing, a founding member of the Basic Income Earth Network, said, "the grants led to more labour and work, with a shift from casual wage labour to more own-account (self-employed) farming and business activity" (Kundu, 2016). There was no income effect observed in the labour market and a positive result was noticed. Therefore, according to the survey where respondents are welcoming UBI, past pilot studies and assessing UBI as a safety net for the future, the implementation of such a scheme is necessary.

Methodology

This study employs a quantitative empirical survey to investigate perceptions and potential behavioural responses to particular Universal Basic Income (UBI) scenarios in India. This study aims to explore public perceptions about UBI specifically focusing on its feasibility, desirability, and potential impact on various socio-economic groups in India.

The primary data collection method was an online survey. This approach enabled the efficient gathering of data from a diverse group of respondents across different geographic locations. The survey consisted of 19 questions and collected data from respondents about their demographic status and perceptions of UBI. A stratified random sampling technique was employed to ensure a balanced representation across relevant subgroups within the target population. A total of 260 respondents participated in the survey: approximately 120 from an educational institution in Dehradun, India (comprising approximately 40 teaching staff and 80 non-teaching staff), and 140 from various tier 3 cities in India, which were selected randomly. This approach enhances the study's ability to draw meaningful conclusions from diverse perspectives within the Indian population.

The questionnaire consisted of two sections. The first section included six questions that aimed to collect demographic information such as age, gender, income level, employment status, educational background, and marital status. This helped explore potential differences between the perceptions of different participant groups. The latter part of the questionnaire was inspired by Hamilton et al. (2021), with multiple-choice questions collecting information about the participant's perceptions of UBI, which were adapted to fit the Indian context. Respondents were asked about their awareness of the UBI scheme, their views on the feasibility of such a scheme, and their perceptions and preferences of UBI implementation. The questionnaire also asked the respondents about their potential usage of the basic income and how they thought others would spend the basic income provided. The full list of questions is provided in the appendix.

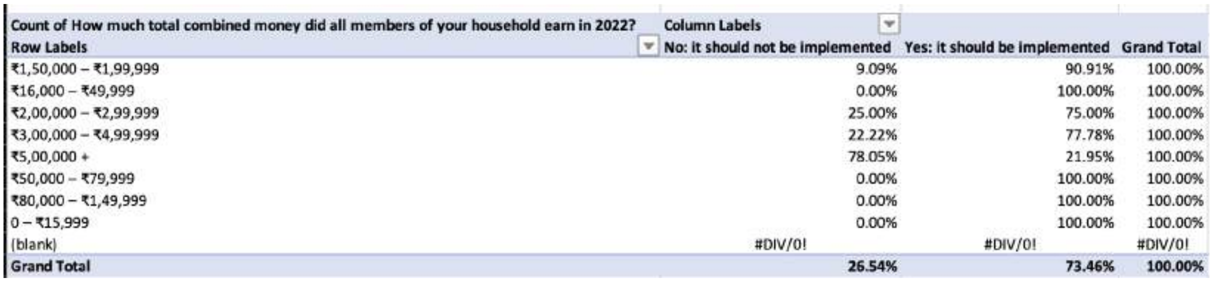
The study took place over a period of four weeks in November and December 2023. A Google Form was made to be filled by the respondents which had all the necessary questions. This approach was selected due to the user-friendliness, security, and wide device compatibility of this platform. The survey took the format of a researcher-administered survey and each participant took approximately 8-12 minutes to complete the survey. A major challenge encountered during data collection was ensuring that participants fully understood the concept of UBI due to the lack of awareness of such a scheme in remote areas and among low-income groups. This limitation was addressed by providing a brief explanation of the scheme and its components at the beginning of the survey.

The results of the survey were exported from Google Forms and analysed using Microsoft Excel. The data analysis involved creating pivot tables and calculating frequencies to identify patterns and trends in the responses. This analysis aimed to reveal insights into public perceptions of UBI which included variations across different demographic groups. By employing these methods, the study sought to provide a comprehensive understanding of the potential acceptance and impact of UBI in India by evaluating the public perception. Participation was completely voluntary and the survey obtained informed consent from the participants. Due to the design of the survey, data collection was anonymous as no identifiable or personal data about individuals was collected.

Results and Discussion

This survey was completed by 261 respondents, of which, 52.9% were male, 45.6% were female and 1.5% were Non-binary/third gender. 22.6% of the respondents were in the 40-49 age group, 21.8% in the 21-29 age group, 21.1% in the 30-39 age group, 13.4% in the 50-59 age group, 11.5% in the 18-20 age group and 9.6% of the respondents in the 60+ age group. The majority of the respondents were employed (91.1) and the rest 8.9% were unemployed. A detailed break-up of the demographics has been provided in Appendix 1.

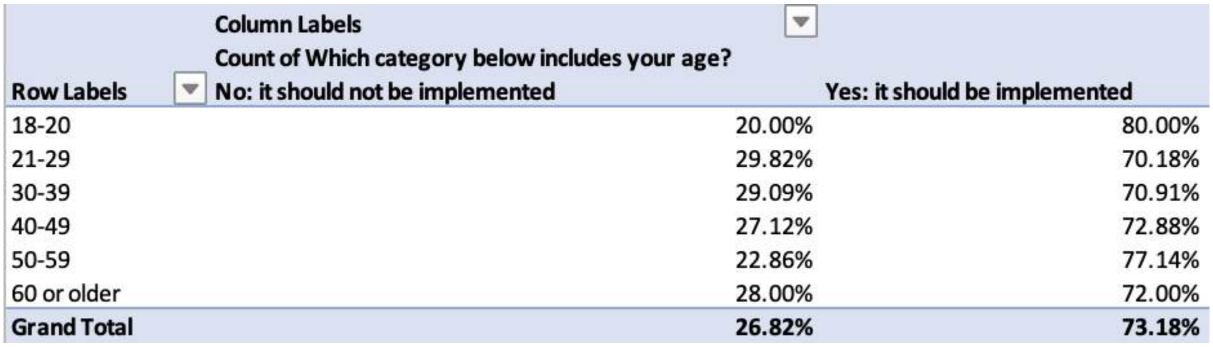
The vast majority of respondents in the ₹1,50,000 - ₹1,99,999 (90.91%) income group and all the respondents in the ₹80,000 - ₹1,49,000 income group believe that a UBI policy should be implemented in India. On the other hand, a majority of the respondents in the ₹5,00,000+ income group (78.05%) believe it would not be appropriate to implement a UBI policy in India. This is shown in Table 1. This shows that the low-earners will welcome the implementation of a UBI scheme.



Count of How much total combined money did all members of your household earn in 2022?	Column Labels		
Row Labels	No: it should not be implemented	Yes: it should be implemented	Grand Total
₹1,50,000 – ₹1,99,999	9.09%	90.91%	100.00%
₹16,000 – ₹49,999	0.00%	100.00%	100.00%
₹2,00,000 – ₹2,99,999	25.00%	75.00%	100.00%
₹3,00,000 – ₹4,99,999	22.22%	77.78%	100.00%
₹5,00,000 +	78.05%	21.95%	100.00%
₹50,000 – ₹79,999	0.00%	100.00%	100.00%
₹80,000 – ₹1,49,999	0.00%	100.00%	100.00%
0 – ₹15,999	0.00%	100.00%	100.00%
(blank)	#DIV/0!	#DIV/0!	#DIV/0!
Grand Total	**26.54%**	**73.46%**	**100.00%**

Table 1: Relationship between income level (Q6) and preference for UBI (Q19)

Table 2 illustrates the relationship between the age of the respondents and their preference for a UBI scheme. In the ages 18-20, we see a higher percentage of respondents advocating in favour of a UBI scheme. Smedley (2017) argues that younger individuals, those with lower incomes, and those in precarious employment situations showed higher levels of support for UBI, perceiving it as a safety net against economic instability. This claim can be supported by the responses in the survey. UBI was particularly more popular in the lower income group and was preferred more by individuals in the 18-20 age group.



	Column Labels Count of Which category below includes your age?	
Row Labels	No: it should not be implemented	Yes: it should be implemented
18-20	20.00%	80.00%
21-29	29.82%	70.18%
30-39	29.09%	70.91%
40-49	27.12%	72.88%
50-59	22.86%	77.14%
60 or older	28.00%	72.00%
Grand Total	**26.82%**	**73.18%**

Table 2: Relationship between Age Group (Q1) and preference for UBI (Q19)

To assess the potential relationship between the level of education and their response to whether they want a UBI scheme to be implemented, we enquired the respondents about their educational background. While 64.63% of people in the ₹5,00,000+ income group had a postgraduate degree, 97.01 % of the respondents in the ₹80,000 - ₹1,49,000 group had a high school degree or lesser. People with a postgraduate degree have a better understanding of the workings of an economy and can better evaluate factors weighed in public policy. and decision-making. This precise population group (postgraduate degree holders with ₹5,00,000+ annual income) of the respondents were against the idea of UBI. A plausible reason for this population group being against the proposal of a UBI scheme in India is the GDP mobilisation caused by the implementation of UBI and the opportunity cost. In this case, the opportunity cost of implementing a UBI scheme would be the removal of multiple subsidies and the potential reduction of the budget allotted to government welfare schemes. For instance, the Pradhan Mantri Ayushman Bharat Yojana (National Health Authority, 2019) provides insurance coverage to more than 50,00,00,000 citizens in India up to ₹5,00,000 per family and this provides more financial security to families than a ₹20,000 UBI scheme.

While evaluating the relationship between the income group of the respondents and their preference between a UBI Scheme or welfare scheme we figured that 81.71% of the respondents in the ₹5,00,000+ income group preferred welfare schemes over UBI whereas 90.91% of the respondents in the ₹1,50,000 – ₹1,99,999 income group would rather opt for the implementation of a UBI scheme. Radhakrishna (2017) with respect to this point of contention reflected upon the availability of schemes such as the ones under the Integrated Child Development Services (ICDS) programme. Supplementary Nutrition Programme (SNP) under (ICDS) provides supplementary food and ration to children up to 5 years of age and provides financial support in their pre-education and healthcare. The implementation of UBI might dismantle such schemes where the impact is not just quantitative but qualitative and intangible. Schemes such as MGNREGA that provide employment and improve the wages of unemployed wage-seekers will also take a backseat when a large-scale investment and the burden of executing a UBI Scheme come onto the shoulders of the government.

The study found that 79.3% of the respondents favoured a UBI scheme for people below or near the poverty line. In addition, 40.24% of the respondents in the ₹5,00,000+ income group found it more viable to have a UBI scheme for people near or below the poverty line and 97.01% of respondents in the ₹80,000 - ₹1,49,000 income group agreed to the same. Although a quasi-UBI scheme restricted to be served to people near the poverty line was preferred by the respondents of the study, it would call for more burden on the government. With the current schemes in place, identifying beneficiaries, leakages and corruption are the dominant pain points for the government. A quasi-UBI scheme will not solve this, rather, it would open more scope for leakages and corruption as it will be prone to a larger share of the budget. In lieu of a quasi-UBI scheme for people near the poverty line, a scheme can be implemented on a particular social criteria (widows, orphans, poor elderly). In such a scheme beneficiaries can be easily identified and it will act as an inception point for a large-scale UBI policy. A large-scale UBI policy where the level of the basic income has to be equivalent to the poverty line will cost India 12.5% of the GDP as predicted by the Economic Survey 2016-2017 (Kumar and Kanojia, 2017). This will account for a huge share of the Union government's budget and revenue expenditure and hence cannot be implemented. A quasi-UBI would be a better and more feasible solution for India and a great starting point to curb income inequality and poverty. Radhakrishna (2017) while commenting on a question about whether UBI should be universal or targeted, concludes that this has not attracted unanimity yet. Joshi (2017) in light of a targeted UBI scheme suggested that if budget and expenditure problems arise with respect to implementation of a UBI scheme then it would be a feasible solution to start with a quasi-UBI scheme targeting only women. This will also reduce the administrative burden of identifying beneficiaries. In contrast to our findings, Bardhan (2017) envisions UBI as a basic right for all whereas the opposition political party of India, the Indian National Congress proposed a UBI-like scheme but for a specific audience (specific income group or social criteria) which seconds our respondents' opinions.

Although this has its setbacks, Radhakrishna (2017) stated that if UBI is targeted to a specific audience then other "problems associated with identifying the beneficiaries are likely to arise" (p. 201). However, the current welfare schemes do require the identification of beneficiaries, for instance, the PM-JAY scheme has rigorous identification criteria through which they recognise beneficiaries and have already identified 30 crore beneficiaries. Hence, some infrastructure has already been developed and well-practised with respect to the identification of beneficiaries, so although there can be problems associated with identifying beneficiaries in a quasi-UBI scheme, India is prepared for it.

The impact of a UBI scheme on employment and labour-force participation has been in the limelight more often than not when considering the implementation of a UBI Scheme.

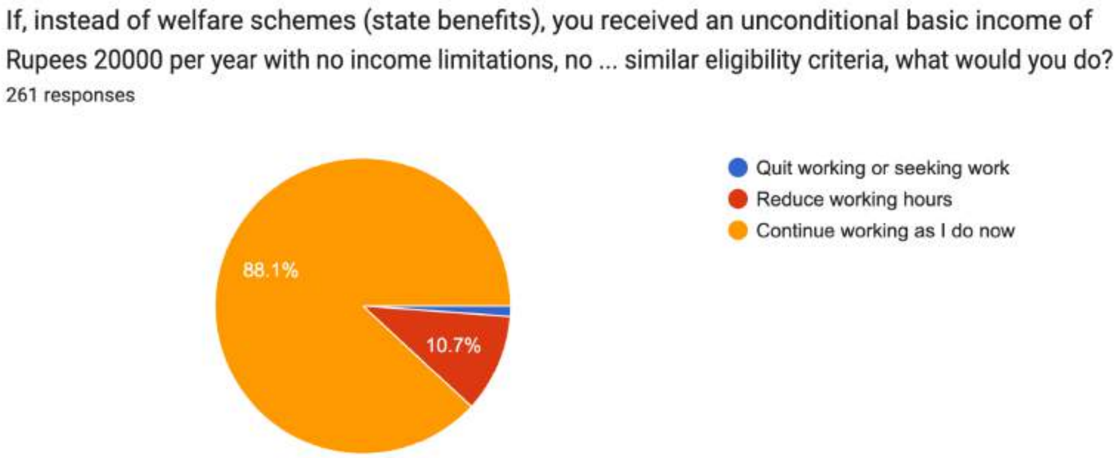
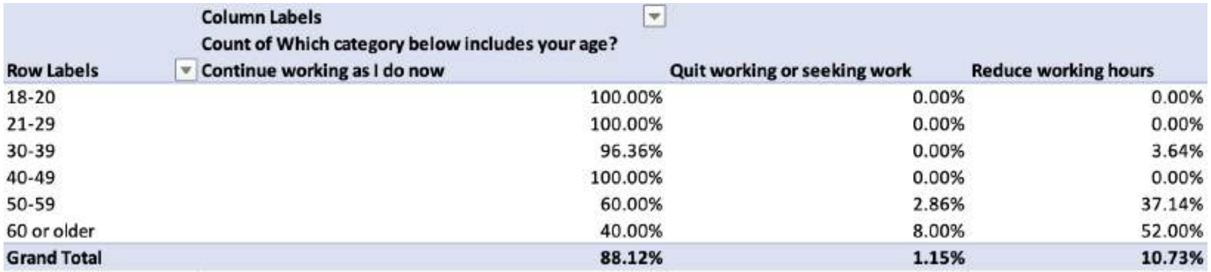


Figure 1: (Q13) Effect on Employment after UBI Implementation

The modest level of proposed income gain through a UBI scheme does not favour quitting jobs or seeking unemployment as a viable option. A much-contested argument claims that female labour force participation will reduce, on asking our respondents results concluded that 89.08% of female respondents favoured working similarly as they do now and only 9.24% would reduce their working hours. We find a similar pattern in the male respondents' response (Q13) to the question where 86.96% of them agreed to the same. Therefore, we can conclude that gender does not have a visible impact on labour participation after the implementation of a UBI scheme. The results second the claims made in Joshi's (2017) study where he claimed that income effect against work is likely to be quantitatively negligible.

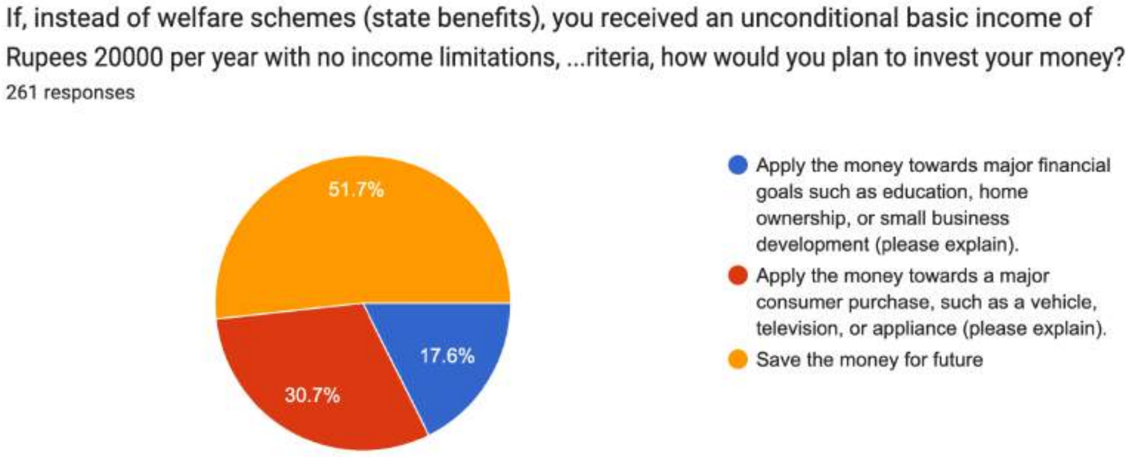
The current research project also explored whether the age of the respondent does vary their response to Q13. Notably, 52% of respondents above the age of 60 and 37.14% of the respondents in the 50-59 year age group would reduce their working hours whereas 0% of respondents in the 40-49 year age group opted to reduce their working hours. This has been shown in Table 3. Hence, we can argue that age does have an important impact on considering reducing their working hours. The plausible reason for the higher age groups opting to reduce their working hours is their deteriorating working capabilities and healthcare.



	Column Labels Count of Which category below includes your age?		
Row Labels	Continue working as I do now	Quit working or seeking work	Reduce working hours
18-20	100.00%	0.00%	0.00%
21-29	100.00%	0.00%	0.00%
30-39	96.36%	0.00%	3.64%
40-49	100.00%	0.00%	0.00%
50-59	60.00%	2.86%	37.14%
60 or older	40.00%	8.00%	52.00%
Grand Total	**88.12%**	**1.15%**	**10.73%**

Table 3: Relationship between Age Group (Q1) and Employment (Q5)

Examining how respondents are likely to spend their UBI-generated income, the survey found that 51.7% of the respondents wanted to save the money for the future. The results are illustrated in Figure 2.



If, instead of welfare schemes (state benefits), you received an unconditional basic income of Rupees 20000 per year with no income limitations, ...riteria, how would you plan to invest your money?

261 responses

Figure 2: (Q14) Probable investment of income gained from a UBI scheme

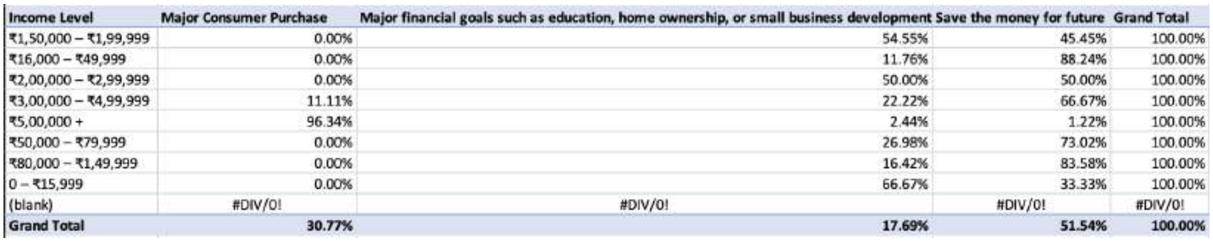
Out of the 51.7% of respondents who wished to save the UBI-added income for their future, 71.52% belonged to the ₹50,000 - ₹1,49,000 income group. 73.02% of the respondents in the ₹50,000 - ₹79,999 income group and 85.07% of the respondents in the ₹80,000 - ₹1,49,000 income wished to save the money for their future. Again we notice a contrasting pattern of responses with respect to the ₹5,00,000+ income group where 79.27% of the respondents wished to apply the money towards a major consumer purchase. Hamilton et al. (2023) did not discover much variation in their findings as to how the respondents of their survey would spend their money. In their research majority of the respondents would pay off debts in case of a one-time UBI payment and alternatively use money for regular day-to-day expenses in case of monthly UBI payments. People in the ₹5,00,000+ income group are more likely to spend their money on major consumer purchases as they already have enough cash flow to save money periodically unlike the lower income groups. Lower-income groups do not have enough cash flow to save enough money and hence majority of them opted to save their money.

Conclusion

This paper aimed to evaluate the public's perceptions regarding the implementation of a Universal Basic Income scheme in an Indian context. The research was based on an empirical survey and data has been assessed based on the evaluation of responses. In conclusion, the survey helps provide insights into the contentious dynamics surrounding the implementation of a Universal Basic Income (UBI) scheme in India. The varied responses from the respondents showcase the perceptions of different demographic groups with diverse income ranges. While a majority of the respondents favour UBI, there has been considerable opposition from a particular demographic group.

The opposition to UBI among high earners, in particular those high-earners with high levels of education, highlights concerns about the potential disruption of current welfare schemes and the opportunity cost associated with implementing UBI. The reluctance to UBI, particularly by the high-earners, is because the value of income received from a UBI scheme in low-income groups is higher than those higher up in the income distribution. The study recruited only a small number of high-earners and therefore, a larger proportion of high-income group respondents could have led to a significant change in the overall outcome of the survey. Upon further questioning, respondents felt that the freedom to spend their money with the potential introduction of a UBI scheme was more lucrative than opting for any other government scheme. Although it is evident that government schemes, such as the likes of Pradhan Mantri Jan Arogya Yojana, offer more monetary benefits than the UBI scheme, the freedom of the populace to choose how the income is spent is why they opt for a UBI Scheme. Therefore, respondents in the lower-income groups suggest a perceived need for direct financial assistance to alleviate poverty and enhance financial security.

The debate between targeted and universal UBI schemes reflects the trade-offs between administrative efficiency and inclusivity. There are two perspectives to this: first, for high-earners, the value of UBI is insignificant; and, second, if they received a UBI payment, this would be used in major consumer purchases, as shown in Table 4. As per the survey, the respondents believe that a targeted or quasi-UBI scheme would suit them best and benefit those who need it the most. There have always been prerequisites and intricate eligibility criteria for the existing government schemes. Therefore, UBI identification and distribution would indeed initially put an additional burden on the administration but gradually it will take the shape of any other government scheme that exists.



Income Level	Major Consumer Purchase	Major financial goals such as education, home ownership, or small business development	Save the money for future	Grand Total
₹1,50,000 – ₹1,99,999	0.00%	54.55%	45.45%	100.00%
₹16,000 – ₹49,999	0.00%	11.76%	88.24%	100.00%
₹2,00,000 – ₹2,99,999	0.00%	50.00%	50.00%	100.00%
₹3,00,000 – ₹4,99,999	11.11%	22.22%	66.67%	100.00%
₹5,00,000 +	96.34%	2.44%	1.22%	100.00%
₹50,000 – ₹79,999	0.00%	26.98%	73.02%	100.00%
₹80,000 – ₹1,49,999	0.00%	16.42%	83.58%	100.00%
0 – ₹15,999	0.00%	66.67%	33.33%	100.00%
(blank)	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
Grand Total	**30.77%**	**17.69%**	**51.54%**	**100.00%**

Table 4: Relationship between Income level (Q6) and probable investment of UBI-gained income (Q14)

Analysing the spending behaviour of respondents is essential for UBI implementation. The survey findings indicate that a significant proportion of respondents intend to save their UBI-generated income, particularly those in lower-income brackets. This propensity for saving reflects a desire to build financial security and resilience against future uncertainties. However, there are notable variations in spending behaviour across income groups, with higher-income individuals more inclined to allocate funds towards major consumer purchases. This distinctly exacerbates the need for financial security in the lower-income groups. Retroceding to the discussion on the implementation of a quasi-UBI scheme, we see that the higher income groups are likely to make a major consumer purchase which reflects the existing financial security they have. Hence, implementing a targeted UBI scheme for people near the poverty line would allow for greater effectiveness in terms of raising standards of living in India as they do not have financial security. Policymakers should consider these spending patterns when designing UBI payment structures and accompanying financial literacy initiatives to promote responsible financial management and long-term economic stability.

The common proposition of income effect on the labour market can be considered as a flawed argument. The majority of the respondents across income groups wanted to continue working the same way they do even when a UBI scheme is implemented. However, there was an observable pattern in the results: the responses varied with age and it can be concluded that the elderly (age 60+) in the lower income brackets would prefer quitting work or reducing working hours when a UBI is offered to them. Labour participation is likely to remain the same, although, the labour supply can potentially increase. 70.5% of the respondents said that they would contribute more monetarily to the education of their children and families if a basic income is potentially offered to them. A rise in literacy rates will increase the supply of high-skilled labour.

A number of factors and parameters can be researched to incorporate quality future research on this avenue. First, longitudinal studies can be conducted to examine the long-term effects of UBI on economic behaviour, labour market dynamics, and social well-being. These studies should track participants over an extended period to observe changes in employment patterns, income distribution, and quality of life. This can potentially provide a more robust analysis of UBI's sustained impact. Second, it is paramount to investigate the psychological and social effects of UBI. Understanding how receiving a basic income influences individuals' mental health, social relationships, and community engagement can provide a more comprehensive view of its benefits and drawbacks. Qualitative research methods, such as in-depth interviews and focus groups, could complement quantitative surveys to capture such impact. Moreover, future research should delve into the fiscal implications of UBI implementation. Detailed cost-benefit analyses and simulations of various funding models (e.g., tax reforms, and reduction of existing subsidies) can help policymakers and government to understand the financial feasibility and economic trade-offs involved. Exploring the potential for digital financial inclusion, especially in remote and underserved areas, is also critical to ensuring efficient and equitable distribution of UBI. Another promising avenue is examining the intersection of UBI with technological advancements and automation. As automation continues to disrupt traditional employment sectors, research should explore how UBI can mitigate job displacement and support reskilling initiatives. Investigating the role of UBI in fostering innovation and entrepreneurship, particularly among marginalized groups, can also reveal pathways for inclusive economic growth.

Works Cited

Aizenman, N. (2023, December 7). *Key findings released in Kenya Universal Basic Income Experiment*. Alaska Public Media. https://alaskapublic.org/2023/12/07/key-findings-released-in-kenya-universal-basic-income-experiment/

Bamney, A., & Tiwari, D. (2020). Study on willingness to use Non-motorized modes in a tier 3 city: A case study in India. *Transportation Research Procedia*, *48*, 2280-2295

Bardhan, P. (2017). Universal basic income–its special case for India. *Indian Journal of Human Development*, *11*(2), 141-143.

Bonnet, F., Vanek, J., & Chen, M. (2019). Women and men in the informal economy: A statistical brief. International Labour Office, Geneva, 20, 22.

Colledge, Mike, and Chris Martyn. 2017. Public Perspectives: Universal Basic Income. *Ipsos*. June 14. Available online: https://www.ipsos.com/sites/default/files/2017-06/public-perspectives-basic-universal-income-2017-06-13-v2.pdf

Drèze, J., & Khera, R. (2017). Recent social security initiatives in India. World Development, 98, 555-572.

Furman, J., & Seamans, R. (2019). AI and the Economy. Innovation policy and the economy, 19(1), 161-191.

Ghose, A. K. (2017). Universal Basic Income in India: Some Comments. *Indian Journal of Human Development*, *11*(2), 177-179.

Government of India, National Health Authority, (2019). About Pradhan Mantri Jan Arogya Yojana (PM-JAY). Available at: https://nha.gov.in/PM-JAY.html

Hamilton, L., Despard, M., Roll, S., Bellisle, D., Hall, C., & Wright, A. (2023). Does frequency or amount matter? An exploratory analysis the perceptions of four universal basic income proposals. Social Sciences, 12(3), 133.

Hamilton, L., Yorgun, M., & Wright, A. (2021). "People Nowadays Will Take Everything They Can Get": American Perceptions of Basic Income Usage. *Journal of Policy Practice and Research*, 1-19.

Joshi, V. (2017). Universal basic income supplement for India: a proposal. Indian Journal of Human Development, 11(2), 144-149.

Kumar, V., & Kanojia, S. (2017). The Idea of Universal Basic Income in India: An Analysis. Economic Survey, 2016, 17.

Kundu, T. (2016, August 27). *Is it time for a universal basic income in India?*. Mint. https://www.livemint.com/Sundayapp/ACG8sl4BvWzWBCH3orUMjO/Is-it-time-for-a-universal-basic-income-in-India.html

Ministry of Social Affairs and Health. (2020, May 6). *Results of the basic income experiment: Small employment effects, better perceived economic security and mental wellbeing*. Ministry of Social Affairs and Health.

https://stm.fi/en/-/perustulokokeilun-tulokset-tyollisyysvaikutukset-vahaisia-toimeentulo-ja-psyykkinen-terveys-koettiin-paremmaksi

Radhakrishna, R. (2017). Is India Ready to Implement Universal Basic Income Scheme?. Indian Journal of Human Development, 11(2), 200-202.

Rengasamy, K., & Kumar, B. S. (2012). State level performance of MGNREGA in India: A comparative study. *International Multidisciplinary Research Journal*, *1*(10).

Richards, D. R., & Steiger, T. L. (2021). Value orientations and support for guaranteed income. *Social Science Quarterly*, *102*(6), 2733-2751.

Richardson, N. & Marie Antonello. (n.d.). People at Work 2023: A Global Workforce View. In *People at Work 2023: A Global Workforce View* (pp. 3–7). https://www.adpri.org/wp-content/uploads/2023/04/People-at-Work-2023-A-Global-Workforce-View-1.pdf

Rycroft, R. S. (Ed.). (2017). The American Middle Class [2 volumes]: An Economic Encyclopedia of Progress and Poverty [2 volumes]. Bloomsbury Publishing USA.

Scott Santens, "Evidence and More Evidence of the Effect on Inflation of Free Money," November 2014, https : //medium.com/basic-income/evidence-and-more-evidence-of-the-effect-on-inflation-of-free-money-a3dcc2a9ea9e.

Smedley, S. (2017, September 8). *Half of UK adults would support universal basic income in principle.* IPSOS. https://www.ipsos.com/en-uk/half-uk-adults-would-support-universal-basic-income-principle

Teather-Posadas, E. (2017). Universally Basic: An Ethical Case for Universal Basic Income.

Van Parijs, Philippe (2004). Basic Income: A Simple and Powerful Idea for the Twenty-First Century. Politics and Society 32 (1):7-39

World Bank Report. (2016). *India's Poverty Profile.* Washington DC: The World Bank.

World Bank Group. (2024). *Poverty and Equity Briefs.* Available from: https://www.worldbank.org/en/topic/poverty/publication/poverty-and-equity-briefs

How to introduce microalgae in wastewater treatment and as a fuel in developing countries

By Chiho Hayashi

Abstract

A technology that uses microalgae cultivated in sewage treatment plants to produce oil is now attracting attention around the world. Not only is this oil a sustainable source of energy, but this means of sewage treatment also reduces the labor involved in removing environmentally harmful nutrients and heavy metals from sewage. Developing countries tend to be price-conscious and use fuels that are not environmentally friendly, but these countries should transition to sustainable energy to help halt overall climate change. In addition, some parts of developing countries have underdeveloped sewage systems, and this infrastructure is urgently needed. However, little research has been done specifically on cultivating microalgae in sewage and using them as a biological oil in developing countries. By reviewing literature on current problems of energy production and sewage systems in developing countries, as well as barriers and solutions to implement microalgae in both wastewater treatment and fuel production, this paper shows why microalgae should be used to treat wastewater and produce fuel in developing countries, and how this can be achieved. Wastewater treatment using microalgae and also the use of energy from the resulting biomass has many benefits, including energy security, resource conservation, and sustainability of the energy supply. Also, this is a promising approach to tackle two of the main problems existing in developing countries, namely transitioning to renewable fuels and sewage treatment. This is not widespread because of the lack of technical, political and business expertise. For the introduction of microalgae technology into developing countries, we need trained engineers who plan the microalgae cultivation, competent and technically aware politicians who support these enterprises, and social awareness of this technology's benefits. We need to train technicians by promoting Science, Technology, Engineering, Mathematics, Arts (STEAM) education and educate politicians who understand the importance of technology in these countries to improve the technology, politics, and business related to microalgae. The use of microalgae fuels should be encouraged in developing countries, as the widespread use of this technology will contribute to a sustainable society, lead to fuel independence in developing countries, and improve sanitation without incurring costs.

Keywords: Microalgae/Bio oil/Developing country/sustainability

Introduction

Microalgae are currently attracting worldwide attention due to their wide range of applications and potential. Their uses range from fuel production, water treatment, food, cosmetics, biopharmaceuticals (Khan et al., 2018). Microalgae are photosynthetic organisms that can live in a variety of environments, such as lakes, seas, ponds, and rivers (Mandal & Mallick, 2014). Microalgae have the potential to collect solar energy 10–50 times more efficiently than terrestrial organisms, even though the photosynthetic mechanism of microalgae itself is comparable to that of other plants (Mandal & Mallick, 2014). There exist two types of microalgae cultivation methods: open-pond systems and closed-culture systems. Open systems consist of open raceway ponds and allow conventional growth under natural sunlight. Closed systems use photo bioreactors (PBR), interrupting the exchange of air between the inside of the device and the outside world, thus preventing contamination and control culture-parameters. Both have advantages and disadvantages, so they are used differently depending on the cultural location and the purpose of the culture. Importantly, they are also of interest as a third-generation biofuel that does not compete with food. This is because photosynthesis allows microalgae to produce high-energy lipids in their bodies. The lipids are extracted and converted to produce biofuel (Alishah Aratboni et al., 2019). In addition, microalgae are also used in environmental applications because they can absorb CO_2 and nutrients, such as phosphorus. For example, they can be used for nutrient removal from wastewater and CO_2 sequestration, making them indispensable from an environmental perspective (Khan et al., 2023). Microalgae could be particularly useful in developing countries, where wastewater treatment is often not adequate and fossil fuels are still relied on for energy.

Water is indispensable for all human activities and economic development. However, activities using water, such as agriculture, industry, and power generation, produce dirty wastewater (Silva, 2023). Because the demand for clean water is increasing rapidly due to the rapid increase in the world's population, clean water is in shortage and wastewater facilities are unable to keep up with the volume of wastewater in some areas (Qadir et al., 2010). In developing countries in particular, untreated water is often discharged into rivers and other bodies of water (Jones et al., 2021). As sewage water contains high levels of organic matter and toxic substances, it pollutes the water environment and causes a variety of environmental and health problems (Kesari et al., 2021).

Furthermore, with ongoing climate change, it is becoming increasingly important for countries to transition to renewable energy. This represents all forms of energy that originate from solar, geophysical, and biological sources are replenished by nature at a rate faster than they are utilized (Adenle et al., 2013). Also, it is clearly stated in COP28 as a global goal to triple the renewable energy generation capacity and double the rate of energy efficiency improvements by 2030 (COP28: Global Renewables and Energy Efficiency Pledge, 2023). Subsequently, annual investment and energy efficiency would have to increase by more than five times the current level (Vanegas Cantarero, 2020). However, as it stands, the pace of conversion to renewable energies is also slow due to technical, economic and social factors (Vanegas Cantarero, 2020). Here, developing countries are a major key to more rapid implementation of radical energy reforms, as their energy demand is also increasing rapidly and their energy installations are just beginning to be built. Developing countries are thus looking for new energy sources and fuels that are not conventional fossil fuels (Zou et al., 2016).

Given these problems in developing countries and the various outstanding capabilities of microalgae, microalgae could be an ideal solution to both fuel production and wastewater treatment problems in developing countries (Nwoba et al., 2020). This is an active area of research, with China, Spain, Brazil, the USA, Japan, Taiwan, Israel and Germany being major producers of microalgae biomass and derived products, and the USA, China, Spain, France, Australia and India are known as the most productive countries in the microalgae sector (Silva et al., 2020). Although research using microalgae is active throughout the world, there tends to be less research on microalgae in many developing countries where the technology could be most useful in solving their countries' serious fuel and wastewater treatment problems by using microalgae (Adenle et al., 2013). To facilitate research in this area, I therefore ask: "What are the main problems of fuel production and wastewater treatment in developing countries, and how can microalgae solve these problems?" If this microalgae technology becomes widespread in developing countries, it will contribute to a sustainable society, fuel independence in developing countries, and improve sanitation in a cheaper way.

First, I outline the current state of fuel generation and wastewater treatment in developing nations with a survey of relevant literature. I analyze the benefits and limitations of using microalgae to address wastewater treatment and fuel production issues in underdeveloped nations. Additionally, I explore the possibilities for using microalgae to solve fuel concerns. Finally, I argued that increased technical knowledge, environmentally conscious politicians, social acceptance of biofuels, and educational improvements are the key to installing microalgae oil in developing countries.

Problems of fuel production in developing countries and the future trend

Developing countries hold the key to the energy transition. This is because the demand for electricity is estimated to double in these rapidly developing economies (Vanegas Cantarero, 2020). Such a large economic base and energy system will allow for more rapid and systematic reforms worldwide as a whole. However, many developing countries continue to rely on fossil fuels and other energy sources, despite investing heavily in new energy sources (Vanegas Cantarero, 2020). In fact, the amount of oil used is increasing in developing countries like China, India, and Brazil, even though it is decreasing or stable in developed countries (Holechek et al., 2022). This causes many problems for those countries and the world as a whole, such as environmental contamination, causing disparities between countryside and urban areas, and economic dependence on other countries.

The first reason why fossil fuels are bad is environmental impact. Indeed, fossil fuels are estimated to be responsible for more than a third of global greenhouse gas emissions (Rajesh & Majid, 2020). Additionally, they cause many problems, such as climate change and air pollution (Adenle et al., 2013). For example, according to a web blog of Center for Global Development, more than 2/3 of the world's GHG gasses are emitted by developing countries, and developing countries are a major cause of climate change today (Busch, 2015). Therefore, fossil fuels have a negative impact on the environment.

A second reason is that the use of fossil fuels creates various disparities, not only in environmental terms. For example, in remote and rural areas of developing countries, residents tend to lack access to modern, clean, and efficient fuels. As a result, they must pay more for energy used inefficiently, such as in transportation, where the amount of energy available is limited. This causes problems of social inequality and impediments to social progress within developing countries (Vanegas Cantarero, 2020). In addition, the indoor use of conventional biomass (e.g., dung, crop residues, firewood) and coal in daily life causes indoor pollution, with disproportionate negative effects on the health of women and children (Röder et al., 2020). The health of children and fetuses is particularly negatively affected because their defensive systems are immature (Perera, 2017). Ultimately, the use of fossil fuels leads to health inequalities between men and women and between generations.

The third reason is energy security and socio-economic aspects. Due to the imbalance in the production area of conventional fossil fuels, some countries are forced to rely on imports from other countries, leading to supply insecurity and price fluctuations, and hindering the independence of developing countries (Adenle et al., 2013; Alemzero et al., 2021). Since the financial crisis in 2008, the energy market and the financial market have become more and more connected (Wen et al., 2021). Therefore, the economies of oil-producing countries, such as those that continue to adhere to the old system of exporting oil and importing it again in hard currency, are highly volatile, heavily influenced by fluctuations in global oil prices (Adenle et al., 2013; Alemzero et al., 2021). For these reasons, developing countries' reliance on fossil fuels destabilizes their economies.

Alternatively, renewable energies have a low environmental impact due to low CO_2 emissions and can be produced in the country itself, leading to an increase in energy self-sufficiency (Vanegas Cantarero, 2020). For the merits of renewable energy being more sustainable and self-sufficient than conventional fuels, a transition to renewable energy sources in these developing countries is necessary for the sustainable development of the world. A major additional benefit of using sustainable energy is that it helps protect people from health hazards. In addition, as the international community encourages countries to use more sustainable fuels, energy transition is necessary to avoid giving the international community a negative image of our country as unsustainable (Owusu & Asumadu-Sarkodie, 2016).

Problems of wastewater treatment in developing countries

Not only do developing countries suffer problems from energy production, but they also are negatively impacted by wastewater treatment problems. Developing countries are more likely than developed countries to discharge and use untreated sewage. This is due to lack of facilities, political, technical and economic reasons, and also influenced by rapid population growth and economic development (Owusu & Asumadu-Sarkodie, 2016). As sewage contains many hazardous substances, untreated sewage has a negative impact on the global environment and human health (Karri et al., 2021).

A major problem with wastewater treatment in developing countries is that they produce more untreated sewage than developed countries (Edokpayi et al., 2017). The total amount of wastewater generated is likely to depend on Gross Domestic Product (GDP), the degree of sanitation expansion, and total population. Alternatively, the amount of untreated wastewater released depends on the low degree of economic development and urban population density (Jones et al., 2021). For example, developing countries in southern and southeastern Africa have the lowest wastewater collection and treatment rates in the world. Unfortunately, the untreated dirty water that is released ends up being used for drinking and agricultural purposes because it is a water resource that is available year-round, is high in nutrients, and requires less energy and cost (Jones et al., 2021). In addition, even when that water is treated, water plants are often only capable of treating it at lower capacities than their estimated performance due to lack of funding, poor maintenance, and inappropriate treatment in developing countries. Furthermore, the population in developing countries is growing so rapidly and industrial activity is getting so invigorated that these countries are not able to keep up with the development of infrastructure and sanitary facilities (Kesari et al., 2021). This ultimately leads to large amounts of untreated wastewater.

There are many reasons wastewater remains untreated, including problems with facilities, technology, policy, and the economy. For example, the cost of conventional wastewater treatment and management may be too expensive for rural communities and they may disapprove of the construction; if not, they may not utilize it (Huang et al., 2022). Given the cost problems and lack of human resources, governments may have difficulty making reuse schemes and policies. Institutional arrangements like sharing information and cooperating with each other to solve this water problem may be lacking, too (Akpan et al., 2020; Ayesha & Ayesha, 2023). In addition, lack of dependable energy sources, insufficient funding, outdated equipment, and an imbalance between urbanization and wastewater treatment capacity can all contribute to this insufficient handling of wastewater treatment (Ayesha & Ayesha, 2023).

Wastewater has a lot of negative impact on environments, eco-systems, and industries. The first negative impact of wastewater is on the environment. Wastewater contains a lot of organic matter, as it also contains fertilizers and other substances such as plant residues, leftover food (Antil, 2012). If this organic matter is discharged as it is, marine pollution and eutrophication, such as red tides and blue-green algae, occur, and water quality declines (Centre for Environment and Natural Resource Management, F-4, Models Exotica, St Inez, Panaji- Goa, 403001, India et al., 2018). In addition, if improperly treated, wastewater is used for irrigation or other purposes, groundwater contamination and soil salinity will occur. This can also result in reduced crop yields (Shakier et al., 2017). Thus, wastewater pollutes the ocean and groundwater, having a severe impact on the environment.

The second negative effect of untreated wastewater is damage to the health of organisms, including humans, and the destruction of ecosystems. When microplastics, which are smaller pieces of plastic, are released into rivers and oceans from sewage, marine organisms swallow them and accumulate, threatening the lives of these organisms and the humans who eat them. The harmful chemicals it contains also leach into the ocean and threaten a wide range of marine life (Some et al., 2021; Verla et al., 2019). The untreated release of pesticides into agricultural water is also ecotoxic to aquatic organisms. Furthermore, heavy metals such as lead, zinc, and mercury in industrial water are non-biodegradable and carcinogenic, causing serious health problems for all organisms, including humans (Ayesha & Ayesha, 2023; Qasem et al., 2021). Wastewater run-off contaminated with bacteria makes people who drink it more susceptible to infectious diseases. Hence, untreated wastewater with high amounts of hazardous chemicals can be harmful to individuals and ecosystems if discharged untreated.

The third negative impact is on industries such as fishing and tourism. The deterioration of the coastal environment causes large numbers of fish to die, become smaller, or decrease in quantity due to a lack of spawning grounds, negatively affecting the fishing industry and forcing some fishermen out of business (Andrews et al., 2021; Zohdi & Abbaspour, 2019). The sight of rubbish drifting into the sea and littering the beaches also has a negative impact on the tourism industry (Williams et al., 2016). This is because, for example, plastic garbage and poisonous compounds, when released in their current form, have a negative influence on the entire ecosystem and the attractiveness of the terrain. Importantly, the health of the ocean ecosystem attracts more marine recreation tourists. Snorkeling, scuba diving, and marine animal watching are popular activities because of the stunning coral reefs, great water quality, and variety of marine animals. The demise of coral reefs and the decline in the population of marine animals result in fewer tourists. Furthermore, toxics in the ocean pose considerable health risks to vacationers. Large plastic debris may hurt divers or create accidents, too. Such potential reductions in tourism are also huge blows to employment opportunities. For these reasons, the untreated wastewater has a detrimental effect on fishing and tourism, and employment.

As wastewater production continues to increase with population and economic growth, wastewater management and reuse practices, which are currently inadequate in developing countries, become more important (Jones et al., 2021). The world as a whole needs to continue researching wastewater treatment methods that are easy to adopt in developing countries, as well as making it easier for developing countries to adopt them themselves (WWAP, 2017). Therefore, cleaning the wastewater by cultivating microalgae can be a promising way to solve the wastewater problems.

Merits of microalgae oil production using wastewater

Using oil from microalgae grown in wastewater can provide energy as a way of solving energy and wastewater problems in developing countries. Wastewater treatment using microalgae and the use of energy from its microalgae have many benefits, including energy security, resource conservation, and sustainability of the energy supply. In addition, microalgae culture is particularly suitable for most developing countries for several reasons. Therefore, this way of using microalgae is a promising approach to tackling two of the main problems existing in developing countries mentioned above.

Wastewater treatment using microalgae has many advantages over conventional methods, including being cheaper, environmentally friendly, and nutrient-rich (Nwoba et al., 2020). First, the cultivation of microalgae under sewage can improve and maintain water quality at a lower cost and be more environmentally friendly than previous water purification technologies. Microalgae culture can also contribute to reducing secondary water pollution, improving environmental balance, and reducing CO2 emissions (Merlo et al., 2021; Razzak et al., 2017). In addition, microalgal cultivation has many advantages beyond the cost and environmental aspects that are often cited as benefits. For example, it requires less space and water for cultivation, can be grown locally and sustainably, and can create employment opportunities for women as well as men. Cultivation and distribution facilities are already in place. Microalgae used for wastewater treatment can be used for secondary purposes including fuel production (Merlo et al., 2021; Razzak et al., 2017).

The use of sewage to cultivate microalgae and use it as fuel leads to independence and sustainability in energy supply, local revitalisation and resource conservation (Merlo et al., 2021). Firstly, from a socio-economic perspective, it enables security, diversity, independence and continuity of energy supply. As mentioned earlier, fossil fuels have so far been unevenly distributed, so countries without resources have had to rely on other countries for much of their energy. Because of the uncertain outlook for world affairs due to wars and other factors, vulnerability of transport infrastructure, exchange rate fluctuations and geopolitical tensions could lead to serious energy shortages (Khan et al., 2021; Zakeri et al., 2022). However, sewage can be generated in any region as long as people live there, so microalgae can be cultivated in any country. In addition, renewable energy sources such as wind power are susceptible to climate and weather influences and may be introduced and effective in some areas and not in others, or may not provide energy for a long period of time (Breslow & Sailor, 2002). Microalgae, on the other hand, can withstand weather changes to a certain extent and thus provide a stable energy supply especially in closed bioreactors. Therefore, adding the option of microalgae biomass to conventional fossil fuels and renewable energies is socio-economically useful (Merlo et al., 2021).

Second, securing energy from microalgae is a socially beneficial option because it promotes local employment that can revitalize local industry. This is because microalgae can be produced in the countryside, allowing energy to be generated in rural areas where there is sufficient land. This makes it cheaper to transport energy within rural areas than to source energy from distant urban areas. For example, factors such as the ability to promote more renewable energy, the potential to create new sectors with employment potential should be important. Also, good press coverage of algae, and the landscape and odors associated with algae production will have a significant impact on whether the microalgae production is socially acceptable. (Efroymson et al., 2017). However, fuels produced from algae seem to be socially acceptable, as demonstrated in the following experiments. One consumer experiment involved decisions regarding gasoline. In four California cities, 20% of customers apparently chose biodiesel made from algae because of its sustainability factor (Efroymson et al., 2017). Sales increased by 35% at filling stations that began selling gasoline made from algae. Because algae-derived fuels offer environmental benefits over other fuels that do not offer benefits, 92% of respondents said they would be willing to purchase them (Propel Fuels, 2013). Additionally, according to a poll conducted in 2013, 40% of customers said they would be willing to pay more for fuels made from algae (summarized in Efroymson et al., 2017). Such studies show that microalgae fuel could be a widely accepted alternative for fossil fuels.

Finally, from an environmental perspective, microalgae cultivation can contribute to resource conservation and CO_2 fixation (Xu et al., 2023). Even if microalgae is used as fuel and carbon dioxide is generated, microalgae itself absorbs carbon dioxide and photosynthesises it, so it is more eco-friendly than fossil fuels that only release, and do not absorb, CO_2. Furthermore, the aspect of wastewater treatment not only preserves the water in oceans and rivers, but also the land (Merlo et al., 2021). This is because some of the water that flows into rivers and streams is absorbed into the soil as groundwater, and if wastewater is released as it is, the groundwater will also be contaminated. Groundwater contamination can also lead to soil contamination (Brunke & Gonser, 1997). Moreover, cultivating microalgae is suitable in developing countries because of its environmental factors. The amount of light, temperature, nutrient concentration, CO_2 concentration, and pH of the culture medium determine the culture efficiency of microalgae. Therefore, environmental conditions such as solar irradiance, temperature, precipitation and evaporation are important not only for open pond systems, but also for closed culture systems in terms of air conditioning costs and water costs to maintain the culture medium concentration. In particular, microalgae require solar irradiance of 4.0 kWh m-2 or more per day, temperatures of 5-35°C, and a stable sunshine duration of 6 hours or more per day (Nwoba et al., 2020). It is also desirable to have precipitation that offsets evaporation. Notably, developing countries in South Asia and Africa are often located in regions with long hours of sunshine, high temperatures, high precipitation, and suitable solar radiation concentrations (Vanegas Cantarero, 2020). These climates are often suitable for microalgal cultivation. Developing countries, moreover, tend to have more undeveloped land than developed countries. This is also advantageous for the construction of sewage treatment plants for further cultivation of algae (Nwoba et al., 2020).

Importantly, harnessing energy from microalgae could solve the problems of untreated wastewater and the need for a transition to sustainable energy in developing countries (Nwoba et al., 2020). Specifically, microalgae's ability to utilize nitrogen and phosphorus in wastewater to produce biomass and high-value compounds could be exploited, with the by-product high-value compounds also being sold to use the biomass as fuel. Municipal wastewater can promote the growth of various types of photosynthetic microalgae (Khan et al., 2018). In the perspective of energy security, resource conservation, sustainability, and environmental suitability, microalgae fuels are promising options for developing countries.

Overcoming challenges of microalgae oil production using wastewater

The development and management of the domestic algae industry presents several challenges for emerging nations. The nations and organizations that create the technology and those who put it into practice have reduced access to human capital with knowledge and data in the areas of technology, education, politics, and business than other developed countries (Vanegas Cantarero, 2020). To solve these problems, it is necessary to foster more highly skilled technicians by promoting Science, Technology, Engineering, Mathematics, Arts (STEAM) education, which is an educational concept that combines science and mathematics education with creativity education. Moreover, it is also important to educate politicians who understand the importance of technology in these countries (Brannstrom et al., 2022; Malamatenios, 2016). By educating these technicians and politicians, we must develop technologies to more accurately simulate changes in microalgae culture environments, produce economically attractive by-products (Merlo et al., 2021).

In the first place, technological advancements are needed. Since microalgae are living organisms, the rate of sewage treatment, biomass production depend on the organism's growing conditions. This is modified by technical equipment, environmental factors such as outdoor conditions, seasonality, and lighting, geographical conditions cannot be predicted without location-specific simulations (Merlo et al., 2021). Therefore, engineers must plan facility construction and other activities under conditions appropriate to the country's climate and infrastructure maturity while individualizing, optimizing, and simulating growing conditions (Zhu et al., 2014). In addition, microalgal fuel is currently expensive when compared to other fuels. This is because it is very costly to separate the algal biomass from the growth medium, such as wastewater, and extra money is spent if unwanted species are mixed in. Therefore, technology must be developed to produce microalgal fuel at a lower cost while taking energy efficiency and sustainability into consideration (Vanegas Cantarero, 2020). For example, microalgal fuels can be made economically viable by producing nutraceuticals, fertilizers, livestock feed, cosmetics, and other products as a secondary source of income (Rizwan et al., 2018). Also, given that huge capital investments are often difficult, it may be better to first improve the efficiency, affordability, and reliability of energy systems to introduce algal fuels in developing countries, adopting technologies that are already commercially available, and then gradually scaling up to achieve mass production and domestic diffusion (Adenle et al., 2013).

In order to develop advanced technology in microalgae cultivation, good technicians are necessary. This means it requires the training and education of superior technicians. And to meet society's needs, these educations must be more rapidly adapted to this demand. This is because it is expected that 60% of the human resources involved in renewable energy will not be those who have worked in manufacturing or construction, but those with advanced knowledge who have completed four or more years of college (Brannstrom et al., 2022). To achieve this, it is necessary to increase the percentage of students who go on to college. Currently, most developing countries have a university enrollment rate of less than 40%, and some countries have a rate of a rate of less than 10%. This is in contrast to developed countries, where many countries have over 60–70% and the shortage of engineers in the renewable energy industry is due to the lack of engineering majors (Malamatenios, 2016). To construct microalgae facilities, design engineers with special knowledge are also needed. For this purpose, people with expertise in electrical engineering, urban engineering, mechanical engineering, environmental and geo-environmental studies, and information engineering are needed. And these specialties are not only necessary for new college students but also for less specialized professionals to work in the renewable energy world.

Another reason for the lack of progress in spreading this education is on the part of education providers. The huge investment required to open and equip new universities and faculties is one of the reasons why universities and governments are slow to establish new renewable energy-related faculties, and why universities are slow to make the decision to do so in the first place. Even if a new department is established, the teachers' understanding of the field may not have caught up with the new discipline, and new media and practical training may be burdensome for the teachers (Widya et al., 2019). This may make it difficult to properly assess and point out issues to appropriate students and may hinder student motivation and confidence (Mulang, 2021). In this way, to make higher education more accessible to people, the government and private sector need to increase scholarships and developed countries need to support and create more schools. In addition, governments and the world must continue to communicate more about the importance of education in order to change people's attitudes. The government must also create a new system, such as making training mandatory for all university teachers to regularly teach what is being done at the cutting edge of the world, about the environment, information technology, and similar topics.

Developing countries need not only engineers but also technically conscious politicians who can help promote microalgae fuels. According to a technology news article, human resource needed is the people who know not only the economic effects, and international and national government projects but also the importance of technologies, such as artificial intelligence, social media and data collection (Cawley & Cawley, 2020). This allows politicians to build governments that can provide financial incentives, tax exemptions, utility regulations, auctions, and tenders to help microalgae fuel companies keep costs low and promote the widespread use of microalgae fuels as energy (Vanegas Cantarero, 2020). At the same time, these politicians would be able to legislate, organize, regulate, and control the establishment of algae biofuel companies. These will lead to the promotion of imported and genetically modified technology research, good treatment in the business market, and more efficient and lower costs for the algae industry. To educate technology-oriented politicians, arts and sciences fusion education and interdisciplinary education are needed. These education will surely enable people to view things from various perspectives and widen their view and knowledge (Klaassen, 2018). This is a global trend, but it will produce human resources who have some understanding of both society and science and technology.

Though I discussed how to promote microalgae oil , I found that most algae have the ability to absorb heavy metals and that there are seldom microalgae that do not absorb harmful heavy metals at all. This heavy metal absorption is thought to involve the expression of intracellular ligands to make metal complexes, the operation of efflux pumps to excrete metal ions, and the abduction of heavy metals via polyphosphate, metallothionein, and phytochelatin, which bind metals and surface proteins (Alishah Aratboni et al., 2019). It was also found that the use of algae cultivated in sewage as feed, cosmetics, or health food may be regulated in some countries due to health concerns (Su et al., 2023). Therefore, if the concentration of heavy metals is too high, dilution is required to reduce the concentration of heavy metals in sewage to levels that are not harmful to health. This dilution requires clean pure water or rainwater, which can be a bit costly additionally. Heavy metal enrichment in microalgae may also adversely affect even microalgae. The reason why this is the problem is that it can affect the properties of the microalgal fuels and the composition of their emissions. It is necessary to determine the concentration of heavy metals in sewage at the limit of their use for fuel production and also to carefully study the effects of these heavy metals on fuels and apply them to cultivation. If the wastewater has very high metal concentrations, conversely, it may be a good idea to remove the metals for reuse (Das & Poater, 2021).

Microalgae fuels, which can be cultivated cheaply by using wastewater in this way and already generate funds from wastewater treatment, would be economically feasible by commercializing the by-products. Engineers would have to develop more efficient technologies in terms of time, cost-effectiveness and many other aspects to develop a product that consumers would like (Merlo et al., 2021). Also important for the introduction of this oil to developing countries are the technicians who actually plan the cultivation of microalgae and the technically aware politicians who manage these enterprises. In order to train these individuals, STEAM education and interdisciplinary education are necessary (Malamatenios, 2016). These solutions will make it possible to popularize microalgae fuels in developing countries in the near future.

Conclusion

Developing countries, with their rapidly growing populations, are key to the world's energy transition, but their continued use of fossil fuels is causing environmental pollution and dependence on other countries (Vanegas Cantarero, 2020; Holechek et al., 2022; Adenle et al., 2013). In addition, wastewater, which is harmful to the environment and human health, is still being released untreated in these countries due to political, technological, and economic reasons (Jones et al., 2021). A promising solution to address both problems is to use the microalgae to treat sewage water and utilize the produced lipids as a sustainable fuel (Merlo et al., 2021; Razzak et al., 2017; Nwoba et al., 2020). Improvement of technology, more technologically-minded politicians, and higher education is necessary to implement this solution (Vanegas Cantarero, 2020).

Though I discussed the benefits of microalgae oil using sewage sludge, how we can make more safe microalgae fuel and other by-products should be researched more (Su et al., 2023). Dilution is one of the possible ways to solve this, but research is needed to find more effective ways to make this microalgae oil safe in an economically feasible manner. In conclusion, since developing countries have problems in the areas of fuel production and wastewater treatment, the use of microalgae fuels using sewage should be encouraged (Nwoba et al., 2020). This is because the widespread use of this technology will contribute to a sustainable society, lead to fuel independence, and improve sanitation without incurring costs. To realize this, technical, political, and educational improvements are necessary.

In this paper, I showed that non-renewable energy used in most developing countries leads to major problems, such as pollution and energy insecurity. Secondly, I illustrated that wastewater in developing countries is still untreated, causing health problems and environmental degradation (Vanegas Cantarero, 2020; Jones et al., 2021). To solve these two fundamental issues, I proposed using microalgae to treat wastewater and use it as a renewable energy source. This can be a cheaper, environmentally-friendly method of wastewater treatment, and stably supplied, locally-produced fuel for developing countries (Nwoba et al., 2020). However, the effective cultivation of microalgae is still difficult due to environmental variability and technical constraints, the fuel is more expensive than conventional fuels, and policies to promote this fuel are insufficient. Finally, I argued that technical, political, social, economic,and educational improvements are the key to installing microalgae oil in developing countries.

Acknowledgement

Research guidance and editorial support was provided by Nathalie Chardon of Indigo Research, and funding for these services was provided by Junichi Hayashi. I attest that the ideas, graphics, and writing in this paper are entirely my own.

Works Cited

Adenle, A. A., Haslam, G. E., & Lee, L. (2013). Global assessment of research and development for algae biofuel production and its potential role for sustainable development in developing countries. *Energy Policy*, *61*, 182–195. https://doi.org/10.1016/j.enpol.2013.05.088

Akpan, V. E., Omole, D. O., & Bassey, D. E. (2020). Assessing the public perceptions of treated wastewater reuse: Opportunities and implications for urban communities in developing countries. *Heliyon*, *6*(10), e05246. https://doi.org/10.1016/j.heliyon.2020.e05246

Alemzero, D. A., Sun, H., Mohsin, M., Iqbal, N., Nadeem, M., & Vo, X. V. (2021). Assessing energy security in Africa based on multi-dimensional approach of principal composite analysis. *Environmental Science and Pollution Research*, *28*(2), 2158–2171. https://doi.org/10.1007/s11356-020-10554-0

Alishah Aratboni, H., Rafiei, N., Garcia-Granados, R., Alemzadeh, A., & Morones-Ramírez, J. R. (2019). Biomass and lipid induction strategies in microalgae for biofuel production and other applications. *Microbial Cell Factories*, *18*(1), 178. https://doi.org/10.1186/s12934-019-1228-4

Andrews, N., Bennett, N. J., Le Billon, P., Green, S. J., Cisneros-Montemayor, A. M., Amongin, S., Gray, N. J., & Sumaila, U. R. (2021). Oil, fisheries and coastal communities: A review of impacts on the environment, livelihoods, space and governance. *Energy Research & Social Science*, *75*, 102009. https://doi.org/10.1016/j.erss.2021.102009

Antil, R. S. (2012). *Impact of Sewage and Industrial Effluents on Soil-Plant Health*. https://doi.org/10.5772/37403

Ayesha, T., & Ayesha, M. (2023). Untreated Wastewater Reasons and Causes: A Review of Most Affected Areas and Cities. *International Journal of Chemical and Biochemical Sciences*, *23*.

Brannstrom, C., Ewers, M., & Schwarz, P. (2022). Will peak talent arrive before peak oil or peak demand?: Exploring whether career choices of highly skilled workers will accelerate the transition to renewable energy. *Energy Research & Social Science*, *93*, 102834. https://doi.org/10.1016/j.erss.2022.102834

Breslow, P. B., & Sailor, D. J. (2002). Vulnerability of wind power resources to climate change in the continental United States. *Renewable Energy*, *27*(4), 585–598. https://doi.org/10.1016/S0960-1481(01)00110-0

Brunke, M., & Gonser, T. (1997). The ecological significance of exchange processes between rivers and groundwater. *Freshwater Biology*, *37*(1), 1–33. https://doi.org/10.1046/j.1365-2427.1997.00143.x

Busch, J. (2015, August 15). *Climate change and development in three*. Center for Global Development. Retrieved June 21, 2024, from https://www.cgdev.org/blog/climate-change-and-development-three-charts

Cawley, C., & Cawley, C. (2020, November 2). *Politicians are too out of touch to make laws about tech*. Tech.co.

https://tech.co/news/politicians-out-of-touch-laws-regulate-tech-2018-10

Centre for Environment and Natural Resource Management, F-4, Models Exotica, St Inez, Panaji- Goa, 403001, India, Sonak, S., Patil, K., The Energy and Resources Institute, Alto-St Cruz, Bambolim, Goa, 403 202, India, Devi, P., & Bioorganic Chemistry Lab, National Institute of Oceanography, CSIR, Dona Paula, Goa, India. (2018). Causes, Human Health Impacts and Control of Harmful Algal Blooms: A Comprehensive Review. *Environmental Pollution and Protection*, *3*(1), 40–55. https://doi.org/10.22606/epp.2018.31004

COP28: Global Renewables and Energy Efficiency Pledge. (2023). https://www.cop28.com/en/global-renewables-and-energy-efficiency-pledge

Das, T. K., & Poater, A. (2021). Review on the Use of Heavy Metal Deposits from Water Treatment Waste towards Catalytic Chemical Syntheses. *International Journal of Molecular Sciences*, *22*(24), 13383. https://doi.org/10.3390/ijms222413383

Edokpayi, J. N., Odiyo, J. O., & Durowoju, O. S. (2017). Impact of Wastewater on Surface Water Quality in Developing Countries: A Case Study of South Africa. In H. Tutu (Ed.), *Water Quality*. InTech. https://doi.org/10.5772/66561

Efroymson, R. A., Dale, V. H., & Langholtz, M. H. (2017). Socioeconomic indicators for sustainable design and commercial development of algal biofuel systems. *GCB Bioenergy*, *9*(6), 1005–1023. https://doi.org/10.1111/gcbb.12359

Holechek, J. L., Geli, H. M. E., Sawalhah, M. N., & Valdez, R. (2022). A Global Assessment: Can Renewable Energy Replace Fossil Fuels by 2050? *Sustainability*, *14*(8), 4792. https://doi.org/10.3390/su14084792

Huang, Y., Wu, L., Li, P., Li, N., & He, Y. (2022). What's the cost-effective pattern for rural wastewater treatment? *Journal of Environmental Management*, *303*, 114226. https://doi.org/10.1016/j.jenvman.2021.114226

Jones, E. R., Van Vliet, M. T. H., Qadir, M., & Bierkens, M. F. P. (2021). Country-level and gridded estimates of wastewater production, collection, treatment and reuse. *Earth System Science Data*, *13*(2), 237–254. https://doi.org/10.5194/essd-13-237-2021

Karri, R. R., Ravindran, G., & Dehghani, M. H. (2021). Wastewater—Sources, Toxicity, and Their Consequences to Human Health. In *Soft Computing Techniques in Solid Waste and Wastewater Management* (pp. 3–33). Elsevier. https://doi.org/10.1016/B978-0-12-824463-0.00001-X

Kesari, K. K., Soni, R., Jamal, Q. M. S., Tripathi, P., Lal, J. A., Jha, N. K., Siddiqui, M. H., Kumar, P., Tripathi, V., & Ruokolainen, J. (2021). Wastewater Treatment and Reuse: A Review of its Applications and Health Implications. *Water, Air, & Soil Pollution*, *232*(5), 208. https://doi.org/10.1007/s11270-021-05154-8

Khan, K., Su, C.-W., Tao, R., & Umar, M. (2021). How do geopolitical risks affect oil prices and freight rates? *Ocean & Coastal Management*, *215*, 105955. https://doi.org/10.1016/j.ocecoaman.2021.105955

Khan, M. I., Shin, J. H., & Kim, J. D. (2018). The promising future of microalgae: Current status, challenges, and optimization of a sustainable and renewable industry for biofuels, feed, and other products. *Microbial Cell Factories*, *17*(1), 36. https://doi.org/10.1186/s12934-018-0879-x

Khan, S., Thaher, M., Abdulquadir, M., Faisal, M., Mehariya, S., Al-Najjar, M. A. A., Al-Jabri, H., & Das, P. (2023). Utilization of Microalgae for Urban Wastewater Treatment and Valorization of Treated Wastewater and Biomass for Biofertilizer Applications. *Sustainability*, *15*(22), 16019. https://doi.org/10.3390/su152216019

Klaassen, R. G. (2018). Interdisciplinary education: A case study. *European Journal of Engineering Education*, *43*(6), 842–859. https://doi.org/10.1080/03043797.2018.1442417

Malamatenios, C. (2016). Renewable energy sources: Jobs created, skills required (and identified gaps), education and training. *Renewable Energy and Environmental Sustainability*, *1*, 23. https://doi.org/10.1051/rees/2016038

Merlo, S., Gabarrell Durany, X., Pedroso Tonon, A., & Rossi, S. (2021). Marine Microalgae Contribution to Sustainable Development. *Water*, *13*(10), 1373. https://doi.org/10.3390/w13101373

Mulang, H. (2021). The Effect of Competences, Work Motivation, Learning Environment on Human Resource Performance. *Golden Ratio of Human Resource Management*, *1*(2), 84–93. https://doi.org/10.52970/grhrm.v1i2.52

Nwoba, E. G., Vadiveloo, A., Ogbonna, C. N., Ubi, B. E., Ogbonna, J. C., & Moheimani, N. R. (2020). Algal Cultivation for Treating Wastewater in African Developing Countries: A Review. *CLEAN – Soil, Air, Water*, *48*(3), 2000052. https://doi.org/10.1002/clen.202000052

Owusu, P. A., & Asumadu-Sarkodie, S. (2016). A review of renewable energy sources, sustainability issues and climate change mitigation. *Cogent Engineering*, *3*(1), 1167990. https://doi.org/10.1080/23311916.2016.1167990

Perera, F. (2017). Pollution from Fossil-Fuel Combustion is the Leading Environmental Threat to Global Pediatric Health and Equity: Solutions Exist. *International Journal of Environmental Research and Public Health*, *15*(1), 16. https://doi.org/10.3390/ijerph15010016

Qadir, M., Wichelns, D., Raschid-Sally, L., McCornick, P. G., Drechsel, P., Bahri, A., & Minhas, P. S. (2010). The challenges of wastewater irrigation in developing countries. *Agricultural Water Management*, *97*(4), 561–568. https://doi.org/10.1016/j.agwat.2008.11.004

Qasem, N. A. A., Mohammed, R. H., & Lawal, D. U. (2021). Removal of heavy metal ions from wastewater: A comprehensive and critical review. *Npj Clean Water*, *4*(1), 36. https://doi.org/10.1038/s41545-021-00127-0

Razzak, S. A., Ali, S. A. M., Hossain, M. M., & deLasa, H. (2017). Biological CO2 fixation with production of microalgae in wastewater – A review. *Renewable and Sustainable Energy Reviews*, *76*, 379–390. https://doi.org/10.1016/j.rser.2017.02.038

Rizwan, M., Mujtaba, G., Memon, S. A., Lee, K., & Rashid, N. (2018). Exploring the potential of microalgae for new biotechnology applications and beyond: A review. *Renewable and Sustainable Energy Reviews*, *92*, 394–404. https://doi.org/10.1016/j.rser.2018.04.034

Shakir, E., Zahraw, Z., & Al-Obaidy, A. H. M. J. (2017). Environmental and health risks associated with reuse of wastewater for irrigation. *Egyptian Journal of Petroleum*, *26*(1), 95–102. https://doi.org/10.1016/j.ejpe.2016.01.003

Silva, J. A. (2023). Wastewater Treatment and Reuse for Sustainable Water Resources Management: A Systematic Literature Review. *Sustainability*, *15*(14), 10940. https://doi.org/10.3390/su151410940

Silva, S. C., Ferreira, I. C. F. R., Dias, M. M., & Barreiro, M. F. (2020). Microalgae-Derived Pigments: A 10-Year Bibliometric Review and Industry and Market Trend Analysis. *Molecules*, *25*(15), 3406. https://doi.org/10.3390/molecules25153406

Some, S., Mondal, R., Mitra, D., Jain, D., Verma, D., & Das, S. (2021). Microbial pollution of water with special reference to coliform bacteria and their nexus with environment. *Energy Nexus*, *1*, 100008. https://doi.org/10.1016/j.nexus.2021.100008

Su, M., Bastiaens, L., Verspreet, J., & Hayes, M. (2023). Applications of Microalgae in Foods, Pharma and Feeds and Their Use as Fertilizers and Biostimulants: Legislation and Regulatory Aspects for Consideration. *Foods*, *12*(20), 3878. https://doi.org/10.3390/foods12203878

Vanegas Cantarero, M. M. (2020). Of renewable energy, energy democracy, and sustainable development: A roadmap to accelerate the energy transition in developing countries. *Energy Research & Social Science*, *70*, 101716. https://doi.org/10.1016/j.erss.2020.101716

Verla, A. W., Enyoh, C. E., Verla, E. N., & Nwarnorh, K. O. (2019). Microplastic–toxic chemical interaction: A review study on quantified levels, mechanism and implication. *SN Applied Sciences*, *1*(11), 1400. https://doi.org/10.1007/s42452-019-1352-0

Wen, J., Zhao, X.-X., & Chang, C.-P. (2021). The impact of extreme events on energy price risk. *Energy Economics*, *99*, 105308. https://doi.org/10.1016/j.eneco.2021.105308

Widya, Rifandi, R., & Laila Rahmi, Y. (2019). STEM education to fulfil the 21 st century demand: A literature review. *Journal of Physics: Conference Series*, *1317*(1), 012208. https://doi.org/10.1088/1742-6596/1317/1/012208

Williams, A. T., Rangel-Buitrago, N. G., Anfuso, G., Cervantes, O., & Botero, C. M. (2016). Litter impacts on scenery and tourism on the Colombian north Caribbean coast. *Tourism Management*, *55*, 209–224. https://doi.org/10.1016/j.tourman.2016.02.008

Xu, P., Li, J., Qian, J., Wang, B., Liu, J., Xu, R., Chen, P., & Zhou, W. (2023). Recent advances in CO2 fixation by microalgae and its potential contribution to carbon neutrality. *Chemosphere*, *319*, 137987. https://doi.org/10.1016/j.chemosphere.2023.137987

Zakeri, B., Paulavets, K., Barreto-Gomez, L., Echeverri, L. G., Pachauri, S., Boza-Kiss, B., Zimm, C., Rogelj, J., Creutzig, F., Ürge-Vorsatz, D., Victor, D. G., Bazilian, M. D., Fritz, S., Gielen, D., McCollum, D. L., Srivastava, L., Hunt, J. D., & Pouya, S. (2022). Pandemic, War, and Global Energy Transitions. *Energies*, *15*(17), 6114.

https://doi.org/10.3390/en15176114

Zhu, L. D., Hiltunen, E., Antila, E., Zhong, J. J., Yuan, Z. H., & Wang, Z. M. (2014). Microalgal biofuels: Flexible bioenergies for sustainable development. *Renewable and Sustainable Energy Reviews*, *30*, 1035–1046. https://doi.org/10.1016/j.rser.2013.11.003

Zohdi, E., & Abbaspour, M. (2019). Harmful algal blooms (red tide): A review of causes, impacts and approaches to monitoring and prediction. *International Journal of Environmental Science and Technology*, *16*(3), 1789–1806. https://doi.org/10.1007/s13762-018-2108-x

Zou, C., Zhao, Q., Zhang, G., & Xiong, B. (2016). Energy revolution: From a fossil energy era to a new energy era. *Natural Gas Industry B*, *3*(1), 1–11. https://doi.org/10.1016/j.ngib.2016.02.001

The Effect of Mass Displacement of Syrians to Turkey due to the Syrian Civil War on Turkey's Relations with the European Union: How has the mass displacement of Syrians to Turkey caused by the Syrian civil war during Recep Tayyip Erdogan's presidency since 2014 affected Turkey’s relations with the European Union? By Aleksandra Taratorina

1. Introduction

The Syrian civil war is an ongoing and multi-faceted conflict that has escalated from peaceful protests to one of the worst humanitarian crises since World War II, spilling far beyond Syria's borders. While it is an intrastate conflict that is concentrated within one state, it is also a proxy war as many international state actors such as the United States and Turkey utilize it to advance their foreign policy interests.[130] The civil war broke out in 2011, after Syria's Bashar al-Assad's crackdown on peaceful protests sparked by the Arab Spring.[131] Syria's neighbor, Turkey, is one of the countries that has borne the grievous consequences of the conflict. One of its major ramifications has been the mass displacement of Syrians, and Turkey has been the “primary destination” for Syrian refugees.[132] Many refugees have been able to reach Europe through Turkey, which caused the 2015 European migrant crisis, and prompted the European Union (EU) and Turkey to cooperate despite having a historically complex relationship.

This paper seeks to explore how the mass displacement of Syrians to Turkey caused by the Syrian civil war during Recep Tayyip Erdogan's presidency since 2014 has affected Turkey’s relations with the European Union. Mass displacement is a global challenge expressed in “large-scale, sudden population movements, prompted by both rapid-onset ‘natural’ disasters such as floods and ‘man-made’ disasters like conflict.”[133] The question is worthy of investigation because mass displacement of Syrians to Turkey marked a significant turning point in Turkey-EU relations as well as in the power dynamic between the two actors. Ankara has used the conflict in Syria and its ramifications to its advantage as a leverage over the EU, demanding funding and closer cooperation.[134] Moreover, Turkey`s overall foreign policy has become more assertive. Finally, although the conclusions of this paper should not be speculative, they may be able to inform the potential ramifications of similar cases in the future. This paper will answer the question that has been posed by outlining the causes of the Syrian civil war, evaluating Turkey-EU relations prior to the war, discussing the 2015 European migrant crisis, and assessing the post-crisis Turkey-EU relations.

Qualitative research will be employed due to its holistic approach that allows for in-depth analysis of the case study. Qualitative approach and case studies are the typical methodology used in the field of politics. It also accounts for the fact that conflicts have many unintended consequences; therefore, a holistic approach may assist in discovering links between events or dynamics that seemed detached before. Thematic analysis of a variety of sources will be

[130] Petrini et al., *The Civil War in Syria: An Intractable Conflict with Geopolitical Implications* (2021)
[131] Polk, *Understanding Syria: from Pre-Civil war to Post-Assad* (2013)
[132] World Bank, *10 Years On, Turkey Continues Its Support for an Ever-Growing Number of Syrian Refugees* (2021)
[133] Kirbyshire et al., *Mass Displacement and the Challenge for Urban Resilience* (2017)
[134] Siccardi, *How Syria Changed Turkey’s Foreign Policy* (2021)

conducted, focusing on making connections between them to produce new insights. Scholarly articles and books are used due to the authors` expertise. Additionally, sources such as the World Bank and the United Nations are used due to their status as non-governmental organizations, meaning that they are not affiliated with any governments discussed in the paper. All these sources offer pre-existing interpretations and historical descriptions of the Turkish regime, Turkey-EU relations, and the Syrian civil war. This paper will produce new insights about the topic through thematic analysis of existing findings.

2. Historical Background: The Arab Spring & The Assad Regime

The Syrian civil war started in 2011 and is ongoing today, 12 years later. It began after the Syrian dictator Bashar al-Assad attempted to forcefully shut down peaceful protests sparked by the Arab Spring, triggering them to quickly grow into a violent insurgency.[2] The Arab Spring was a movement comprised of a series of initiatives, from peaceful protests to violent clashes between the people and the government, all against authoritarianism and for human rights across the Middle East. It started in Tunisia in 2010 and led to the removal of oppressive rulers such as Egypt's President Hosni Mubarak.[135] Soft power in the form of social media enabled the movement to spread across the region and galvanize a large number of supporters.[136]

However, in Syria, the movement took a turn. The Assad regime has been in place since 1970, when Bashar`s father, Hafez al-Assad, took power. His son became the president after his death in 2000 and adopted his authoritarianism.[137] Although his administration's policies were once regarded as innovative, they have largely proven to be economically unsuccessful and have been involved in several human rights controversies. Today, though the causes of the Syrian uprising of 2011 are complex, these failures are largely credited for bringing the Arab Spring to Syria.

In the early 2000s, Al-Assad implemented several economic liberalization programs which exacerbated economic inequality in the country. Historically, the Syrian regime has been viewed as socialist due to its social protection strategies from the 1970s onward, which resulted in the expansion of state institutions such as schools and the public sector, which ensured employment.[138] However, by the 2000s, Syria's economy had become stagnant, so Bashar al-Assad decided to marketize it and gradually dismantle the public sector, which increased poverty and economic inequality in the country due to the sudden take away of social protections such as subsidies. Moreover, the number of jobs available to Syrians decreased along with their wages while the cost of living was gradually increasing, creating more economic disparity.[9] According to William R. Polk, a former US foreign policy consultant for the Kennedy administration, people in rural areas were also affected during that time due to the four-year-long drought that Syria went through, beginning in 2006, and their dependence on agriculture.[2] In

[135] QadirMushtaq and Afzal, *Arab Spring: Its Causes And Consequences* (n.d.)
[136] Elakawi, *The Geostrategic Consequences of the Arab Spring* (2014)
[137] Weeks, *Dictators at War and Peace* (2014)
[138] Abboud, *Economic Liberalization and Social Transformations in Pre-War Syria* (2019)

summary, the effects of the drought and increasing economic inequality fed people's political discontent.

However, the Syrian uprising of 2011 cannot be attributed to only economic factors. Another failure of Bashar Al-Assad's regime was its authoritarianism and repressive measures against opposition. Prior to Bashar`s disproportional reaction to the Arab Spring, he had already been accused of human rights violations such as forcefully shutting down the Damascus Spring, a period of intense opposition activism in Syria in the early 2000s, and imprisonment of opposition members involved in the movement.[139] Ultimately, such severe and violent repressions by the state combined with economic struggles and other factors led Syrians to take to the streets in 2011, which quickly escalated into a civil war.

3. The European Union & Turkey: Pre-Crisis Relations

Two major international actors in the Syrian civil war are Turkey and the European Union. The European Union is a global governance organization founded in 1992 with the Maastricht Treaty.[140] Currently at 27 member states, the European Union is a post-World War II attempt to integrate European economies and prevent future conflict in the region.[141] Its aim to facilitate cooperation to ensure peace adheres to the international relations theory of liberalism and its main assumptions, which are that the world is a global community, and cooperation rather than conflict is the norm. Today, all EU members enjoy capabilities such as unified trade policy and free trade agreements, freedom of movement, and a shared currency.[11] As for Turkey, stretched from southeastern Europe to southwest Asia, it connects two continents together. In the context of the Syrian civil war, Turkey's location made it possible for Syrian refugees to reach the European Union. According to the UNHCR, refugees are defined as “someone who has been forced to flee his or her country because of persecution, war or violence.”[142]

Since 2014, when Recep Tayyip Erdogan was elected as Turkey's president (and has been re-elected since), he largely focused on two issues: economic development and the expansion of democracy to maintain his popularity.[143] Joining the European Union, which meant fulfilling the Copenhagen criteria, would allow Erdogan to improve democracy and human rights in Turkey, so he eagerly presided over the fulfillment of the criteria, deepening Turkey’s ties with the EU.[14] The criteria defined at the Copenhagen European Summit in 1993, also known as the Copenhagen Criteria, serve as guidelines for the countries wishing to become a European Union member. The EU starts negotiations with candidate countries only if they fulfill the requirements known as the Copenhagen “Political” Criteria, which include democracy, the rule of law, human rights, and the existence of institutions that guarantee minority rights.[144] These criteria adhere to the Democratic Peace Theory, which argues that democracies are less likely to go to war with

[139] Carnegie Middle East Center, *The Damascus Spring* (n.d.)

[140] The World Factbook, *European Union* (2021)

[141] U.S. Mission to The European Union, *About the Mission* (2023)

[142] *USA for UNHCR, What Is a Refugee? Definition and Meaning* (n.d.)

[143] Yavuz, *Secularism and Muslim Democracy in Turkey* (2009)

[144] Republic of Türkiye Ministry of Foreign Affairs, *Enlargement of the European Union* (2022)

one another due to shared values, separation of powers in the government, and more accountability that democratic leaders are held to, therefore, reiterating the European Union`s efforts to contain conflict in Europe. Erdogan supported Turkey's application for full membership of the EU; however, the application has faced significant obstacles due to his inconsistency when it comes to prioritizing democracy and protecting human rights in Turkey, causing his relations with the EU to deteriorate.

As a political leader, Erdogan is pragmatic but not ideologically committed; he struggles to choose between a West-oriented democracy and authoritarianism. M. Hakan Yavuz, Professor of Political Science at the University of Utah, describes Erdogan`s domestic policy as "local-based politics on a national scale," meaning that he strives to provide "justice for all."[14] Yavuz`s argument reveals Erdogan`s leadership style: a pro-European conservative who has to constantly adapt to the changing environment inside and outside Turkey to remain relevant and appeal to voters across the political spectrum as well as various political actors internationally. For example, he refers to Israel as a "terrorist state" but still visits due to pressure from the US; hence, he tries to appeal to actors with a variety of perspectives.[14] Therefore, when it comes to Erdogan's regime, he is torn between a democratic republic, which would strengthen Turkey's relations with the EU, and authoritarianism, which would grant him more power.

In recent years, Turkey has been experiencing democratic backsliding due to its shift to the presidential system of governance after the failed 2017 coup attempt against Erdogan and his unwillingness to relinquish some of his power. The transition granted more powers to the president, leaving many experts concerned about the future of Turkish democracy.[145] Juan Linz, a political scientist who taught at Yale University, argues that presidential systems are inherently less stable than parliamentary systems;[146] hence, they are more vulnerable to democratic backsliding. According to the Sigma Iota Rho Online Journal of International Relations, Erdogan has already been dismantling "institutions of democracy that contradict his message" and suppressing his opposition, which showcases his authoritarianism.[147]

From a realist point of view, which is one of the major theories in international relations, it would be in Turkey's national interest if Erdogan strengthened Turkish democracy as it would increase the country's chances of receiving EU membership, which comes with many benefits. However, Erdogan seems to be unwilling to promote democracy because he would have to relinquish some of his power.

Moreover, looking at Turkey as an emerging middle power provides more clarity as to why Erdogan does not fully commit to the EU and its values. Middle powers are "states that are neither great nor small in terms of international power, capacity and influence, and demonstrate a propensity to promote cohesion and stability in the world system."[148] According to The Hague Centre for Strategic Studies, there are established and emerging middle powers. Established

[145] Quamar, *The Turkish Referendum and Its Impact on Turkey's Foreign Policy* (2017)
[146] Linz, *The Perils of Presidentialism* (1990)
[147] Baghdady, *Turkey's Electoral Authoritarianism* (2020)
[148] Jordaan, *The Concept of a Middle Power in International Relations: Distinguishing between Emerging and Traditional Middle Powers* (2003)

middle powers strive to uphold liberal-democratic norms domestically and internationally, while emerging middle powers are "not necessarily committed to the current arrangement in world affairs" due to their dissatisfaction with the status quo and regional or global hegemons.[149] Turkey is an example of an emerging middle power, with its apparent regional influence and struggle to choose between two world orders, democratic and authoritarian. Erdogan seems to prioritize firm leadership over compliance to Western ideals in order for Turkey to solidify itself as its own, influential political actor, which is also expressed in Turkey's current assertive foreign policy that will be discussed further in the paper.

Altogether, Erdogan`s lack of fixed principles has prevented Turkey from being accepted into the EU, making the relations between the two actors extremely complex. The European Union has condemned Erdogan's authoritarian tendencies, with Turkey's accession talks being stalled due to its democratic backsliding. However, Erdogan continues to uphold his regime to advance Turkey as an influential political actor.

4. The 2015 European Migrant Crisis & Turkey: The Cost of the War

Despite Erdogan`s continued authoritarian rule, he has been able to deepen Turkey`s cooperation with the European Union. Mass displacement has been one of the Syrian civil war`s major ramifications and has significantly affected Turkey and the EU. According to the Council on Foreign Relations, "nearly thirteen million people—more than half the country's [Syria`s] prewar population—have been displaced" as a result of the conflict.[150] As of 2021, Turkey is the largest host of refugees in the world and the "primary destination" for Syrian refugees: it hosts a total of 4 million people, with 3.6 million from Syria.[3]

Such a high number of refugees requires a large amount of financial spendings and other resources, especially considering that, as of 2021, 98.5% of Syrians under temporary protection live in Turkish communities rather than isolated camps and hope to fully integrate into the Turkish society, obliging the Turkish government to ensure their long-term well-being.[3] Many of the communities that the refugees joined have already been facing "significant development obstacles;" therefore, "providing adequate services and support" such as education, housing, and employment for "millions of additional people" has been a "monumental challenge," the World Bank reports.[3] According to the University of British Columbia, from 2011 to 2015, Turkey spent $7.6 billion caring for 2.2 million Syrian refugees.[151] Ultimately, it proves challenging and costly for Turkey to ensure the well-being of such a massive, vulnerable population.[152]

Due to the extent of the issue, it is often referred to as Turkey`s migrant crisis, which falls under the larger 2015 European migrant crisis. The 2015 European migrant crisis was a period of an influx of refugees and asylum seekers arriving to Europe, especially to the European Union.[22] According to the Organisation for Economic Co-operation and Development, during the crisis,

[149] Strategic Monitor 2018-2019, *A Balancing Act* (n.d.)
[150] Laub, *Syria's Civil War* (2023)
[151] The University of British Columbia, *The 2015 European Refugee Crisis* (n.d.)
[152] International Crisis Group, *Turkey's Refugee Crisis: The Politics of Permanence* (2016)

the number of refugees reached the most in a single year since World War II.[153] Tim Hatton, a professor of Economics at the University of Essex, has estimated that more than 2 million refugees have entered Europe between 2015 and 2016.[154] As a result, the European Union faced challenges such as inability of border controls to manage the flow of migrants and asylum applications being distributed very unevenly among EU member states with most Syrians arriving in Germany.[155] However, it is important to note that the Syrian civil war was not the only cause of the intensified migration: the Arab Spring has caused instability in states other than Syria as well.[25] Altogether, the 2015 European migrant crisis threw the European asylum system into disarray, exposing weaknesses of the EU`s foreign policy toward refugees and forcing the organization to initiate improvement. The European Union`s cooperation with Turkey on the matter is one of the areas of policy that has seen significant changes since the crisis.

5. Shifts in Turkey-EU Relations

It is not typical for a non-democratic regime such as Turkey to host such a large number of refugees.[156] One interpretation of Turkey`s willingness to do so is that Turkey's Erdogan weaponizes refugees to gain leverage over the European Union and challenge the status quo of power distribution in the region, with power defined as the “ability to affect others to get the outcomes one wants” by Joseph Nye.[157] Prior to 2015, Turkey's relations with the EU had been complex with Turkey wanting to join the organization but not qualifying politically. During and after the 2015 European migrant crisis, Erdogan saw an opportunity to advance his interests through foreign policy, using the refugees that Turkey hosts to push for more funding from the EU through bilateral deals and to establish a more assertive policy in the Eastern Mediterranean.

5.1 Turkey-EU Refugee Deals

Turkey-EU cooperation aiming to ease the challenges that the heavy flow of refugees and asylum seekers has created is largely facilitated through refugee deals. While refugees have been defined earlier, the EU defines asylum seekers as “a third-country national or stateless person who has made an application for protection under the Geneva Refugee Convention and Protocol in respect of which a final decision has not yet been taken.”[158] In 2015, the EU and Turkey announced the implementation of a joint Action Plan with the goal to increase cooperation in order to support Syrian refugees and enhance migration management.[159] The plan stated that the EU would provide immediate humanitarian and financial assistance of 3 billion euro to Turkey.[160] Turkey, on the other hand, would improve its support of Syrian refugees and tighten its

[153] The Organisation for Economic Co-operation and Development, *Migration Policy Debates* (2015)
[154] Hatton, *European Asylum Policy Before and After the Migration Crisis* (2020)
[155] World Economic Forum, *Europe's Refugee Crisis Explained* (2015)
[156] Higashijima and Woo, *Political Regimes and Refugee Entries: Motivations behind Refugees and Host Governments* (2020)
[157] Nye and Goldsmith, *The Future of Power* (2011)
[158] Migration and Home Affairs, *Asylum Seeker* (n.d.)
[159] European Commission, *EU-Turkey Joint Action Plan* (2015)
[160] European Commission, *EU-Turkey Cooperation: A €3 Billion Refugee Facility for Turkey* (2015)

borders to prevent further Syrian migration to Europe. In return, Turkey was set to receive further political concessions, including “a revitalization of talks for the country to join the EU” and “visa-free travel for Turkish citizens.”[22] The deal ensured that refugees who arrived in Greece after March 20th, 2016, were sent back to Turkey due to its designation as a “safe third country.”[22] According to the University of British Columbia, when referring to the deal, Frans Timmermans, First Vice-President of the EU Commission for Better Regulation, Interinstitutional Relations, the Rule of Law, and the Charter of Fundamental Rights at the time, said: “In dealing with the refugee crisis, it is absolutely clear that the European Union needs to step up its cooperation with Turkey and Turkey with the European Union,” emphasizing the importance of cooperation between the two actors to the EU[22], as an institution built on the values of liberalism.

The second deal of March 2016 further strengthened the cooperation between Turkey and the EU. It granted that all new irregular migrants crossing to Greece would be returned to Turkey, and for every Syrian readmitted by Turkey, a Syrian from Turkey would be resettled to the EU. Additionally, it promised acceleration in the implementation of the EU-Turkey visa liberalization process, of preparations for the opening of new chapters in Turkey’s accession negotiations, and of disbursement of funds through the Facility for Refugees in Turkey as well as an increase in the fund amounts.[161] However, according to the report from Global Turkey in Europe, visa liberalization remains “on hold” and there has been “very limited progress regarding the opening of new chapters as part of Turkey’s EU accession process” due to the EU`s uncertainty about the future of democracy in Turkey.[162]

Altogether, these deals allow the EU to externalize migration and its management through facilitating its cooperation with Turkey. This approach still prevails in their foreign policy today, likely due to the continued disagreement regarding Turkey's accession in the EU among members.[5] In June 2021, European leaders reaffirmed their determination to build partnerships with refugees’ countries of origin, which, according to Francesco Siccardi, a senior program manager at Carnegie Europe, “undoubtedly plays into Turkey’s hands and gives Ankara leverage over the EU and its member states.”[5] The Syrian civil war allowed Turkey to demand funding from the EU as well as other privileges such as visa-free travel and acceleration of the accession process.

However, Turkey is also often the only party that fulfills its portion of the agreement, while many of the privileges that have been promised to Turkish citizens in return have not been fulfilled, which creates tension between Turkish leaders and the EU.[33] Therefore, an argument could be made that Turkey does not benefit from the refugee deals as much as it contributes to them.

5.2 Turkey's New Foreign Policy Objectives in the Eastern Mediterranean

[161] The European Council, *EU-Turkey Statement* (2016)

[162] Seeberg, The EU-Turkey March 2016 Agreement As a Model: New Refugee Regimes and Practices in the Arab Mediterranean and the Case of Libya (2016)

The Syrian civil war allowed Turkey to set new foreign policy objectives. The main objective in relation to the EU that Erdogan pursued is becoming more assertive in the Eastern Mediterranean, opposing two EU members: Cyprus and Greece.[5]

Historically, Turkey has had a territorial dispute with Cyprus (formally known as the Republic of Cyprus), which deteriorated its relations with both Cyprus and Greece, who support each other.[34] Moreover, the European Union`s “ambivalent” attitude towards Turkey as if it is a “buffer zone” between Europe and the Middle East as well failing to fulfill the conditions of the refugee deals provoked Turkey to shift into more assertive foreign policy in the region.[163] The Syrian civil war enabled Erdogan to implement it as he was able to utilize Turkey's refugee population to threat and pressure the EU: “Failure to share Turkey’s burden may result in fresh waves of migration towards Europe,” he reminded the EU leaders in March 2021, on the tenth anniversary of the start of the Syrian civil war.[5]

The discovery of gas fields under the Mediterranean Sea exacerbated the tensions further.[34] Erdogan, once again, empowered by his importance for the EU when it comes to refugee management, boldly shifted into “neo-ottoman” policy, confronting his “Western allies as well as regional actors” in order to pursue its share of energy resources.[34] Turkey intensified its activity in the region through acts such as “deploying expeditions into Greece’s and Cyprus’ waters” and “blocking Cyprus’ vessels.”[34] It significantly deteriorated Turkey's relations with the EU as the organization sided with its member states.[34] For instance, France demanded more comprehensive sanctions against Turkey and sent navy as well as participated in military exercises in the region together with Greece and Cyprus.[34] Altogether, the Syrian civil war allowed Erdogan to become more assertive towards EU member states and advance Turkey's interests in the Eastern Mediterranean due to Turkey's important role in preventing migration from the Middle East to Europe, despite the EU`s vocal support of its member states.

6. Conclusion

The mass displacement of Syrians to Turkey due to the Syrian civil war during Erdogan's presidency impacted Turkey-EU relations in a number of ways. Triggered by Bashar al-Assad's crackdown on peaceful protests, the conflict displaced millions of Syrians, and many reached the EU through Turkey, which resulted in the 2015 European migrant crisis. Prior to 2015, relations between Turkey and the EU had been complex due to Turkey`s democratic backsliding. However, the crisis prompted them to cooperate more closely in order to mitigate it. The two actors signed bilateral refugee deals to facilitate cooperation, which granted Turkey funding to care for its refugee population, though there also were drawbacks of these deals for Turkey due to the EU`s failure to grant Turkey some of the privileges it was set to receive.

Moreover, since the 2015 crisis, Turkey, as an emerging middle power, had utilized its refugee population and the EU`s dependency on it when it came to migration management to its advantage as a leverage over the EU. It has been able to implement a significantly more assertive foreign policy in the Eastern Mediterranean to advance its interests against Cyprus and Greece.

[163] Ivanovic, *Turkey and the Eastern Mediterranean: A Chance for Cooperation or A Warning of Conflict?* (2022)

At large, Turkey was able to challenge the status quo of the power distribution between them and the EU through effectively weaponizing refugees.

The shifts in Turkey`s foreign policy can be explained with the fact that as a middle power, it benefits from multipolarity because it allows freedom to maneuver between the great powers, where Erdogan's lack of fixed principles benefits Turkey.[20] Therefore, it is useful for Turkey to challenge the EU`s dominance in Europe and the Eastern Mediterranean to keep the regions where it has the most influence multipolar. Overall, the mass displacement of Syrians to Turkey due to the Syrian civil war during Erdogan's presidency gave Turkey leverage over the EU, though it was limited. On one hand, Turkey was able to advance its interests in the Eastern Mediterranean. On the other hand, it has not gained enough influence to force the EU to fulfill all conditions that the Turkey-EU refugee deals of 2015 and 2016 included. Turkey-EU relations remain complex: Turkey has been able to establish itself as a more influential and assertive regional actor and to strengthen its ties with the EU, though it did not become more likely to receive EU membership. Finally, several topics that have been discussed such as the causes of the Syrian civil war and aspects of Turkey-EU relations are multi-faceted and could not be captured in their entirety for the sake of maintaining focus on answering the research question that was posed, which may be reductionist. Further research may yet challenge the arguments presented within this paper.

Works Cited

10 Years On, Turkey Continues Its Support for an Ever-Growing Number of Syrian Refugees. (2021, June 22). World Bank. Retrieved December 8, 2023, from https://www.worldbank.org/en/news/feature/2021/06/22/10-years-on-turkey-continues-its-support-for-an-ever-growing-number-of-syrian-refugees

A Balancing Act | Strategic Monitor 2018-2019. (n.d.). Retrieved December 8, 2023, from https://www.clingendael.org/pub/2018/strategic-monitor-2018-2019/a-balancing-act/

Abboud, S. (2019, October 1). *Economic Liberalization and Social Transformations in Pre-War Syria*. Crisis Magazine. Retrieved December 8, 2023, from https://crisismag.net/2019/10/01/economic-liberalization-and-social-transformations-in-pre-war-syria/

About the Mission. (2023, May 24). U.S. Mission to The European Union. Retrieved December 8, 2023, from https://useu.usmission.gov/about-the-mission/

Asylum Seeker. (n.d.). Migration and Home Affairs. Retrieved December 8, 2023, from https://home-affairs.ec.europa.eu/networks/european-migration-network-emn/emn-asylum-and-migration-glossary/glossary/asylum-seeker_en

Baghdady, G. (2020, December 28). *Turkey's Electoral Authoritarianism — SIR Journal*. SIR Journal of International Relations. Retrieved December 8, 2023, from https://www.sirjournal.org/research/2020/12/28/turkeys-electoral-authoritarianism

Elakawi, S. (2014, November 22). *The Geostrategic Consequences of the Arab Spring*. openDemocracy. Retrieved December 8, 2023, from https://www.opendemocracy.net/en/north-africa-west-asia/geostrategic-consequences-of-arab-spring/

Enlargement of the European Union. (2022, November 22). Republic of Türkiye Ministry of Foreign Affairs. Retrieved December 8, 2023, from https://ab.gov.tr/enlargement_109_en.html

European Union. (2021, December 28). The World Factbook. Retrieved December 8, 2023, from https://www.cia.gov/the-world-factbook/about/archives/2021/countries/european-union/

Europe's Refugee Crisis Explained. (2015, November 12). World Economic Forum. Retrieved December 8, 2023, from https://www.weforum.org/agenda/2015/11/europes-refugee-crisis-explained/

EU-Turkey Cooperation: A €3 billion Refugee Facility for Turkey. (2015, November 24). European Commission. Retrieved December 8, 2023, from https://ec.europa.eu/commission/presscorner/detail/en/IP_15_6162

EU-Turkey Joint Action Plan. (2015, October 15). European Commission. Retrieved December 8, 2023, from https://ec.europa.eu/commission/presscorner/detail/en/MEMO_15_5860

EU-Turkey Statement. (2016, March 18). The European Council. Retrieved December 8, 2023, from https://www.consilium.europa.eu/en/press/press-releases/2016/03/18/eu-turkey-statemen

Hatton, T. J. (2020). European Asylum Policy Before and After the Migration Crisis. *IZA World of Labor*. https://doi.org/10.15185/izawol.480

Higashijima, M., & Woo, J. (2020). *Political Regimes and Refugee Entries: Motivations behind Refugees and Host Governments*.

Ivanovic, F. (2022, April 16). *Turkey and the Eastern Mediterranean: A Chance for Cooperation or A Warning of Conflict?* E-International Relations. Retrieved December 8, 2023, from https://www.e-ir.info/2022/04/16/turkey-and-the-eastern-mediterranean-a-chance-for-cooperation-or-a-warning-of-conflict/#google_vignette

Jordaan, E. (2003). The Concept of a Middle Power in International Relations: Distinguishing between Emerging and Traditional Middle Powers. *Politikon: South African Journal of Political Studies*, *30*(1), 165–181. https://doi.org/10.1080/0258934032000147282

Kirbyshire, A., Wilkinson, E., Le Masson, V., & Batra, P. (2017). *Mass Displacement and the Challenge for Urban Resilience*.

Laub, Z. (2023, February 14). Syria's Civil War. *Council on Foreign Relations*. Retrieved December 8, 2023, from https://www.cfr.org/article/syrias-civil-war

Linz, J. (1990). The Perils of Presidentialism. In *Journal of Democracy*.

Migration Policy Debates. (2015, September). The Organisation for Economic Co-operation and Development.

Nye, S., Jr., & Goldsmith, L. (2011). The Future of Power. *Bulletin of the American Academy of Arts and Sciences*, *64*(3), 45–52. https://www.jstor.org/stable/41149419?seq=1

Petrini, B., Fischer, M., & Hokayem, E. (2021, December 14). *The Civil War in Syria: An Intractable Conflict with Geopolitical Implications*. The International Institute for Strategic Studies. Retrieved December 8, 2023, from https://www.iiss.org/online-analysis/online-analysis/2021/12/the-civil-war-in-syria-an-intractable-conflict-with-geopolitical-implications

Polk, W. R. (2013, December 10). Understanding Syria: from Pre-Civil war to Post-Assad. *The Atlantic*. Retrieved December 8, 2023, from https://www.theatlantic.com/international/archive/2013/12/understanding-syria-from-pre-civil-war-to-post-assad/281989/

QadirMushtaq, A., & Afzal, M. (n.d.). *Arab Spring: Its Causes And Consequences*.

Quamar, M. M. (2017, May 22). *The Turkish Referendum and Its Impact on Turkey's Foreign Policy*. E-International Relations. Retrieved December 8, 2023, from https://www.e-ir.info/2017/05/22/theturkish-referendum-and-its-impact-on-turkeys-foreign-policy/#google_vignette

Seeberg, P. (2016). The EU-Turkey March 2016 Agreement As a Model: New Refugee Regimes and Practices in the Arab Mediterranean and the Case of Libya. In *Global Turkey in Europe*. Global Turkey in Europe.

Siccardi, F. (2021, September 14). *How Syria changed Turkey's foreign policy*. Carnegie Europe. Retrieved December 8, 2023, from https://carnegieeurope.eu/2021/09/14/how-syria-changed-turkey-s-foreign-policy-pub-85301

The 2015 European Refugee Crisis. (n.d.). The University of British Columbia. Retrieved December 8, 2023, from https://cases.open.ubc.ca/the-2015-european-refugee-crisis/

The Damascus Spring. (n.d.). Carnegie Middle East Center. Retrieved December 8, 2023, from https://carnegie-mec.org/diwan/48516?lang=en

Turkey's Refugee Crisis: The Politics of Permanence. (2016, November 30). International Crisis Group. Retrieved December 8, 2023, from https://www.crisisgroup.org/europe-central-asia/western-europemediterranean/turkey/turkey-s-refugee-crisis-politics-permanence

Weeks, J. L. P. (2014). *Dictators at War and Peace*. Retrieved December 8, 2023, from https://muse.jhu.edu/book/57681

What is a Refugee? Definition and Meaning | USA for UNHCR. (n.d.). Retrieved December 8, 2023, from https://www.unrefugees.org/refugee-facts/what-is-a-refugee/

Yavuz, M. (2009). *Secularism and Muslim Democracy in Turkey*. Cambridge University Press.

How does music affect cognitive functioning? By Jacqueline Ye

Abstract

Music is a widespread form of communication that has been around for over 35,000 years and is prominent in various cultures and identities. Even today, people often listen to music at various times throughout the day. The objective of this paper is to explore whether music affects a person's cognitive functioning, such as enhanced focus, memory retention, and emotion regulation, via a literature review. Understanding the role of music in our brains is important due to its prevalence in our lives, from educational environments to social settings or our psyche. The ability to utilize the potential of music can be valuable to achieving further success.

Section Summaries

Music and attention	People have limited attention available for use. Attempting to allocate that attention to multiple cognitive activities lowers the quality of performance for each activity. Individuals should be mindful of what cognitive activities they are participating in and how much attention each activity requires. For example, songs with lyrics will demand more attention than songs without lyrics. Research suggests that introverts should minimize auditory distractions, and extroverts should work in noisy conditions to *improve* learning performance.
Music and memory	**Active recall/encoding specificity principle:** memory recall improves if the contextual conditions when memory is made and retrieved are the same.
Music and mood	Music can enhance or change an individual's current mood. Music therapy, for instance, can help patients manage symptoms of mood disorders, lower risk of depression, and reduce stress. The two types of music therapy are: ● Active: patients engage directly in musical activities (e.g. singing, dancing, playing instruments) ● Passive: patients don't engage directly in musical activities (e.g. mindfully listening to a piece of music)
Suggested applications	Implementation of whole-brain learning in traditional school systems. Matching auditory condition (quiet and noisy) with introversion/extroversion. Pair vocabulary words with a melody to improve

	memorization
Future directions	How would music listening impact the attention quality and length of those with attention deficit disorders? Further study into introversion and extroversion in relation to music listening and performance. Strengthening the hippocampus Neurologically observing the responses of certain areas of the brain when different genres of music are being played ● Furthering suggested applications based on the final results

Definitions

Music:
A strategic arrangement of sound and rhythm to create melody and harmony.
Mozart Effect:
A temporary increase in performance in individuals after listening to music composed by Mozart.
Spatial-temporal reasoning:
A person's ability to visualize 3D objects and mentally manipulate them.
Context-dependent memory:
A psychological learning phenomenon where memory recall is improved when the environments during memory encoding and recall are repeated.
Mood:
A longer-lasting feeling or state of mind that may be underlying without the individual knowing what prompted the state.
Emotion:
A subjective response resulting from an individual's environment.
Nonsense syllable:
An arbitrarily formed syllable, which doesn't have any prior meaning, that psychologists use to study memory and information retention.
Semantic memory:
The memory of meanings, understandings, and concept-based knowledge that are not attached to specific events.

Introduction

Music is ubiquitous across cultures and time. As a result, understanding what role music may play in cognitive function is certainly warranted. With the cognitive effects of music ranging from providing pleasure to enhancing productivity, there is debate and curiosity surrounding whether these effects of music are fact or myth. There are many possible interpretations and applications from existing literature on this topic, from implementing suggestions within academic settings to influencing emotions inside and outside of therapeutic contexts. Music has the potential to be a valuable tool for enhancing cognitive function, specifically attention and focus, mood regulation, and memory retention. This paper will examine the effects of music in terms of three specific cognitive functions: enhanced focus, memory retention, and emotion regulation. Selecting these three topics narrows the field of study, allowing for a more in-depth understanding. With the power of music, the results have the ability to be applied all around the world at individual and group levels, inside and outside of schools, and within mental health and other medical contexts.

I predict that the literature will support the following hypotheses:

1) The effect of music on an individual's attention depends on the cognitive task the individual is doing and the genre of music being played.
2) Music improves memory when it is being used for context-dependent memory, and there are ways to optimize music to improve memory.
3) The effect of music on mood depends on the genre of music an individual plays, and also why they listen to music.

Impacts of music on attention

Most studies that are focused on exploring the connection between music and attention—specifically the distribution of attention—refer to Kahneman's Limited Capacity Theory (Lang, 2000). The theory states that a person has a limited amount (capacity) of attention, or information-processing resources. Attention can be allocated and invested into cognitive tasks, which compete for the same information-processing resources because of the limited supply. The subconscious decision allotting how much attention to invest, and where to invest, is dependent on the two main theories of the Limited Capacity Theory: (1) the amount of attention poured into an activity is dependent on how much arousal level an activity has. More arousal energy means more attention invested into the activity; (2) the difficulty and energy demand of the activity influence the amount of attention a task requires. According to the Limited Capacity Theory, there are two types of interference that can affect performance: **capacity interference** and **structural interference** (Chou, P. T. M., 2010).

Capacity interference occurs when two cognitive activities done simultaneously are competing for the same information-processing resources, but the total amount of attention available is not enough to meet the demands of both activities. If the demand for attention exceeds the capacity, performance quality worsens. This phenomenon is frequently looked into when studying multitasking because performing tasks simultaneously can lead to errors, slower response rates, and reduced efficiency. Specifically, when two simultaneous cognitive tasks require the same amount of information-processing resources, and there are not enough resources available, neither task can be completed unless one task receives more. This distribution of the information-processing resources can be described using structural interference.

Structural interference is a phenomenon that occurs when the attention capacity is exceeded and attention needs to be distributed between the two simultaneous tasks. Mental activities require different levels of attention. When structural interference occurs and the amount of attention required is more than what can be provided, performance for both tasks worsens. In other words, it is important to keep a balance of attention. While attempting a cognitive task that requires more attention, such as learning a brand new topic, having no background music would contribute to effective learning. On the contrary, cognitive tasks with lower attention requirements, such as doing simple, repetitive math problems, can be done with some background music.

A survey (Ballard, 2003) was done in the U.S. with students from two midwestern states who self-reported their media habits. The results showed the participants believed all kinds of media, including cell phones, television, CD players, computers, and iPods, had a negative effect on academic performance, considering it as a "source of major distraction." As technology evolved, personal laptops have been more commonly implemented into classrooms as assistance, but the study questioned whether they could become a hindrance. Computers amplify the study habits of the students because, according to Ballard (2003), multitasking is more likely to occur if the students have no interest in what they are learning. If the student wants to learn, the computer is an amazing tool to further their learning. On the other hand, if the student wants to stray away from the class, the computer gives them access to the entire internet, which they may use however they'd like. A possible way to prevent this level of distraction in students is by limiting how much of the internet they can have access to. However, this is not a foolproof method; since this method has been so widely used—school and work are good examples—there exists websites designed to be unblockable, such as games encoded into a Google Site. Restrictions are useful in limiting the number of distractions across the internet, but it does not have the ability to completely get rid of them. Another common source of media, used in a similar way as the computer, is television (TV).

TV, often used by students while studying, is considered to be a more potent source of distraction than radio or silence as it stimulates more arousal energy. The presence of background TV pushes the actual task at hand back to be the secondary task, as the student aims to focus more on understanding the TV. However, studies (Armstrong et al., 1991; Cool and Yarbrough, 1994; Pool et al., 2010) show both TV and radio affect the performance of mathematical and reading tasks, as well as the time it takes for the students to complete them. Background TV hindered students' abilities to recall information from difficult passages. Surprisingly, the results also showed that the time spent wasn't a sign of lowered performance quality; the difference in the amount of time spent on the task when background TV was on, versus when it wasn't, was exactly equal to the amount of time the participant was looking at the computer (Pool et al., 2010). While performance did decrease, it would have only been due to switching modes from focusing on the TV and on the task. On the contrary, if students are only listening to music—which only consists of an audio component—they will be *less* distracted since their eyes will be focused on the task instead of looking around, trying to understand the visual components of types of media such as TV. Both forms of distraction hindered the individuals' performances, but TV served as a more significant distraction than radio did. Therefore, it is important to be aware of what sort of media to indulge in when carrying out tasks dependent on cognitive ability, especially those needing information processing resources and attention.

Another approach to observing the effects of music on attention is in relation to the level of extroversion or introversion of the individual(s). To explore this, Furnham and Strbac (2010) and Belojevic et al. (2001) conducted studies to observe the differences in impact that background-music listening has on introverts versus extroverts, with introversion measured using the Eysenck Personality Questionnaire (EPQ). Both studies show that compared to introverts, extroverts tend to produce *better* results when in a noisier environment and are able to work significantly faster under noisy conditions. Specifically, the introverted participants in the study by Belojevic et al. (2001) reported their levels of fatigue and concentration problems were more prominent when in a noisy environment compared to when in a quiet environment. For both personality types, the accuracy of mental processes was not impacted by noise. Furnham and Strbac (2010) provided results similar to those of Belojevic. The methodology tested reading comprehension, prose recall, and mental arithmetic. Their data showed a trend favoring extroverts' performance in noisy conditions, but performance of both personality types were similar under quiet conditions. However, more studies should be done to corroborate these findings. Additionally, there is a flaw that lies within the categorization of introversion and extroversion. Personality tests—such as the Myers-Briggs Type Indicator (MBTI) and the EPQ—are not 100% accurate (Randall et al., 2017; Caruso and Edwards, 2001). Therefore, there is no guarantee that an introvert is being tested as an introvert, or an extrovert as an extrovert. Overall, this study provides conclusions on music's impact on attention from a different approach than the previously discussed studies. As opposed to focusing on different types of auditory distraction, this experiment groups individuals based on their own personality types. This pattern provides further insight into how music can be utilized for further improvement in cognitive function, as well as opening the door to future study into introversion vs extroversion, potentially spreading it to other fields of cognitive psychology.

Impacts of music on memory

A study (Smith, 1985) was conducted with the goal of testing active recall. The experiment explored whether the presence of background music would affect the ability of participants to recall a given set of words. The results of the study supported the notion of context-dependent memory, which is a theory stating recall of memory is improved when the environment during memory encoding and recall are the same. The experiment's methodology went as follows: 54 volunteer participants took part in two experimental sessions. During the first session, the participants saw a list of words. The words were each printed on an index card, which the experimenter switched out every five seconds. Immediately after seeing the cards, the participants participated in a five-minute active recall "test." After two days, the participants returned to the same space for a final five-minute active recall test. What differentiated the first session and the final test was the acoustics in the background. During the first session, one-third of the participants heard a Mozart piano concerto (M) in the background, another third heard jazz

(J), and the last third had no background acoustics (Q). In the second session, the three groups (M, J, Q) were again split into thirds, each of the subgroups listening to one of the options (M, J, Q) for the final test. The combinations were categorized as SC groups (MM, JJ, QQ) if the same acoustics were playing in the background for both sessions or DC groups (MJ, MQ, JM, JQ, QM, QJ) if different acoustics were playing for the two sessions. The results showed recall was improved when music was added during session one and was not replaced or taken away during session two. However, the recall abilities were not enhanced when quiet conditions were kept across both sessions. Adding or removing music did not prove to have a negative effect on recall. The findings of this study showed an applicable way of optimizing active recall. Should the environments in which an individual is taking in information and attempting to actively recall that information be the same—specifically acoustically speaking in relation to this study—the ability to recall the information drastically improves. Active recall is also commonly named the encoding specificity principle, which can be viewed as an intersection between context-dependent memory and active recall. The encoding specificity principle states that memory recall is enhanced when the contextual factors of when memory is encoded and when it is being recalled are the same (Tulving and Thomson, 1973). This principle is seen clearly in Smith's experiment. When there is an auditory stimulus that is repeated when information is being encoded and when it is being recalled, the memory recall ability improves.

Another study (Zhang, 2020) was conducted to test the influence of music listening on academic performance in terms of time taken and errors made. The eight selected students from a Chinese international high school were divided into four categories in terms of academic performance in school: Good, Fair, Limited, and Weak. There were two rounds of color memorization tests: one before listening to music and one after. Results show that after listening to music, the students in the Good category had a significant decrease in errors made. A positive correlation was found between academic performance and response time, indicating that students with higher academic performance tended to have shorter response times. Similarly, the same was also true with the number of errors made: the better the academic performance, the fewer errors made. However, the results failed to provide any convincing evidence that listening to music could improve memory retention and decrease the number of errors made as the data patterns did not hold true for each category of students. This study's methodology found excelling students can perform better after listening to music, but there is no overarching conclusion to be made.

Looking at applications of optimizing music to enhance learning, Bulgarian scientist and neurologist Georgi Lozanov, the "father of accelerated learning," was a leading figure in the study of the effects of music on memory and learning. He found that music has the ability to induce a state of relaxed alertness, or "psycho relaxation." During this state of "psycho relaxation," the brain shows a large increase in alpha brain waves. Alpha brain waves (frequencies of 8-13 Hz) are one of the five electrical waves produced by the brain. When they dominate brain activity, it is usually a signal of relaxed alertness. The benefits of increasing the amount of alpha brain waves include: improved levels and maintenance of focus, heightened memory, reduction of stress and anxiety, and elevated creativity levels; this significant increase due to psycho relaxation, according to Ushi Felix (1993) should result in higher achievement. To achieve this state of psycho relaxation, the individual should take part in "whole-brain learning." To introduce this method into classrooms, Lozanov later created a certain teaching system: "Suggestopedia." Whole brain learning is using both the left and right hemispheres of the brain simultaneously to carry out an action. The left hemisphere is responsible for language processing, logic and reasoning, communication, and memory. On the other hand, the right hemisphere oversees music processing, creativity, and artistic skills. As such, reformatting education to create activities merging music processing with reasoning and memory will likely result in heightened performance. Overall, this holistic approach has been shown to optimize an individual's learning. Suggestopedia can be a very useful tool to be implemented, and future research and testing in this method is likely to provide more insight into its abilities and limitations.

Additionally, the German evolutionary biologist Richard Semon first introduced the "engram theory," describing memory storage from a neurological lens. He stated that memory leaves physical changes in the brain. When a memory is being recalled, these neurons storing the memory are reactivated. Semon's theory has since inspired future scientists and modern information on memory. Susumu Tonegawa's lab at MIT was the first to prove the credibility of the engram theory. To cure his seizures, 27-year-old Henry Molaison's hippocampi were removed. As a result, he was unable to make new memories; because of this, the key purpose of the hippocampus was uncovered (Zhang, 2020). Without the hippocampus, new episodic memories cannot be formed. As such, strengthening the hippocampus has the potential to enhance an individual's memory capacity. Due to the hippocampus's role in memory, it is a region of interest in memory studies involving music (for review., see Toader et al., (2023)). From a neuroscientific point of view, Hotz (1998) and Molnar-Szakacs and Overy (2006) reviewed evidence that music stimulates specific regions of the brain responsible for memory, such as the amygdala and medial prefrontal cortex, which work with the hippocampus, lateral prefrontal cortex, and parietal cortex (Buchanan, 2007). The activation of these centers leads to the retrieval of emotions corresponding to pieces of memory. Similarly, the amygdala, one sector noted above to be stimulated by music, is responsible for encoding the emotional associations of information as it is transferred from short-term storage to long-term memory. The amygdala, along with the hippocampus, prefrontal cortex, and parietal cortex, is located in the cerebrum, which makes up 80% of the brain (Ackerman, 1992) and is responsible for memory, emotional response, and learning. The hippocampus, situated deep within the brain's central region, is crucial for learning and memory formation (Eichenbaum, 1999). Its primary job is storing short-term memories and transferring them to long-term storage. The emotional connection to certain pieces of memory has the ability to enhance memory storage through the release of hormones that stimulate regions of the brain (Fogarty, 1997). Listening to music, along with other forms of music and auditory stimulation, has been shown to strengthen brain synapses and neural pathways, specifically by inducing LTP (Long-Term Synaptic Potentiation/Long-Term

Potentiation) in the brain (Chatterjee et al., 2021). LTP, when prompted, increases signal transmission between the neurons. When this is triggered in the hippocampus, which is responsible for forming and retrieving memory, its ability to carry out its memory-related duties is heightened. De Deus et al. (2017) conducted a study on how short-term loud sounds impact on synapse plasticity and function in the hippocampus of rats. The experiment found that daily exposure to sound stimulation of 60 db led to enhanced learning performance and the increase of BDNF (neurotransmitter modulator important for memory and learning and plays a key role in LTP). This study provides a habit that is easily incorporated. Adding daily exposure to certain sounds to benefit cognitive function can be a less attention-demanding replacement for background music.

Impacts of music on mood

While the two terms are often interchangeable, many believe there is a root difference between "mood" and "emotion." While emotions arise from a scenario-specific cause, moods are the underlying feelings an individual experiences and have longer durations (APA). Moods are considered to be tools that "provide information about internal states," while emotions provide hints to the "states of the environment" (Saarikallio, 2007). When discussing emotional reaction, most emotion theorists believe it is composed of three types of experiences: subjective, behavioral, and physiological (Saarikallio, 2007). The subjective experience refers to the emotional and mental experience—the feelings that arise. The behavioral component describes how the individual chooses to express their emotions. Lastly, the physiological element represents the reaction of the rest of the physical body (Saarikallio, 2007). Music has shown the capability to affect all three of these components. Due to its extensive reach, music can be optimized as a tool for individuals to regulate their moods and emotions. However, 'mood regulation' is more often used when discussing music-related regulation. Mood regulation refers to a process that aims to change or maintain the current mood that an individual is in (Saarikallio, 2007). Music-related regulation is more targeted to the subjective experience as opposed to the behavioral and physiological responses. Multiple studies have observed results proving its impact. Thayer et al. (1994) found that in a self-assessment, listening to music was a tool ranked second in effectiveness to change bad moods. The Wells group noted that in their study group, 85% of women and 74% of men use music to change their mood (1990). In addition to being used to change their moods, some participants have also reported using music to express their current state. The Benhe group (1997) found that adolescents chose happy music when happy, aggressive music to vent when angry, and emotional music for comfort. In conclusion, it is possible to use music to enhance certain moods as desired.

Music therapy is a widely practiced form of therapy. It is a treatment method that focuses on optimizing music activities to improve the mental and physical health of individuals. Physiologically, music therapy makes use of external factors to stimulate the sections of the brain responsible for mood and emotions. When used in conjunction with drug therapy, the effects prove to be significant, particularly on patients with senile depression, schizophrenia, and preoperative anxiety (Zhang, 2020). There are two types of music therapy: active and passive. In active music therapy, patients are actively participating in musical activities; this includes singing, dancing, and playing musical instruments. On the other hand, passive music therapy activities can include listening to music with extra care, really feeling the piece internally. With these methods, music therapy has been used to manage depression, specifically post-stroke depression in parts of China (Zhang, 2020). When paired with exercise/physical therapy, music therapy can improve paralysis, with benefits extending even to patients with early stages of a stroke.

Discussion

Overall, the findings of the studies provide important information on the relationship between music and cognitive function. With multiple methods of use, music listening has the ability to enhance and maintain positive moods, and improve active recall. Its effects on quality and duration of attention—as well as memory retention—has shown to be slim to none, given the level of distraction caused by the music does not exceed the individual's capacity of attention. This information can be used by anyone who relies tremendously on the three aspects of cognitive function mentioned in this paper. This includes students as they choose between listening to background music while studying, educators as they debate whether to allow students to listen to music while working, and parents when they see their child working with headphones on.

Some suggested applications include implementing Suggestopedia or other forms of whole-brain learning in school systems and other academic settings (Waluyo, 2018). For example, playing soft background music during lectures or learning vocabulary by repeating words with a melody (the Alphabet song and the Periodic Table of Elements song are good examples of this), stimulating both hemispheres of the brain and making it easier for students to encode a series of words into long-term memory (Anderson, 2000). A possible application best used by those in studying conditions is optimizing learning by matching the amount of auditory distractions to the individual's personality type in terms of introversion and extroversion: more auditory distractions for extroverts and less for introverts. However, there is room left for further study into music and cognitive function. In addition to being a relatively new science, psychology studies complex, and often unpredictable, creatures. What environment works for an individual's learning one day will likely not be the same the next, as human cognition is affected by many other factors, to begin with.

There exists limitations with the research done in this paper. There is no inclusion of studies with results that can generalize individuals with different attention capacities than the average population, for example, those with attention disorders. This would be a valuable field of study as individuals with attention disorders will likely react to surrounding auditory stimuli differently than those without, making this a useful direction for future study. Another interesting addition would be the study of introversion/extroversion and susceptibility to auditory distractions from a neurological perspective: analyzing if the brains of introverts and extroverts have different responses to auditory distractions, and *why* this correlation or causation exists. It's possible that this deeper look into the reason behind this phenomenon can provide better insight and applications. Another direction for neurological research is observing which different parts of the brain are stimulated when an individual listens to different genres of music or silence. This study could be worthwhile, especially if there are differences in behavior that hint at causation. Another path would be further exploration of how music therapy impacts the brain; which parts of the brain does it appeal to in order for the effects to combat depression? Additionally, another beneficial research focus would be studying ways to strengthen the hippocampus. A limitation of this paper is the lack of suggestions for improving attention capacity. A way to apply the findings of this paper would be analyzing the difference in one's memory capacity before and after strengthening the hippocampus through exercise, obtaining more knowledge, and any other form of mental stimulation.

Conclusions

In this study, I reviewed the literature on attention, memory, and mood as they relate to music. Overall, I found support for the hypothesis of:

1. The effect of music on attention is dependent on the cognitive task the individual is doing and the genre of music that is being played.
2. Music will improve memory when it is being used for context-dependent memory, and there are ways to optimize music to improve memory.
3. The effect of music on mood will depend on the genre of music an individual plays, and also what their goal for listening to music is.

In conclusion, music has a large impact on cognitive function. The majority of existing studies analyzing the connection/causation between music and cognitive function are focused on how music affects *attention*, relating well to other areas of studies on multitasking. Participating in multiple cognitive tasks simultaneously—specifically music listening and another cognitive task—can be done, as long as the amount of attention is balanced. An individual can listen to music in the background if they are not actively attempting to listen and if the other cognitive task does not demand more attention than they have available. Introverts study better in quiet conditions, and the opposite holds true for extroverts. Repeated musical or auditory conditions during memory encoding and recall can improve (context-dependent) memory recall. Suggestopedia can be implemented by individuals and in academic environments alike to improve memory retention. Daily exposure to sounds of 60db can improve LTP and, in turn, enhance memory and learning. Music can also be used as a tool to enhance, release, or change an individual's mood. It is also widely-used for music therapy to treat patients with mood disorders. These findings are applicable to anybody, with applications extending from studying to improving memory to being used as a coping method.

Future studies in this field can focus on neuroscientific approaches to understand why the brain reacts the way it does to music, as well as expanding the ranges of experiment subjects, such as researching people with different attention and memory capacities and different mental health conditions.

Works Cited

Ackerman S. (1992). Discovering the Brain. Washington (DC): National Academies Press (US).

Anderson, S., Henke, J., McLaughlin, M., Ripp, M., & Tuffs, P. (2000). Using Background Music To Enhance Memory and Improve Learning.

Bandral, N., & Kaur, R. (2018). A Psychological Inquiry into the Role of Music in Video Games. *Language in India*, 18(5).

Belojevic, G., Slepcevic, V., & Jakovljevic, B. (2001). Mental performance in noise: The role of introversion. *Journal of environmental Psychology*, 21(2), 209-213.

Buchanan, T. W. (2007). Retrieval of emotional memories. *Psychological bulletin*, 133(5), 761.

Caruso, John C., and Shana Edwards. (2001) "Reliability generalization of the Junior Eysenck Personality Questionnaire." *Personality and Individual Differences* 31.2: 173-184.

Chatterjee, D., Hegde, S., & Thaut, M. (2021). Neural plasticity: The substratum of music-based interventions in neurorehabilitation. *NeuroRehabilitation*, 48(2), 155-166.

Chou, P. T. M. (2010). Attention Drainage Effect: How Background Music Effects Concentration in Taiwanese College Students. *Journal of the Scholarship of Teaching and Learning, 10*(1), 36-46.

De Deus, J. L., Cunha, A. O. S., Terzian, A. L., Resstel, L. B., Elias, L. L. K., Antunes-Rodrigues, J., ... & Leão, R. M. (2017). A single episode of high intensity sound inhibits long-term potentiation in the hippocampus of rats. *Scientific reports*, 7(1), 14094.

Eichenbaum, Howard, et al. "The hippocampus, memory, and place cells: is it spatial memory or a memory space?." *Neuron* 23.2 (1999): 209-226.

Furnham, A., & Strbac, L. (2002). Music is as distracting as noise: The differential distraction of background music and noise on the cognitive test performance of introverts and extraverts. *Ergonomics*, 45(3), 203-217.

Holmes, S. (2017). *The impact of participation in music on learning mathematics* (Doctoral dissertation, UCL (University College London).

Hotz, R. L. (1998). Music stimulates brain, study finds. The Chicago Sun-Times, p. 34 S.

Lang, A. (2000). The information processing of mediated messages: A framework for communication research. *Journal of Communication*, 52.

Lu, Y., Christian, K., & Lu, B. (2008). BDNF: a key regulator for protein synthesis-dependent LTP and long-term memory?. *Neurobiology of learning and memory*, 89(3), 312-323.

Molnar-Szakacs, I., & Overy, K. (2006). Music and mirror neurons: from motion to 'e'motion. *Social cognitive and affective neuroscience*, 1(3), 235-241.

Musliu, A., Berisha, B., Musaj, A., Latifi, D., & Peci, D. (2017). The impact of music in memory. *European Journal of Social Science Education and Research*, 4(4), 138-143.

Pool, M. M., Koolstra, C. M., & Van der Voort, T. H. (2003). Distraction effects of background soap operas on homework performance: An experimental study enriched with observational data. *Educational Psychology*, 23(4), 361-380.

Purves D, Augustine GJ, Fitzpatrick D, et al. (2001). Neuroscience. 2nd edition. Sunderland (MA): Sinauer Associates. Long-Term Synaptic Potentiation.

Randall, K., Isaacson, M., & Ciro, C. (2017). Validity and reliability of the Myers-Briggs Personality Type Indicator: A systematic review and meta-analysis. *Journal of Best Practices in Health Professions Diversity*, 10(1), 1-27.

Saarikallio, S. (2007). *Music as mood regulation in adolescence* (No. 67). University of Jyväskylä.

Schuster, D.H. (1985). The effect of background music on learning words. *Journal of the Society for Acccelerative Learning and Teaching.* 10 (1), 21-39.

Scripp, L. (2002). An overview of research on music and learning. *Critical links: Learning in the arts and student academic and social development*, 132-136.

Shih, Y. N., Huang, R. H., & Chiang, H. Y. (2012). Background music: Effects on attention performance. *Work*, 42(4), 573-578.

Smith, S. M. (1985). Background music and context-dependent memory. *The American Journal of Psychology*, 591-603.

Thomson, C. J., Reece, J. E., & Di Benedetto, M. (2014). The relationship between music-related mood regulation and psychopathology in young people. *Musicae Scientiae*, 18(2), 150-165.

Toader, C., Tataru, C. P., Florian, I. A., Covache-Busuioc, R. A., Bratu, B. G., Glavan, L. A., ... & Ciurea, A. V. (2023). Cognitive Crescendo: How Music Shapes the Brain's Structure and Function. *Brain Sciences*, 13(10), 1390.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review*, 80(5), 352.

van den Tol, A. J., Coulthard, H., & Hanser, W. E. (2020). Music listening as a potential aid in reducing emotional eating: An exploratory study. *Musicae Scientiae*, 24(1), 78-95.

Waluyo, H. J., Suudi, A., & Wardani, N. E. (2018). Suggestopedia Based Storytelling Teaching Model for Primary Students in Salatiga. *Malaysian Online Journal of Educational Technology, 6*(1), 64-75.

Zhang, S. (2020). The positive influence of music on the human brain. *Journal of Behavioral and Brain Science*, *10*(1), 95-104.

Unveiling Systemic Disparities: The Interplay of Housing Policies and Educational Inequities in America By Shreeya Ram

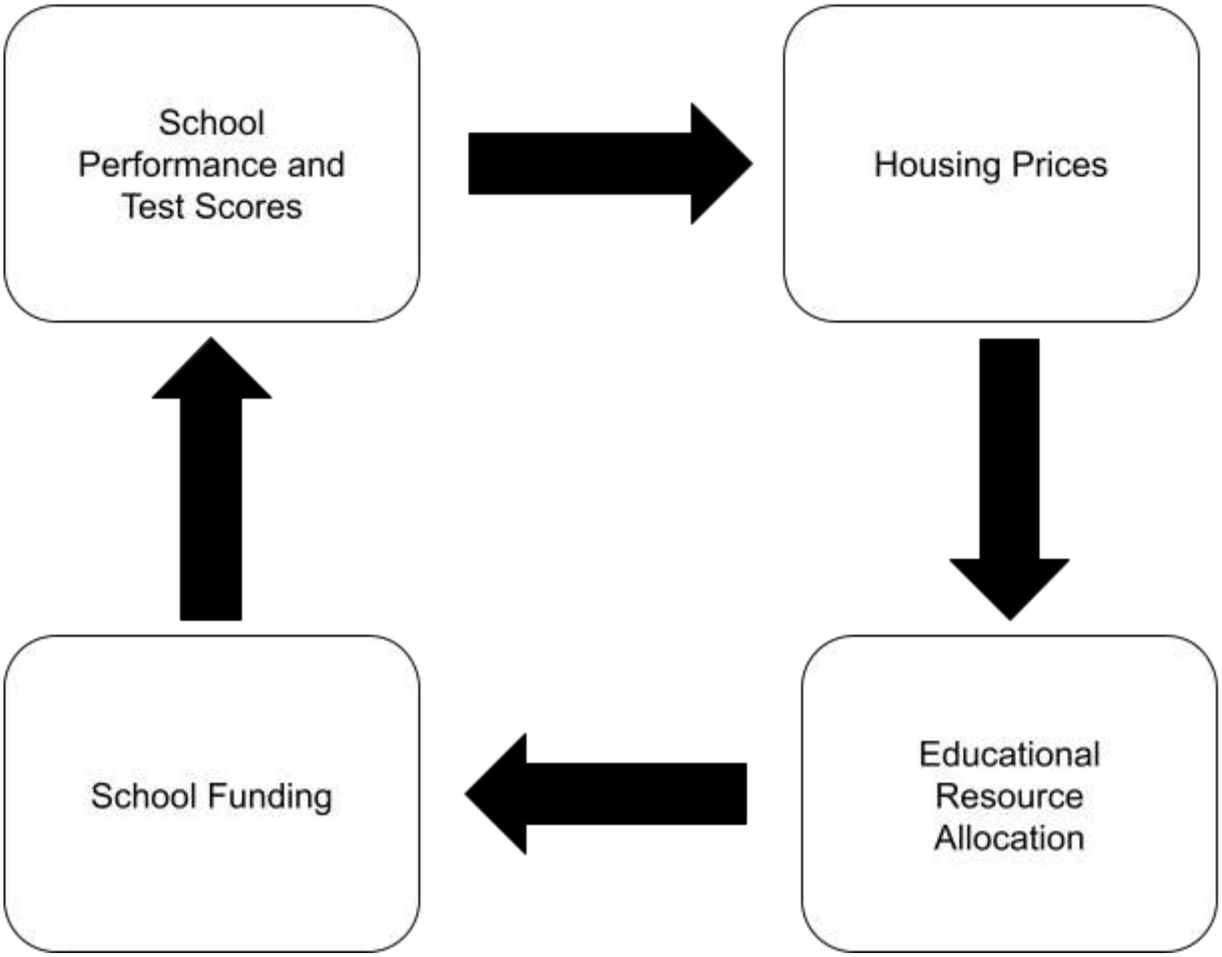
Minority communities in America face many disparities in housing and education. Discriminatory housing and education policies perpetuate this. Historically, the Federal Housing Agency's (FHA) role in redlining perpetuated systemic disparities in housing opportunities and community development. Examining the impact of housing policies on urban communities reveals racially restricted rental housing practices and the displacement of marginalized communities. The link between federal housing policies and educational disparities in marginalized communities is deeply rooted in historical dynamics, where race and income play crucial roles. Understanding these historical aspects of federal housing policies is crucial in understanding the link between housing policies and educational segregation. These historical dynamics perpetuate current systemic inequalities, which continue to shape housing opportunities and educational outcomes in the present day. Examining the present-day situation reveals persistent challenges in achieving equitable housing and education, particularly for communities burdened by historical disparities.

After the 1933 housing shortage, the government began to segregate housing in the United States, focusing on providing housing for white families while pushing African American families and other minorities to urban housing programs (Gross). The Federal Housing Administration (FHA) also refused to insure mortgages for predominantly black neighborhoods (Gross). The FHA institutionalized discriminatory practices by implementing redlining and creating color-coded maps to allocate housing opportunities based on racial compositions (Gerken et al.). Other exclusionary practices included only allowing single-family houses to be built in better-quality neighborhoods (Drew). Because of the financial hardships faced by African Americans post-slavery, this practice specifically targeted their communities, forcing them to live with relatives under one roof as it was not feasible for a single family to live in a house (Davis).

The legacy of redlining extends beyond housing, influencing school zoning and resource allocation, and impacting long-term educational outcomes for marginalized communities (Burke and Schwalbach). While at first glance it may seem as simple as ‘poor people have to go to poor schools,’ educational opportunities are explicitly tied to redlining because redlining prevents families from buying houses or taking mortgages in wealthier neighborhoods, trapping families in low-income neighborhoods. Public school attendance often depends on where families live and students are usually only allowed to attend schools if they live within those specified school district borders (Meckler and Rabinowitz). Catalyzing a direct relationship between current implications of redlining and educational opportunities for nonwhite children. Schools and districts currently located in areas historically placed in redlining maps to be ‘risky’ in terms of loan lending see less revenue per student, more African American and non-white students, less diversity, and worse test scores (Lukes and Cleveland). Moreover, students residing in marginalized neighborhoods face significant barriers to educational attainment, including limited resources, overcrowded classrooms, and underfunded schools. Throughout the United States, school districts with Black, Latino, and Native American students may receive up to $2,700 less than districts with more white students (“School Districts That Serve Students of Color Receive Significantly Less Funding”).

The disparity in school funding also has a direct impact on housing market dynamics. Houses located within district boundaries of better-funded schools see a 10 to 20% increase in their market value, making it increasingly difficult for low-income families to attend schools that receive more resources and funding (Fischel). In fact, for more than 70% of students in the United States, the school they attend is based on the area their family can afford to live in (U.S. Department of Education). Effects of redlining and other racially discriminatory practices have kept generations of families in a state of poverty and confined them within low-income neighborhoods with schools receiving disproportionately less funding. Reports establish a correlation between school funding and housing prices; one study shows that s for every dollar spent on public schools in a community, home values increased by $20 (Chen). Students from low-income and minority backgrounds are disproportionately affected by housing policies that confine them to under-resourced school districts, perpetuating cycles of disadvantage and hindering their academic success. As long as certain schools receive more funding than other schools, home prices will continue to be affected by resource allocation. Just as school funding affects housing prices, housing prices affect school funding. Revenue from property taxes funds schools; schools in high-income neighborhoods receive more funding since almost 60% of school funding comes from local property taxes (Rancaño). This interconnection of housing and education shows that policies that are aimed to address either one will end up impacting both sectors.

Figure 1. Relationship between housing prices and resource allocation


School Performance and Test Scores
Housing Prices
Educational Resource Allocation
School Funding

Housing policies of the 21st century seemingly attempt to create a landscape of equitable housing; however, often the harms have an equal or even greater effect than the supposed benefits. There are multiple ways that affordable housing and voucher programs are put into place, including tenant-based assistance. Public housing has been used as a presumed solution to inequitable housing for decades; houses built in the mid-1900s are still being used as public housing with little to no repairs (Demsas). Only 17% of public housing has been built after 1997, while 42% has not experienced any renewal construction since 1975 (Demsas). Most housing voucher programs are government-sponsored, and, consequently, the lack of quality is attributed to insufficient funding. Additionally, these subsidized units under the Housing Voucher Program are not properly evaluated during health inspections, and poor living conditions are overlooked because of oversight and inadequate budget allocation.

One such unit that was subsidized by a HUD HAP (Housing Assistance Payments) Housing Voucher contract is the Concordia Place Apartments in Chicago, Illinois (Staff). Tenants reported mold growing everywhere, mice running in the walls, and their garbage cans were taken away disallowing tenants to properly dispose of their trash ("U.S. Lawmakers Demand HUD 'Swiftly Examine' Mold, Rodents, Other Issues at Concordia Place Apartments"). However, when the Department of Urban Housing and Development conducted their health inspection for the apartments they assigned it a grade of 94, which would insinuate that the apartments were in a pristine or at least livable condition, given that the passing grade is merely 60 (Hendrix). Inaccurate grading seems to happen quite often as only 10% of public housing that was recently reviewed by the Department of Housing and Urban Development received failing grades (Hendrix). In contrast, as of 2019, 42% of public housing properties had finished their last construction work well before 1975 (Demsas). The houses that were built during that time frame tend to be more prone to poor conditions; however, this has been repeatedly overlooked. Despite the tendency to give passing grades to houses that exhibit poor living conditions including toxic mold and lead, HUD stated that 62,000 public housing units require lead abatement, a process that essentially rids houses of toxic lead (Hendrix). Lead exposure causes learning and behavior problems, brain damage, and slow development, which can lead to lower IQ and lower academic achievement (Centers for Disease Control and Prevention). The Public Housing Agency agrees in the contract to provide financial assistance for tenants by making housing payments to the owner on behalf of the tenants. Increased exposure to lead further increases educational disparities faced by children of marginalized communities. Despite ample evidence that public housing often lacks safety for low-income families, the prevalence of these problems highlights the ongoing struggles that certain communities face when it comes to housing.

After Concordia Apartments received a passing grade, CBS News made several reports exposing the harsh reality of the living conditions, and L+M developers finally acquired and remodeled the apartment one year after these reports and complaints were made. (Staff). Another similar situation is found in Minneapolis, MN, where a public housing building that was on fire ended with the demise of five people, seemingly due to the lack of sprinklers (Demsas). The Department of Housing and Urban Development released a statement saying that moving forward, they plan to ensure the implementation of sprinklers in every building; however, it would take a decade to complete even with the necessary funding, which has not been allocated (Demsas). Many would argue that the current setup for public housing fails to provide equitable housing. First, harsh living conditions are often overlooked, then once they finally do come to light, it takes even more time for developers to consider and approve remodeling and renovations, if ever; all the while, low-income families are put at risk. Dismantling inequitable housing has to include policy reform coupled with adequate management and facilitation.

HUD utilizes a voucher program under the Housing Assistance Payments Contract (HAP Contract) specifically the Housing Voucher Program that assists tenants (Housing Assistance Payments (HAP) Contract, n.d.). In the contract, the Public Housing Agency agrees to provide financial assistance for tenants by making housing payments to the owner on behalf of the tenants (*Housing Choice Voucher Program Guidebook*). The Housing Choice Voucher program provides financial assistance to low-income families on behalf of the federal government ("Housing Choice Voucher (Section 8) | USAGov"). Recipients can apply vouchers to any available townhouses, apartments, or single-family homes for rent. Once given a voucher, the receiving family must find a house themselves, and if the owner agrees to rent out under the voucher, then they can move in (“Housing Choice Voucher Program (Section 8)”). The federal government pays a portion of the rent as outlined through the voucher and then the family is required to pay the difference between the original rent and what was covered by the voucher.

Over 2 million low-income families, consisting of over 5 million low-income people, have utilized the program (Fischer). However, even with Housing Choice Vouchers, there is still discrimination. In a 2018 study conducted by the Urban Institute, their field team combed through over 341,000 rental ads across Fort Worth, Texas, Los Angeles, California, Newark, New Jersey, Philadelphia, Pennsylvania, and Washington DC (Cunningham et al.). They identified 8,735 units that were not only available but also met the testing parameters and local voucher program rent limits (Cunningham et al.). On average, it took them screening 39 ads to find just one potentially suitable unit (Cunningham et al.). The results were eye-opening: landlords turning away voucher holders, with rejection rates varying significantly (Cunningham et al.). In places like Fort Worth, TX, and Los Angeles, CA, denial rates are as high as 78% and 76%, respectively (Cunningham et al.). Philadelphia, is slightly better at 67%, in Newark and Washington, DC, where rejection rates are much lower at 31% and 15%, respectively. The evidence proves that the ability to benefit from housing voucher programs may depend on specific cities or states.

Additional policy and legislative reform is necessary to combat these issues. Laws and regulations at the state and local levels, known as Source of Income (SOI) laws or ordinances, aim to prevent discrimination against renters and home buyers based on where their income originates (“Source of Income Laws”). These regulations typically encompass various income sources, such as federal benefits like Social Security and Temporary Assistance for Needy Families (TANF). Many of these laws consider federal rental assistance, like the Housing Choice Voucher program, as a protected income source (“Source of Income Laws”). Consequently, it is illegal to deny housing to a household solely because they participate in such programs (“Source of Income Laws”). Nineteen states, along with the District of Columbia, as well as numerous local jurisdictions, have enacted such legislation to combat housing discrimination based on the origin of income (“Source of Income Laws”). The specifics of these laws vary considerably. For instance, in Washington, DC, the law specifically mentions "section 8 vouchers" or Housing Choice Vouchers, while others adopt broader language encompassing various income sources (“Source of Income Laws”). Conversely, some laws explicitly exclude housing choice vouchers while safeguarding other forms of income (“Source of Income Laws”). While general anti-discrimination language regarding income sources may offer a framework to prevent discrimination against diverse forms of assistance, such as locally-funded tenant-based aid, it may also lack the clarity necessary to fully address all instances of income source discrimination (“Source of Income Laws”).

Research on the Housing Choice Voucher program indicates that individuals holding housing choice vouchers tend to reside in neighborhoods with schools of similar quality to those accessible to households with comparable incomes but without vouchers (Ellen et al.); these households do not appear to utilize the additional income provided by the voucher to access better educational opportunities, which can be attributed to bias and lack of awareness (Ellen et al.). In the analysis, researchers used HUD data on 1.4 million housing choice voucher recipients across 15 states, with school-level data from 5,841 distinct school districts (Ellen et al.). This effect is more pronounced in urban regions where there is a significant proportion of affordable rental units situated near top-performing schools and in neighborhoods close to these schools (Ellen et al.). Findings indicate that with access to the right information and opportunities, more families utilizing vouchers would choose to relocate to superior schools as their children approach school age (Ellen et al.).

Marginalized communities not only suffer from oversight in safety and housing quality, but public housing programs often further perpetuate the very racial inequities that they attempt to dismantle. Housing programs often allocate permanent housing for low-income families and individuals in an attempt to provide a quick, one-stop solution. (Eide). The process generally involves finding a new, unused area to build new houses or finding cheap neighborhoods that can be subsidized for public housing (Eide). The result is government-created neighborhoods that essentially segregate Black and Latino families into low-income neighborhoods, limiting their educational opportunities (Gross). School districts where the majority of students are not white receive $23 billion less funding than school districts where students are primarily white, even if the districts have the same number of students attending their schools (Guastaferro). In a neighborhood in Manhattan where the population is predominantly whiter and the average income is over $100,000, the public schools have an average math and reading proficiency of 84%, whilst another neighborhood that is predominantly Black and Latino where the average income is just a little over $25,000 have an average math and reading proficiency of 30% and 37% respectively (Guastaferro). This is because when we create communities with a high concentration of poverty and we force those students to attend schools that are not properly funded because of lower taxes it creates a poverty trap. Students who grow up in a poor neighborhood due to racist housing policy, then have to go to a poor school without proper resources, then they do not get a quality education, which has a long-term effect on future earnings, as funding for schools is positively correlated with adult income. A 10% increase in funding for schools would lead to almost an 8% increase in wages and a 9.8% increase in family income (Lafortune). Schools with significantly less funding and lower test scores produce students who receive lower income in their later life bringing them back to these segregated neighborhoods.

It is more apparent than ever that the implementation of policy and reform is necessary to alleviate disparities that affect the way people live and the extent to which children receive education. There have been many solutions over the past decades that have been proposed and implemented. One such initiative is the Building Black Wealth Campaign sponsored by The California Housing Finance Agency (Agency). The campaign aims to close gaps in Black homeownership by increasing access to educational materials, resources for free housing, and downpayment assistance (Agency). The campaign was proposed in 2021 and is relatively new, so the impact and extent of solvency for the program are not clear; however, reports show that Black families do interact with their website and have sought help from the campaign (Saur). Less than three months after the campaign was first initiated it received 807 views on the informational videos, 2,391 visits to the website, of which 1,543 visitors were new, and 670 engagements on Twitter and Facebook, making the combined reach 9,718 (Saur). However, initiatives like these can only go so far, as the initiative aims to help black families take advantage of resources and programs. However, those resources and programs themselves have various issues, as outlined.

Another proposed solution is allowing more leniency in the type of housing that can be built in neighborhoods. California Senator Atkins proposed legislation in 2021 known as the California Housing Opportunity and More Efficiency (HOME) Act, which was signed by Governor Newsom in 2021, and went into effect at the beginning of 2022 ("Senate Leader Atkins Introduces Legislation to Improve Access, Oversight for California HOME Act"). The California Housing Opportunity and More Efficiency (HOME) Act expanded housing opportunities for working families in California ("Senate Leader Atkins Introduces Legislation to Improve Access, Oversight for California HOME Act"). It simplified the process for homeowners to construct a duplex or divide their existing residential lot, enabling up to four units on a single-family property ("Senate Leader Atkins Introduces Legislation to Improve Access, Oversight for California HOME Act"). The legislation inherently works in direct opposition to the historical housing policy that only allowed single-family homes, which were not accessible for black families. An analysis conducted in 2021 by the Terner Center projected that the passage of SB 9 could potentially enable the construction of over 700,000 new homes, factoring in real-world market conditions ("California's HOME Act Turns One: Data and Insights from the First Year of Senate Bill 9"). However, the Terner Center acknowledged that the actual number of homes built would likely fall far short of this estimate ("California's HOME Act Turns One: Data and Insights from the First Year of Senate Bill 9"). Homeowners encounter various obstacles in utilizing SB 9 to subdivide their lots and develop new homes, including steep construction expenses and/or a lack of experience in home construction ("California's HOME Act Turns One: Data and Insights from the First Year of Senate Bill 9"). The research highlighted that local governments have imposed restrictions undercutting the effectiveness of SB 9 ("California's HOME Act Turns One: Data and Insights from the First Year of Senate Bill 9"). A review conducted in June 2022 of a selection of ordinances across the state revealed that certain local regulations, such as limitations on maximum unit size, height restrictions, and other design constraints, could hinder the feasibility of constructing homes under SB 9 ("California's HOME Act Turns One: Data and Insights from the First Year of Senate Bill 9"). Lawmakers can maximize the effectiveness of SB 9 by extending numerous substantive and procedural safeguards, currently provided to accessory dwelling units (ADUs), to SB 9 projects. This could include implementing flexible design standards, reducing impact fees, and simplifying the permitting process ("Is SB 9 Working? Here's What Early Data Reveals"). Statewide laws concerning accessory dwelling units (ADUs) provide more favorable conditions for builders seeking to construct additional housing on single-family lots, offering benefits such as reduced fees, absence of owner-occupancy mandates, standardized design regulations, and streamlined permitting processes ("Is SB 9 Working? Here's What Early Data Reveals"). In contrast, SB 9 projects do not benefit from any of these protections ("Is SB 9 Working? Here's What Early Data Reveals").

Aside from housing-based solutions, a solution that reforms the implicit segregation through the education side of the issue may also have profound effects. As previously mentioned, housing segregation contributes to educational disparities because marginalized communities often can not afford to live in wealthier neighborhoods. This, in turn, affects education as schools establish boundaries based on residential areas, thereby excluding outsiders from attending. Thus solutions that would solve this could come from either the housing side or the education side, or preferably from both aspects. Both housing and education policies have problems, and while the issues are interconnected, the root cause lies in both sectors, necessitating reforms in both aspects. Multiple studies find that school districts may draw their attendance boundaries to further racial segregation (Chang). The drawing of these boundaries may be driven by political or other biases. Districting is largely done by school board members who of course have their own biases and opinions, board members who lean left generally draw boundaries that reduce segregation compared to board members with other views (Chang). The effects of attendance districting are severe as studies have found that segregation in schools exists at the same level that it did in the South almost immediately after Brown v. Board of Education was passed (Chang). School boards can draw districts to further segregation, or they can draw them to increase diversity and reduce segregation. In 2007, a decision made by the Supreme Court prohibited school boards from using a student's racial background to determine their school attendance, however, the judge also stated that school boards would be allowed to use demographic data to draw the district boundaries (Chang).

Redrawing districts to reduce segregation would greatly influence educational segregation. There are a few ways that this can happen. Legislation may be passed that would prohibit the use of demographic data in drawing boundaries, this would prevent school boards from drawing boundaries to determine where students of certain backgrounds may go to school. The proposal or implementation of this would face much opposition citing feasibility and practicality. However, the argument is almost entirely unfounded as numerous magnet, charter, and private schools do not use district boundaries, and they also do not face additional problems (Mathews). Another option is to get rid of attendance boundaries as a whole, however, this would have other unintended consequences, such as overcrowding in certain schools, but issues such as these could be solved by implementing attendance caps. The other impact of eliminating district boundaries is that school funding would dramatically change (Merod). One of the functions of district boundaries is to determine which residential areas funding from taxes can come from, getting rid of district boundaries would have severe implications for this system (Turner et al.). However, this may be beneficial in dismantling the disparities in funding. Rather than residents paying taxes for the schools within their district, these taxes could be potentially collected by the state and then equally distributed amongst all schools. Another solution is to utilize districting and draw attendance boundaries in a certain way that would promote more diversity, but this would be difficult to oversee, due to the aforementioned fact that districts are often drawn based on bias from school board members. Thus to implement this solution the task of districting would have to be assigned to a bipartisan committee dedicated to ensuring that diversity is promoted (Carey). Aside from this, districts should be redrawn every 10 years, similar to congressional districts, following receiving updated information regarding demographics from censuses, to ensure that the district boundaries remain effective (Carey).

The solution to solving housing and education disparities is multi-faceted. A solution only backed by private organizations or initiatives might lack the political support that it needs to truly have a huge reach. However, legislation alone would face obstacles within cities and municipalities and lack widespread recognition. The most effective solution involves a combination of legislation, support from private organizations, and increasing access to these implementations to yield the most effective results. To maximize the issues that it would solve the solution should be twofold with reforms to housing and education policy. The findings highlight the necessity of extending key protections and incentives already established for accessory dwelling units (ADUs) to SB 9 projects, such as flexible design standards, reduced impact fees, and streamlined permitting processes. Additionally, local governments should reassess existing regulations that may hinder the feasibility of SB 9 developments, aiming to create a more conducive environment for increased housing density on single-family lots. In addition, to ensure federal-level benefits, SB9 can be adopted by state leaders with compliance with other state-level laws and policies. State legislation may also be passed to reform attendance districting for the school board. When implementing this solution, it is crucial to take into account the implications it will have on resource allocation for school districts and the changes in demographics that will ensue. To see the long-term benefits of this solution once legislation is implemented, there must be compliance within municipalities and support from private organizations to ensure that the communities that should be benefiting from this proposal have the resources and information they need to access the benefits.

The historical examination of federal housing policies, particularly the Federal Housing Administration's (FHA) role in redlining, elucidates the entrenched systemic disparities in housing opportunities and community development. The repercussions of these policies extend beyond housing, affecting educational outcomes and perpetuating cycles of disadvantage for marginalized communities. The intersectionality of housing and education underscores the need for comprehensive policy reforms to address systemic inequalities. Initiatives such as the Building Black Wealth campaign and programs like the Housing Choice Voucher program offer potential avenues for mitigating disparities, but they must be supplemented with legislative and structural changes to ensure their effectiveness. The implementation of laws prohibiting discrimination based on income sources, as well as measures to promote equitable housing development, like the California Housing Opportunity and More Efficiency (HOME) Act, are vital steps toward dismantling systemic inequities. However, true progress requires a multifaceted approach, encompassing both legislative reforms and community-based initiatives, to create a more just and equitable society where every individual has access to quality housing and education.

Works Cited

Agency, California Housing Finance. "Building Black Wealth." *Www.calhfa.ca.gov*, www.calhfa.ca.gov/community/buildingblackwealth.htm.

Burke, Lindsey, and Jude Schwalbach. "Housing Redlining and Its Lingering Effects on Education Opportunity." *The Heritage Foundation*, 11 Mar. 2021, www.heritage.org/education/report/housing-redlining-and-its-lingering-effects-education-opportunity#:~:text=This%20practice%20of%20redlining%20had.

"California's HOME Act Turns One: Data and Insights from the First Year of Senate Bill 9." *Terner Center*, 18 Jan. 2023, ternercenter.berkeley.edu/research-and-policy/sb-9-turns-one-applications/#:~:text=A%202021%20analysis%20by%20the. Accessed 31 Mar. 2024.

Carey, Kevin. "No More School Districts!" *Democracy Journal*, 10 Dec. 2019, democracyjournal.org/magazine/55/no-more-school-districts/.

Centers for Disease Control and Prevention. "Health Effects of Lead Exposure | Lead | CDC." *Www.cdc.gov*, 2 Sept. 2022, www.cdc.gov/nceh/lead/prevention/health-effects.htm.

Chang, Alvin. "We Can Draw School Zones to Make Classrooms Less Segregated. This Is How Well Your District Does." *Vox*, Vox, 8 Jan. 2018, www.vox.com/2018/1/8/16822374/school-segregation-gerrymander-map.

Chen, Grace. "What Is the Connection between Home Values and School Performance? | PublicSchoolReview.com." *Public School Review*, 13 July 2013, www.publicschoolreview.com/blog/what-is-the-connection-between-home-values-and-school-performance.

Cunningham, Mary, et al. "A Pilot Study of Landlord Acceptance of Housing Choice Vouchers." *Urban Institute*, 20 Aug. 2018, www.urban.org/research/publication/pilot-study-landlord-acceptance-housing-choice-vouchers.

Davis, Flora. "Three-Generation Households: Are They History?" *Silver Century Foundation*, 27 Mar. 2017, www.silvercentury.org/2017/03/three-generation-households-are-they-history/#:~:text=Prosperous%2C%20white%20Victorians%20weren. Accessed 28 Mar. 2024.

Demsas, Jerusalem. "America's Houses Are Old. Low-Income Renters Are Suffering because of It." *Vox*, 22 July 2021, www.vox.com/2021/7/22/22586701/housing-aging-public-housing-section-8.

Drew, Rachel. "A Very Brief History of Housing Policy and Racial Discrimination | Enterprise Community Partners." *Www.enterprisecommunity.org*, 17 Dec. 2020, www.enterprisecommunity.org/connect/blog/very-brief-history-housing-policy-and-racial-discrimination#:~:text=The%20result%20of%20these%20failures. Accessed 28 Mar. 2024.

Eide, Stephen. “Housing First and Homelessness: The Rhetoric and the Reality.” *Manhattan Institute*, 21 Apr. 2020, manhattan.institute/article/housing-first-and-homelessness-the-rhetoric-and-the-reality.

Ellen, Ingrid, et al. “Why Don’t Housing Voucher Recipients Live near Better Schools? Insights from Big Data.” *Furmancenter.org*, June 2016, furmancenter.org/research/publication/why-don8217t-housing-voucher-recipients-live-near-better-schools-insights-f. Accessed 31 Mar. 2024.

Fischel, William A. *Making the Grade : The Economic Evolution of American School Districts*. University of Chicago Press, 2009.

Fischer, Will. “HUD Expands Promising Policy to Support Housing Choice.” *Center on Budget and Policy Priorities*, 1 Nov. 2023, HUD Expands Promising Policy to Support Housing Choice.

Gerken, Matthew, et al. “Addressing the Legacies of Historical Redlining.” *Urban Institute*, 24 Jan. 2023, www.urban.org/research/publication/addressing-legacies-historical-redlining#:~:text=%E2%80%9CRedlining%E2%80%9D%20of%20neighborhoods%2C%20one.

Gross, Terry. “A ‘Forgotten History’ of How the U.S. Government Segregated America.” *NPR*, NPR, 3 May 2017, www.npr.org/2017/05/03/526655831/a-forgotten-history-of-how-the-u-s-government-segregated-america.

Guastaferro, Lynette. “Why Racial Inequities in America’s Schools Are Rooted in Housing Policies of the Past.” *USA TODAY*, 2 Nov. 2020, www.usatoday.com/story/opinion/2020/11/02/how-redlining-still-hurts-black-latino-students-public-schools-column/6083342002/.

Hendrix, Michael. “America’s Failed Experiment in Public Housing.” *Governing*, 10 May 2021, www.governing.com/community/americas-failed-experiment-in-public-housing.

Housig Choice Voucher Program Guidebok. 2021, www.hud.gov/sites/dfiles/PIH/documents/HAP_Contracts_HCV_Guidebook_Chapter_July_2021.pdf.

Housing Assistance Payments (HAP) Contract. www.hud.gov/sites/dfiles/OCHCO/documents/52641ENG.pdf.

“Housing Choice Voucher (Section 8) | USAGov.” *Www.usa.gov*, www.usa.gov/housing-voucher-section-8.

“Housing Choice Voucher Program (Section 8).” *Benefits.gov*, 2019, www.benefits.gov/benefit/710.

“Is SB 9 Working? Here’s What Early Data Reveals.” *California YIMBY*, 22 Feb. 2023, cayimby.org/blog/is-sb-9-working-heres-what-early-data-reveals/. Accessed 31 Mar. 2024.

Lafortune, Julien. "Understanding the Effects of School Funding." *Public Policy Institute of California*, May 2022, www.ppic.org/publication/understanding-the-effects-of-school-funding/.

Lukes, Dylan, and Christopher Cleveland. "The Lingering Legacy of Redlining on School Funding, Diversity, and Performance." *The Lingering Legacy of Redlining on School Funding, Diversity, and Performance*, 2021, pp. 21–363, https://doi.org/10.26300/qeer-8c25.

Mathews, Jay. "Perspective | Why We Must Shed Old Fears of Changing School Boundaries to Help Poor and Minority Kids." *Washington Post*, 19 Dec. 2021, www.washingtonpost.com/education/2021/12/19/school-boundaries-integration/.

Meckler, Laura, and Kate Rabinowitz. "The Lines That Divide: School District Boundaries Often Stymie Integration." *Washington Post*, 16 Dec. 2019, www.washingtonpost.com/education/2019/12/16/lines-that-divide-school-district-boundaries-often-stymie-integration/.

Merod, Anna. "Report: District Boundaries, Affordable Housing Access Fuel Funding Disparities." *K-12 Dive*, 15 Oct. 2021, www.k12dive.com/news/report-district-boundaries-affordable-housing-access-fuel-funding-dispari/608155/. Accessed 11 Apr. 2024.

Rancaño, Vanessa. "The Block That Prop. 13 Built: Public Schools, Public Trust." *Projects.scpr.org*, projects.scpr.org/prop-13/stories/education/.

Saur, Chris. "CalHFA Address Black Homeownership Gap with Building Black Wealth Campaign 2021 Annual Awards for Program Excellence Entry California Housing Finance Agency CalHFA Addresses Black Homeownership Gap with Building Black Wealth Campaign Communications: Integrated Campaign CalHFA Addresses Black Homeownership Gap with Building Black Wealth Campaign Summary." *Suite*, vol. 438, 2021, www.ncsha.org/wp-content/uploads/California-Communications-Integrated-Campaign-2021.pdf. Accessed 31 Mar. 2024.

"School Districts That Serve Students of Color Receive Significantly Less Funding." *The Education Trust*, 8 Dec. 2022, edtrust.org/press-release/school-districts-that-serve-students-of-color-receive-significantly-less-funding/.

"Senate Leader Atkins Introduces Legislation to Improve Access, Oversight for California HOME Act." *Senator Toni G. Atkins*, 20 Mar. 2023, sd39.senate.ca.gov/news/20230320-senate-leader-atkins-introduces-legislation-improve-access-oversight-california-home.

"Source of Income Laws." *Local Housing Solutions*, localhousingsolutions.org/housing-policy-library/source-of-income-laws/.

Staff, Crusader. "L+M Development Partners and SAA|EVI Acquire Concordia Place Apartments in Chicago." *The Chicago Crusader*, 23 Dec. 2022, chicagocrusader.com/lm-development-partners-and-saaevi-acquire-concordia-place-apartments-in-chicago/#:~:text=Originally%20built%20in%201969%2C%20Concordia. Accessed 28 Mar. 2024.

Turner, Cory, et al. “Why America’s Schools Have a Money Problem.” *NPR*, 18 Apr. 2016, www.npr.org/2016/04/18/474256366/why-americas-schools-have-a-money-problem.

U.S. Department of Education. “Digest of Education Statistics, 2018.” *Nces.ed.gov*, 2018, nces.ed.gov/programs/digest/d18/tables/dt18_206.40.asp. Accessed 28 Mar. 2024.

“U.S. Lawmakers Demand HUD ‘Swiftly Examine’ Mold, Rodents, Other Issues at Concordia Place Apartments.” *Www.cbsnews.com*, 16 Mar. 2021, www.cbsnews.com/chicago/news/concordia-place-apartments-mold-mice-housing-urban-development/.

Predictive Modeling of Gun Violence Using Machine Learning: Understanding the Role of Demographic and Socioeconomic Factors at the County Level

Rohan Singhal

Abstract

In a society plagued by the implications of gun violence, the need for effective risk assessment measures has become evident. This project addresses the challenge of predicting gun violence risks at the county level in the United States by harnessing the power of predictive modeling. We hypothesize that demographic and socioeconomic factors, such as unemployment and poverty rates in 2021, influence the likelihood of gun violence in various regions. Through an intricate analysis of multifaceted datasets encompassing crime rates, socioeconomic indicators, and geographic characteristics, this research seeks to uncover the extent to which these factors impact the risk of gun violence at the county level. Analyzing 2021 data, the research identifies counties with specific poverty-related characteristics that significantly elevate the risk of shootings. The study's use of decision trees provides a glimpse into the predictive model's classification process. By decoding the complex dynamics of gun violence, this research opens new avenues for understanding and addressing the factors contributing to shootings, ultimately guiding the way towards a safer society.

Key Words

Gun Violence, Poverty, County, Demographic Factors, Crime Rates, Analysis, Machine Learning

Introduction

Gun violence is a pressing issue in modern society, particularly in the United States, where devastating mass shootings have prompted urgent discussions on prevention. Despite the focus on gun control policies, understanding the multifaceted factors underlying gun violence, especially at the county level, remains a critical gap. This research uses advanced predictive modeling techniques to analyze gun violence risk at the county level, highlighting the intricate relationship between demographic, socioeconomic, and geographic factors that influence these occurrences.

Gun violence poses a complex challenge with no single cause. It is influenced by various factors, including social, economic, and geographic dynamics. By recognizing this complexity, this research delves into county-level data to unveil patterns shaping gun violence in the United States. Leveraging comprehensive datasets, including crime rates, poverty, unemployment, and educational indicators, this analysis seeks to unravel the nuanced relationship between socioeconomic status and gun violence risks in specific regions.

Our hypothesis posits that demographic and socioeconomic factors, among other variables, contribute significantly to varying levels of gun violence risks across counties. To explore this intricate relationship, comprehensive datasets are vital, enabling us to move beyond reactive measures and adopt proactive, data-driven approaches. By employing cutting-edge technology and in-depth analysis, this research contributes to the growing knowledge surrounding gun violence. It informs evidence-based policies and interventions, paving the way for safer communities and bolstering social security.

Literature Review

The prediction of gun violence, particularly school shootings, is a topic of great concern and research interest. Understanding the underlying factors and patterns that contribute to these incidents is crucial for prevention and intervention efforts. This literature review explores recent research that delves into the intricate task of predicting gun violence, with a focus on community-level factors, machine learning approaches, and dynamic factor models.

Blair T. Johnson's study "Community-level factors and incidence of gun violence in the United States, 2014-2017" (2021) investigates the role of community-level factors in the incidence of gun violence. The research examines data from 2014 to 2017 and identifies variables at the community level that may contribute to gun violence. Johnson's work found that the extent to which counties are urban was the most robust predictor of both gun violence incident and casualty rates. Similarly, places characterized by greater income disparity were also more likely to experience higher gun violence rates, especially when high income was paired with high poverty. These findings highlight the complex relationship between urbanization, income disparity, and gun violence, providing valuable insights for future predictive models.

Dana E. Goin's study "Predictors of firearm violence in urban communities: A machine-learning approach" (2018) employs machine learning methodologies to predict firearm violence within urban communities. By analyzing a range of variables, including demographic factors and neighborhood characteristics, Goin's research found that certain variables, such as neighborhood poverty rates and educational attainment, significantly influence firearm violence. This machine learning approach provides a data-driven foundation for developing effective strategies to address gun violence in urban settings.

Salvador Ramallo's research "A dynamic factor model to predict homicides with firearm in the United States" (2023) introduces a dynamic factor model to predict homicides involving firearms. This innovative approach combines dynamic modeling techniques with comprehensive data to forecast firearm-related homicides. Ramallo's work reveals the temporal dynamics of gun violence, indicating that specific factors, such as firearm sales and legislative changes, impact the occurrence of firearm-related homicides. This dynamic factor model offers a novel perspective for predictive modeling in this critical area, emphasizing the importance of considering time-varying factors in gun violence prediction.

These studies represent a diverse range of research endeavors aimed at advancing our understanding of gun violence prediction. They underscore the importance of community-level factors, machine learning, and dynamic modeling in enhancing predictive accuracy. The findings from these studies highlight various factors, including urbanization, income disparity, neighborhood poverty rates, educational attainment, firearm sales, and legislative changes, as significant predictors of gun violence.

Building upon the foundations laid by these scholars, this project aims to contribute to the ongoing discourse on predictive modeling for gun violence by analyzing how various measures of poverty and income disparity play a specifically predictive role in gun violence likelihood. This paper demonstrates more granularity and nuance into which specific measures of poverty and income have strong predictive power in gun violence outcomes across the United States.

Methodology

Data Collection and Selection

This project's foundation lies in the meticulous collection and curation of diverse datasets that capture the various dimensions of gun violence and potential determinants. Multiple sources were explored to acquire relevant data, ensuring a comprehensive representation of the factor influencing gun violence at the county level. The key data sources encompassed the Center for Homeland Defense and Security (CHDS) shooting incidents dataset, the Mother Jones mass shootings dataset, the Integrated Public Use Microdata Series (IPUMS) census data, and county-level crime statistics from the Federal Bureau of Investigation (FBI) Uniform Crime Reporting (UCR) program. The selection of these datasets was driven by their comprehensive coverage of variables related to gun violence, socioeconomic indicators, and demographic factors.

Data Preprocessing

The selected datasets were carefully cleaned and transformed to ensure their compatibility for analysis. Python emerged as the programming language of choice due to its versatility and extensive libraries for data manipulation. The CHDS and Mother Jones datasets were merged, leveraging the city, state, and date variables to create a unified incident timeline. The IPUMS census data was integrated using the Federal Information Processing Standards (FIPS) code, a unique identifier for each county. This step facilitated the alignment of socioeconomic and demographic information with gun violence incidents.

Feature Selection

The complexity of the relationship between gun violence risks and the myriad of predictor variables necessitated a thoughtful approach to feature selection. Key demographic variables, such as population density, education, and racial composition, were chosen due to their potential influence on crime rates. Socioeconomic factors, including poverty rates and unemployment rates for the year 2021, were also included, reflecting their strong connection to community well-being and crime. These features were augmented with crime data, encompassing various offense categories, allowing for a comprehensive assessment of the crime landscape.

Geographic Integration

An essential component of the analysis involved integrating geographic information with the datasets. The Google Geocoding API and Google Maps Geolocation Services were utilized to geocode the location information provided in the datasets and derive latitude and longitude coordinates. These coordinates were then employed to query the API, yielding the FIPS code, which encapsulates the geographical identifier at the county level. This enabled the linkage of gun violence incidents to specific counties, facilitating spatial analysis.

Long-Wide Data Conversion

To extract temporal insights, the data underwent a conversion from wide to long format. This transformation allowed us to extract the year from variables related to unemployment and poverty rates, creating a separate year column. Subsequently, the data was transformed back to its original wide format, now enriched with this temporal information. This approach provided a granular temporal dimension, enabling a closer examination of trends and patterns over time.

Observational Level

The analysis is conducted at the county level, with each county-year combination represented by a unique observation unit. This approach facilitates a granular examination of how different variables contribute to gun violence risks across diverse counties over time. The merged dataset encompasses a comprehensive timeline of gun violence incidents, demographic indicators, socioeconomic metrics, and crime rates, creating an integrated resource for analysis.

By intricately manipulating and integrating diverse datasets, harnessing geographic insights through the Google Geocoding API and Google Maps Geolocation Services, and employing advanced python programming techniques, this methodology provides a robust foundation for the subsequent stages of analysis.

Statistical Methods

Ten-Fold Cross Validation and Class Imbalance

As the analysis progressed, an essential aspect of model selection and evaluation was addressing the challenge of class imbalance with the gun violence dataset. The logistic regression model, initially employed, revealed an unexpected pattern in its predictions. While the model appeared accurate when predicting the absence of a shooting (coded as 0), its performance significantly deteriorated when predicting the occurrence of a shooting (coded as 1). This disparity stemmed from the class imbalance present in the dataset, where instances of shootings were notably less frequent compared to non-shooting instances.

Class imbalance refers to an unequal distribution of classes in a dataset, which can lead to skewed model predictions and compromised accuracy. To overcome this issue, a transition was made to a tree-based model, specifically a Random Forest classifier. This decision was motivated by their inherent capacity to handle nonlinear relationships and adapt to imbalanced datasets.

In the pursuit of refining the model’s performance, ten-fold cross validation emerged as a feasible technique. Ten-fold cross validation involved splitting the dataset into ten subsets or “folds.” The model is trained on nine of these folds and tested on the remaining fold. This process is repeated ten times, ensuring that each fold serves as the testing set once. The average performance across the ten iterations provides a more robust estimate of the model’s generalization ability.

By implementing ten-fold cross validation, the performance of the Random Forest model was evaluated across different subsets of the data. This technique effectively addressed the issue of class imbalance and offered insights into the model's consistency in predicting gun violence incidents. The outcome of this validation process proved promising, demonstrating a marked improvement in the model's accuracy and predictive power compared to the initial logistic regression attempt.

To evaluate the models' performance, various metrics were employed. For the logistic regression model, metrics such as accuracy, precision, and F1 score, specifically weighted F1 and macro F1, were used. These metrics provided a nuanced understanding of the model's predictive power. For the Random Forest model, similar metrics were used, with a particular focus on accuracy.

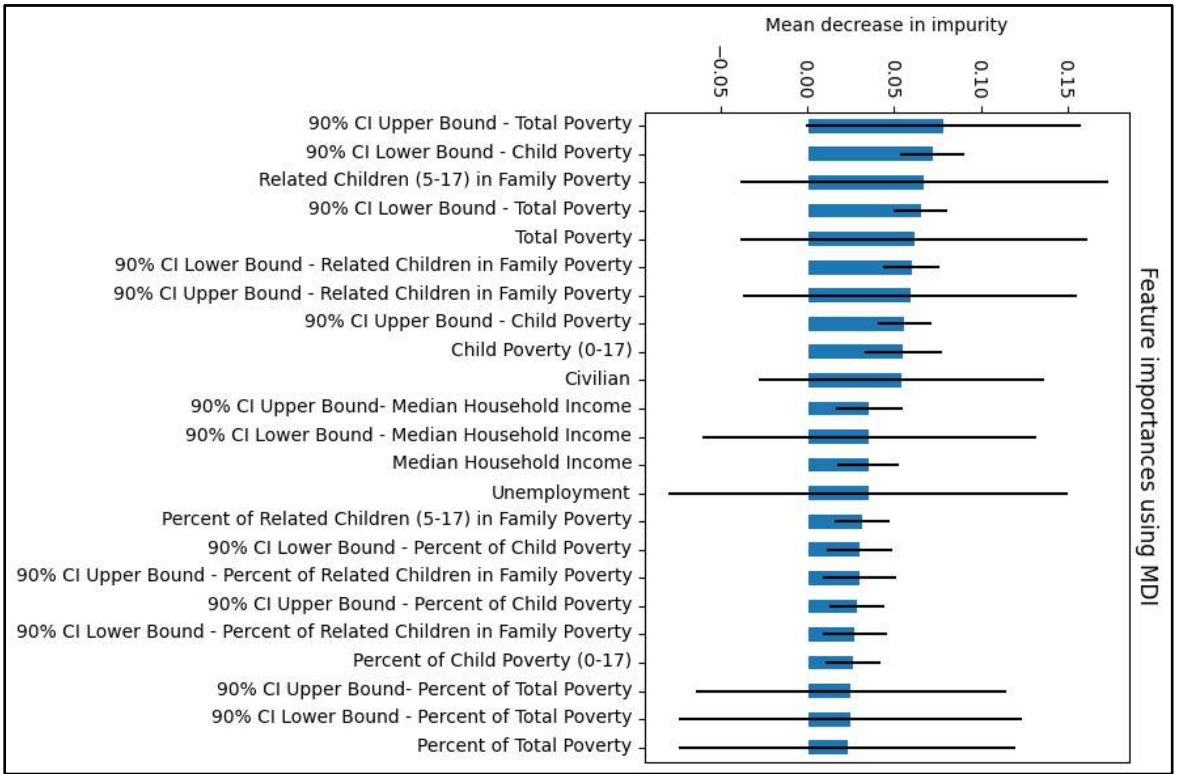
Interpretation

Results

In our analysis of predictors for shootings in 2021, poverty-related variables were the most significant. The feature importance plot shown below highlighted three key factors: counties with wider variations in total poverty (higher upper bounds), consistently high child poverty rates (high lower bounds), and more related children aged 5 to 17 in family poverty faced a significantly elevated risk of shootings. Conversely, the overall percentage of total poverty and its confidence intervals displayed weaker associations with shooting likelihood.

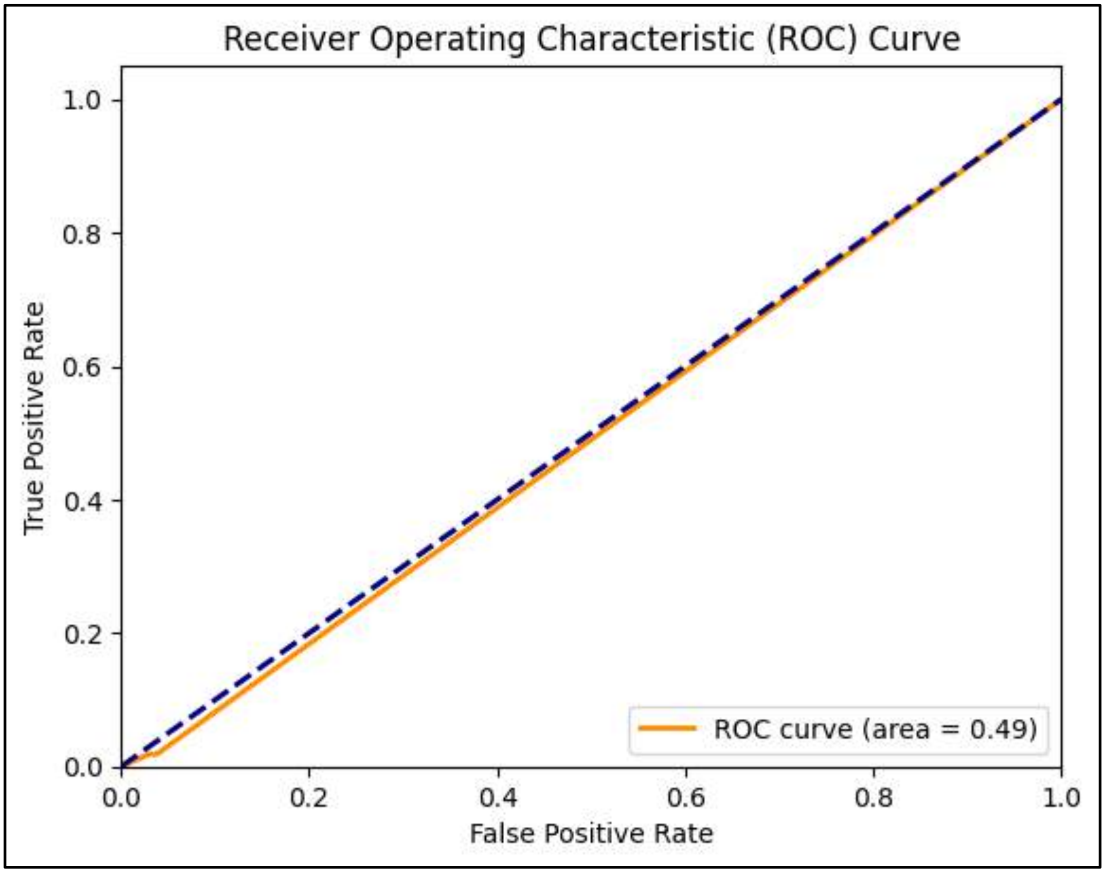
Figure 1

Based on Figure 1, it is evident that the total number of people in poverty is the most influential factor, significantly impacting the likelihood of shootings occurring at the county level.



Surprisingly, the percentage of people in poverty appeared to be the least influential, suggesting a nuanced relationship between poverty rates and gun violence. This observation indicates that the sheer size of a county's population in poverty plays a crucial role in influencing the occurrence of shootings. This finding draws attention to the significance of understanding not only the prevalence but also the demographic scale of poverty within a community, highlighting the multifaceted nature of gun violence risk.

Figure 2



During the exploratory phase of this research, we initially experimented with a logistic regression model to predict gun violence occurrences at the county level. However, the results revealed that logistic regression was not suitable for our dataset due to its limited predictive power. Based on Figure 2, the ROC curve for the logistic regression model showed an area under the curve (AUC) score of 0.49, which is close to random chance (0.50) and indicates that the model's ability to distinguish between shootings and non-shootings was no better than flipping a coin. This lack of predictive power led us to explore alternative machine learning techniques that could handle the complexity and nonlinear relationships within the data. As a result, we transitioned to using a random forest classifier, which demonstrated far better performance in predicting gun violence risk. The decision to use random forest over logistic regression was driven by the random forest model's ability to capture complex interactions and nonlinearities in the data, resulting in a more accurate and robust predictive model.

Figure 3a

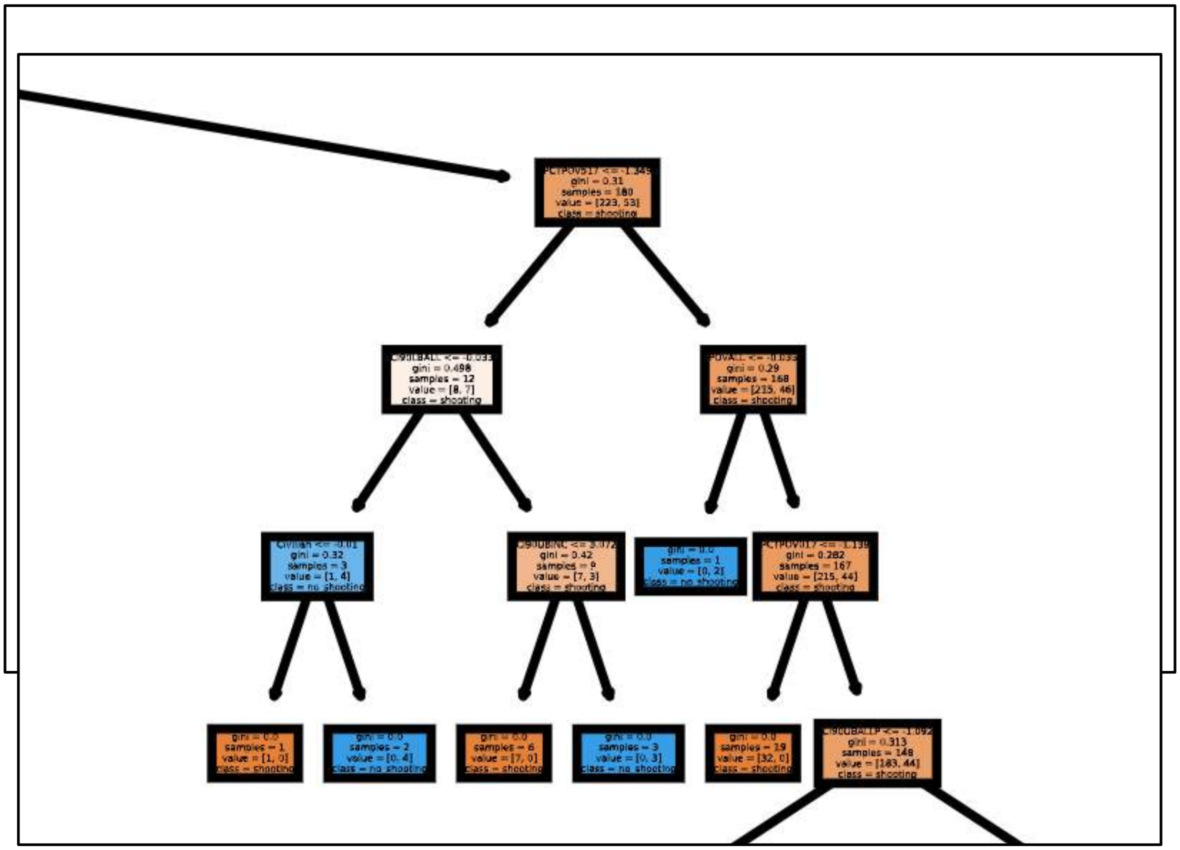


Figure 3b

To enhance our understanding, we used decision trees to help visualize the model's decision logic. Based on Figure 3a, although the tree shown above is one example out of many, it illustrates how the model classifies counties based on the aforementioned factors. Each box in the tree represents a decision node where the data is partitioned based on specific conditions, ultimately categorizing the data as either shooting or no shooting, coded as 0 or 1, respectively. While this decision tree is simplified, it accurately captures how the model classifies counties. The zoomed-in image of the decision tree (Figure 3b) provides a detailed view of the specific logic and classification criteria within a selected segment of the entire decision tree (Figure 3a).

The first variable in each box corresponds to one of the features from our dataset. For instance, it could be a feature like "Total Poverty" or "Unemployment Rate" measured at the county level. The condition associated with this variable is usually a threshold or range (e.g., less than or equal to a specific value) that determines how the data is split at that node.

The Gini coefficient is a measure of impurity or uncertainty within a node. It quantifies the probability of misclassifying a randomly chosen sample, with lower values indicating higher purity or homogeneity in terms of the target variable (shooting vs. non-shooting).

The “samples” variable indicates the number of samples (data points or instances) that fall into the specified condition at that node. It provides insight into the distribution of data and the volume of instances affected by the split condition.

The "value" variable within brackets represents the distribution of classes (shooting and non-shooting) within the samples at that node. For example, if the value is [300, 100], it means there are 300 instances classified as non-shooting and 100 instances classified as shooting.

Based on the split condition, Gini coefficient, and distribution of classes (value), the decision tree algorithm determines the class (shooting or non-shooting) for each node. The majority class or the class with higher representation within the node is often assigned as the predicted class for that subset of data. The decision tree's hierarchical structure allows it to recursively split the data based on the most informative features and conditions, gradually creating branches that lead to the final classification decisions. By following the path from the root node to the leaf nodes, we can trace how the model makes decisions at each step, ultimately predicting whether a county is likely to experience gun violence or not based on the given features and conditions.

Our study emphasizes the pivotal role of poverty-related factors, especially those with wider variations and higher lower bounds, in predicting shootings. These insights offer a better understanding of the dynamics of shooting incidents, which also helps mitigate their occurrence.

The random forest model's performance was evaluated using ten-fold cross-validation. Across the folds, the accuracy was consistently high, ranging from 0.91 to 0.95, and the average accuracy across the 10 folds was 0.93. These accuracy scores highlight the model's ability to correctly predict county-level gun violence risk across diverse datasets.

The model's effectiveness was further measured using the weighted F1 score, which accounts for both precision and recall. The F1 weighted scores ranged from 0.90 to 0.95 across the folds, and the average weighted F1 score across 10 folds was 0.92. This indicates the model's balanced performance in predicting both positive and negative instances.

Additionally, the model's performance was assessed using the macro F1 score, which considers the balance between precision and recall for each class. The F1 macro scores ranged from 0.55 to 0.81 across the folds, with an average score of 0.67. These results demonstrate the model's ability to generalize well to diverse data and maintain consistent performance across different subsets.

Moreover, the analysis of feature importance revealed key factors influencing gun violence risk. Notably, variables related to poverty, such as the 90% Confidence Interval Upper Bound for Total Poverty, exhibited the most substantial impact. Counties with higher upper bounds of total poverty were more likely to experience gun violence incidents. This insight aligns with the broader understanding that socioeconomic disparities significantly contribute to the occurrence of gun violence at the county level.

Limitations and Future Work

A surprising revelation emerged during this study regarding the influence of total population on gun violence occurrences when assessed per capita. This unexpected finding adds a layer of complexity to the analysis, highlighting the nuanced relationship between population density and the likelihood of shootings. While this study offers valuable insights into the predictive modeling of gun violence risk at the county level, there are several limitations that deserve attention for future work.

One significant limitation encountered during this research was data availability and quality. Using data from different sources presented challenges due to missing or incomplete information. Notably, certain variables obtained from the Census Bureau did not cover all FIPS codes, resulting in observations being dropped from the analysis. Additionally, the restriction of the analysis to the year 2021 was necessitated by the availability of relevant data for unemployment and poverty. This temporal limitation implies that the model's predictive ability might be influenced by the unique circumstances of that specific year.

Moreover, the selected predictor variables, while encompassing crucial socioeconomic and demographic factors, represent only a subset of the multifaceted variables that potentially contribute to gun violence risk. The study's focus on poverty and unemployment as primary predictors does not capture the full complexity of underlying factors. Future research could explore additional variables like education levels, mental health resources, and law enforcement presence for a more comprehensive understanding.

As the predictive model's foundation solidifies, future work could involve its application to longitudinal data spanning multiple years. Analyzing trends across time would provide a deeper understanding of how various factors interact and evolve to influence gun violence risk. This longitudinal analysis could uncover patterns, fluctuations, and potential correlations that remain obscured in a single-year analysis. Furthermore, enhancing the model's accuracy could entail employing advanced machine learning techniques, such as ensemble methods and deep learning algorithms. These methods could unlock further predictive power, especially when confronted with complex, nonlinear relationships among variables.

Although this study offers a comprehensive approach to assessing gun violence risk through predictive modeling, it has limitations related to data availability and issue complexity. These limitations underscore the need for continued research that expands the scope of variables, extends analysis across multiple years, and leverages more sophisticated modeling techniques. By addressing these limitations and pursuing future avenues of exploration, we can foster a deeper understanding of the multifaceted factors influencing gun violence at the county level.

Conclusion

From this analysis, poverty emerges as a critical determinant of shootings in 2021. Counties characterized by higher upper bounds of total poverty, elevated lower bounds of child poverty, and a greater number of related children aged 5 to 17 in family poverty face an increased risk of shootings. This highlights the relationship between socioeconomic factors and gun violence, emphasizing the need for approaches that address these poverty-related disparities. While this study provides valuable insights, it also highlights the limitations of focusing solely on one year of data. Future research should encompass broader temporal and spatial scopes, incorporating additional variables such as educational attainment, mental health statistics, and gun ownership sales to gain a more comprehensive understanding of the nature of gun violence.

Appendix

Variable Name	**Description**
Civilian	Civilian
Unemployment	Unemployment
CI90LB017	90% CI Lower Bound - Child Poverty
CI90LB017P	90% CI Lower Bound - Percent of Child Poverty
CI90LB517	90% CI Lower Bound - Related Children in Family Poverty
CI90LB517P	90% CI Lower Bound - Percent of Related Children in Family Poverty
CI90LBALL	90% CI Lower Bound - Total Poverty
CI90LBALLP	90% CI Lower Bound - Percent of Total Poverty
CI90LBINC	90% CI Lower Bound - Median Household Income
CI90UB017	90% CI Upper Bound - Child Poverty
CI90UB017P	90% CI Upper Bound - Percent of Child Poverty
CI90UB517	90% CI Upper Bound - Related Children in Family Poverty
CI90UB517P	90% CI Upper Bound - Percent of Related Children in Family Poverty
CI90UBALL	90% CI Upper Bound - Total Poverty
CI90UBALLP	90% CI Upper Bound- Percent of Total Poverty
CI90UBINC	90% CI Upper Bound- Median Household Income
MEDHHINC	Median Household Income
PCTPOV017	Percent of Child Poverty (0-17)
PCTPOV517	Percent of Related Children (5-17) in Family Poverty

PCTPOVALL	Percent of Total Poverty
POV017	Child Poverty (0-17)
POV517	Related Children (5-17) in Family Poverty
POVALL	Total Poverty

References

Data.gov home. (n.d.). Data.gov. https://data.gov/

Datamade. (n.d.). *GitHub - datamade/census: A Python wrapper for the US Census API.* GitHub. https://github.com/datamade/census

Education. (n.d.). https://data.ers.usda.gov/reports.aspx?ID=17829

Federal Communications Commission. (n.d.). The United States of America. https://www.fcc.gov/

Goin, D. E., Rudolph, K. E., & Ahern, J. (2018a). Predictors of firearm violence in urban communities: A machine-learning approach. *Health & Place*, *51*, 61–67. https://doi.org/10.1016/j.healthplace.2018.02.013

Goin, D. E., Rudolph, K. E., & Ahern, J. (2018b). Predictors of firearm violence in urban communities: A machine-learning approach. *Health & Place*, *51*, 61–67. https://doi.org/10.1016/j.healthplace.2018.02.013

Johnson, B. T., Sisti, A. J., Bernstein, M., Chen, K., Hennessy, E. A., Acabchuk, R. L., & Matos, M. (2021a). Community-level factors and incidence of gun violence in the United States, 2014–2017. *Social Science & Medicine*, *280*, 113969. https://doi.org/10.1016/j.socscimed.2021.113969

Johnson, B. T., Sisti, A. J., Bernstein, M., Chen, K., Hennessy, E. A., Acabchuk, R. L., & Matos, M. (2021b). Community-level factors and incidence of gun violence in the United States, 2014–2017. *Social Science & Medicine*, *280*, 113969. https://doi.org/10.1016/j.socscimed.2021.113969

National Bureau of Economic Research. (2023, September 29). NBER. https://www.nber.org/

Number of mass shootings in the U.S. 1982-2023 | Statista. (2023, September 5). Statista. https://www.statista.com/statistics/811487/number-of-mass-shootings-in-the-us/

Personal Income by County and metropolitan Area, 2021 | U.S. Bureau of Economic Analysis (BEA). (n.d.). https://www.bea.gov/news/2022/personal-income-county-and-metropolitan-area-2021

Population. (n.d.). https://data.ers.usda.gov/reports.aspx?ID=17827

Poverty. (n.d.). https://data.ers.usda.gov/reports.aspx?ID=17826

Python code for transforming lat long into FIPS codes? (n.d.). Geographic Information Systems Stack Exchange. https://gis.stackexchange.com/questions/294641/python-code-for-transforming-lat-long-into-fips-codes

Ramallo, S., Camacho, M., Marín, M. R., & Porfiri, M. (2023a). A dynamic factor model to predict homicides with firearm in the United States. *Journal of Criminal Justice*, *86*, 102051. https://doi.org/10.1016/j.jcrimjus.2023.102051

Ramallo, S., Camacho, M., Marín, M. R., & Porfiri, M. (2023b). A dynamic factor model to predict homicides with firearm in the United States. *Journal of Criminal Justice*, *86*, 102051. https://doi.org/10.1016/j.jcrimjus.2023.102051

sklearn.model_selection.cross_validate. (n.d.). Scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.cross_validate.html#sklearn.model_selection.cross_validate

Talk, M. D. (2023, January 18). Reshaping a Pandas dataframe: Long-to-Wide and vice versa. *Medium*. https://towardsdatascience.com/reshaping-a-pandas-dataframe-long-to-wide-and-vice-versa-517c7f0995ad

Unemployment. (n.d.). https://data.ers.usda.gov/reports.aspx?ID=17828

US Census Bureau. (2021, December 16). *Census Bureau releases small area income and poverty estimates for states, counties and school districts*. Census.gov. https://www.census.gov/newsroom/press-releases/2021/saipe.html

US Census Bureau. (2022, December 9). *2010 - 2021 County-Level estimation details*. Census.gov. https://www.census.gov/programs-surveys/saipe/technical-documentation/methodology/counties-states/county-level.html

US Census Bureau. (2023a, June 20). *County population by characteristics: 2020-2022*. Census.gov. https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-detail.html

US Census Bureau. (2023b, September 29). *Census.gov*. Census.gov. https://www.census.gov/

U.S. Census Population Estimates — County level | Duke University Libraries. (n.d.). https://library.duke.edu/data/sources/popest

USDA ERS - county-level data sets. (n.d.). https://www.ers.usda.gov/data-products/county-level-data-sets/

Welcome to GeoPy's documentation! — GeoPy 2.4.0 documentation. (n.d.). https://geopy.readthedocs.io/en/stable/

Writing a literature review. (n.d.). https://psychology.ucsd.edu/undergraduate-program/undergraduate-resources/academic-writing-resources/writing-research-papers/writing-lit-review.html

Zach. (2021). Pandas: How to Reshape DataFrame from Wide to Long. *Statology*. https://www.statology.org/pandas-wide-to-long/

The Effect of iPhone Proximity on Rapid Eye Movement Sleep By Jana Abualhuda

Abstract

Purpose: This experiment explores the impact of iPhone proximity on the REM sleep patterns of 10 American 16-year-old females, utilizing an Apple Watch for monitoring. REM sleep, constituting 25% of the sleep cycle, is crucial for health. The study aims to assess if iPhone proximity influences REM sleep. Insights gained will inform decisions on technology use, impacting policies and practices concerning adolescent sleep health.

Experimental Procedure: Ten 16-year old females residing in the United States set an 8-hour sleep timer, wearing an Apple Watch. Trials varied iPhone distances (7.62 dm, 15.24 dm, 30.48 dm, 60.96 dm). Sleep data, including REM percentages, were recorded. Separate nights were dedicated to each participant, minimizing carryover effects. Statistical methods identified patterns and correlations between iPhone proximity and REM sleep. Ethical considerations included informed consent and participant confidentiality. Control measures ensured a consistent environment (temperature, noise), using identical beds. Trial order randomization and cross-verification of Apple Watch data with participant logs enhanced experimental rigor.

Conclusion: In this experiment, I explored the relationship between iPhone proximity and REM sleep in 16-year old female participants, which revealed a clear correlation: greater distance from the iPhone during sleep corresponded to higher REM sleep percentages. This aligns with the experiment's hypothesis, highlighting the significance of considering electronic device distance in sleep environments. This correlation is essential due to REM sleep's role in memory, emotion, and cognition. Ultimately, mitigating iPhone proximity's influence on sleep can foster healthier sleep habits well-being.

Humans as Test Subjects Endorsement

Students and sponsors doing a human vertebrate project must complete this form. The signature of the student or students and the sponsor indicates that the project was done within these rules and regulations. Failure to comply with these rules will mean the disqualification of the project at the state level. This form must follow the Safety Sheet in the project paper.

1. Humans must not be subjected to treatments that are considered hazardous and/or that could result in undue stress, injury, or death to the subject.
2. **No** primary or secondary cultures taken directly (mouth, throat, skin, etc.) or indirectly (eating utensils, countertops, doorknobs, toilets, etc.) will be allowed. However, cultures obtained from reputable biological suppliers or research facilities are suitable for student use.
3. Quantities of food and non-alcoholic beverages are limited to normal serving amounts or less and must be consumed in a reasonable amount of time. Normal serving amounts must be substantiated with reliable documentation. This documentation must be attached to the Humans as Test Subjects Endorsement form. No project may use over-the-counter, prescription, illegal drugs, or alcohol in order to measure their effect on a person.
4. The only human blood that may be used is that which is either purchased or obtained from a blood bank, hospital, or laboratory. No blood may be drawn by any person or from any person specifically for a science project. This rule does not preclude a student making use of data collected from blood tests not made exclusively for a science project.
5. Projects that involve exercise and its effect on pulse, respiration rate, blood pressure, and so on are allowed provided the exercise is not carried to the extreme. Electrical stimulation is not permitted. A valid, normal physical examination must be on file for each test subject. Documentation of same must be attached to the Humans as Test Subjects Endorsement form.
6. Projects that involve learning, ESP, motivation, hearing, vision, and surveys require the **Humans as Test Subjects** form.

The signatures of the student or students and sponsor below indicate that the project conforms to the above rules of the Illinois Junior Academy of Science.

Were humans given food? If so, was it a serving size or less?	Not given food
Were humans subjected to exercise? If so, is there evidence of a physical on file for each test subject?	No.
Briefly describe how humans were used in the investigation.	In this experiment, ten 16-year old females in the United States wore Apple watches for sleep monitoring while their iPhones were placed at varying distances during trials. This is to understand the impact of iPhones on REM sleep in adolescents.

Describe the possible risks to humans test subjects.	Describe how each risk was handled or avoided.
1. The experiment itself might influence participants' sleep patterns. 2. The experiment may lead to fatigue if not allocated a specific, healthy amount of sleep . 3. Participants may experience discomfort due to the experimental setup, affecting the reliability of sleep data.	1. Minimize disruptions by conducting the experiment over several nights to account for potential adaptation effects. Emphasize the importance of maintaining a normal sleep routine and provide guidelines on sleep hygiene. 2. Schedule the experiment at a time aligned with participants' regular sleep patterns. Allocate at least 8 hours of sleep for participants, as recommended by the CDC guidelines. Allow sufficient time for recovery between experimental nights. 3. Clearly communicate the experimental procedures to participants in advance. Emphasize the importance of maintaining their regular sleep routine. Consider a gradual acclimatization period to the experimental conditions over several nights to minimize adaptation effects.

Nadia Ismail
(Sponsor)*

Jana Abualhuhda
(Student)

07/07/2024
(Date)

Jana Abualhuda
(Student)

*As a sponsor, I assume all responsibilities related to this project.

The Influence of iPhone Proximity on REM Sleep

Jana Abualhuda

Aqsa School

July 7, 2024

Table of Contents

Acknowledgements

I want to thank my parents for covering all of the expenses for this experiment and for providing me with their guidance wherever I needed it. Thank you to my brother-in-law, Alex, for assisting me in bringing this idea into a reality. I would further like to thank the Aqsa science teachers, Ms. Nadia and Mrs. Razeq, for aiding me in any and all questions I had throughout the course of my experiment. Lastly, I would like to give my thanks to my little sister, Liyana, as she assisted me in aesthetically designing my board to create what it currently encompasses!

Purpose

The purpose of this experiment is to investigate the potential impact of iPhone proximity on the Rapid Eye Movement (REM) sleep patterns of a 16-year-old girl in the United States. As REM sleep constitutes a vital stage, comprising approximately 25% of total sleep duration, understanding the relationship between how close an iPhone is during sleep and its potential disruption of this stage is crucial for assessing health implications (Walker, 2017). By utilizing an Apple Watch for non-invasive sleep monitoring, the experiment aims to reveal whether the proximity of an iPhone, a common electronic device emitting radiation, during sleep influences the percentage of REM sleep achieved. Ultimately, this research seeks to empower individuals, parents, and healthcare professionals with valuable insights to make informed decisions about technology use, with potential ramifications for policies and practices concerning technology and sleep health, particularly in the context of adolescent well-being.

Hypothesis

If the proximity of an iPhone during sleep is reduced, then the percentage of Rapid Eye Movement (REM) sleep will increase because minimizing the proximity of the iPhone will mitigate potential sleep disturbances associated with the device, such as exposure to blue light and the psychological impact of pre-sleep distractions (Chang et al., 2015), and overall lead to a more optimal REM sleep duration.

Variables

The independent variable in this study is the iPhone proximity to each participant's head. The control of this experiment is the iPhone when it is placed outside of the room, meaning it is approximately 60.96 decimeters away from the participant. This is the prime distance for each participant to which the phone is not located in the room at all. The dependent variable is the percentage of REM sleep over 8 hours. The constants of this experiment are the age group regarding the participants (each being 16 years old), the type of iPhone (iPhone 13), the type of Apple Watch (series 3), and the temperature of the room being 69 degrees F.

Review of Literature

Lights Out: Investigating Smartphone Effects on REM Sleep

Sleep is a fundamental aspect of human health, influencing cognitive function, emotional well-being, and physical restoration. Among the various sleep stages, Rapid Eye Movement (REM) sleep stands out as particularly critical, constituting around 25% of the sleep cycle (National Sleep Foundation, 2022). When humans sleep near their phones, it disrupts this continuous cycle. Phones serve as a constant distraction during sleep, overall contributing to a lack of REM sleep, affecting one's overall memory, mood, and learning (Walker, 2017).

The symphony of neurobiological processes that make up the sleep cycle unfolds through various stages, with REM sleep playing a starring role characterized by heightened brain activity resembling wakefulness. REM sleep is pivotal for memory consolidation, emotional regulation, and overall cognitive performance (Walker, 2017). The science behind these sleep stages, as explored by Walker in "Why We Sleep," highlights their multifaceted significance in maintaining optimal health. Further, REM sleep promotes brain development and activates the amygdala, a crucial part of the brain responsible for processing emotions.

In our modern era, smartphones, celebrated for their connectivity and convenience, paradoxically serve as potential disruptors of our sleep cycle. The constant stream of notifications and psychological stimulation induced by digital devices contribute to difficulties in falling asleep and fragmented sleep patterns (Chang et al., 2015). Beyond the psychological impact, the blue light emitted by smartphones during evening usage poses an additional challenge. Studies conducted by Chang et al. (2015) and Hatori et al. (2017) elucidate the harmful effects of blue light on melatonin production, disrupting circadian rhythms and exacerbating sleep disturbances. Being exposed to blue-light before bed can also increase the time that it takes to fully wake up in the morning, causing an overwhelming sense of fatigue. Not only is there an increased sense of fatigue, but research illustrates that insomnia rates have risen rapidly with increased phone use among the general population (Hatori et al., 2017).

While the psychological and circadian disruptions are evident, the impact of smartphone proximity on sleep involves a deeper layer of intricacy. Electromagnetic radiation emitted by smartphones, though subtle, interacts with the delicate balance of neurotransmitters and hormones involved in sleep regulation (Sivertsen et al., 2021). The potential influence on the release of neurotransmitters like serotonin and melatonin may contribute to altered sleep cycles. Electromagnetic radiation can further lead to imbalances that affect how one thinks, feels, and overall performs, as well as provoking chronic illnesses (Divan et al., 2012).

Understanding the multifaceted impact of smartphones on sleep holds profound implications for public health. Sleep disturbances, particularly disruptions to REM sleep, can cascade into a myriad of health issues, ranging from impaired cognitive function to increased vulnerability to mood disorders (Alvaro et al., 2013). As our reliance on digital devices deepens, the need to comprehend and mitigate their effects on sleep becomes increasingly crucial. The recognition of the detrimental impact of smartphones on sleep quality also prompts a call to action. Cultivating healthy sleep environments involves not only acknowledging the psychological and circadian disruptions but also considering the potential impact of subtle yet pervasive factors, such as electromagnetic radiation (Alvaro et al., 2013).

In our contemporary society, where smartphone use has become almost second nature, the implications of disrupted sleep patterns extend beyond the individual level to societal consequences. The rise in insomnia rates, closely correlated with increased phone use, raises concerns about the collective impact on public health (Twenge & Campbell, 2018). Sleep deprivation is not just an individual challenge; it's a societal issue with far-reaching consequences. Impaired cognitive function, decreased productivity, and heightened stress levels among the population may become more prevalent if sleep disturbances become the norm rather than the exception (Twenge & Campbell, 2018).

Moreover, the profound alertness and arousal provoked by smartphones, available at one's service whenever and wherever, contribute to a constant state of vigilance. This heightened state not only interferes with the ability to fall asleep but also impacts the quality of sleep. The science behind the psychological impact of smartphones on sleep is a testament to the need for a collective reevaluation of our relationship with these devices (Gradisar et al., 2013). As we navigate the interconnectedness of the digital age, promoting healthy sleep habits on a societal level becomes a responsibility shared by individuals, communities, and policymakers alike. Encouraging digital literacy, educating the public on the importance of sleep hygiene, and fostering a culture that values and prioritizes sleep are essential steps in mitigating the broader consequences of smartphone-related sleep disruptions (Gradisar et al., 2013).

The dynamic interplay between smartphones and sleep underscores the urgent need for interdisciplinary efforts to address this complex issue comprehensively. Collaborations between scientists, technologists, educators, policymakers, and healthcare professionals are essential to develop evidence-based strategies that mitigate the adverse effects of smartphone use on sleep (Giménez et al., 2014). Furthermore, fostering a culture of digital well-being, where individuals are empowered with knowledge and tools to make informed choices about their technology use, is paramount.

As we reflect on the profound impact of smartphones on our sleep patterns, it becomes clear that the responsibility extends beyond the individual to societal structures and norms. Encouraging a collective reevaluation of our relationship with technology and prioritizing sleep hygiene can contribute to a cultural shift toward healthier sleep practices. Initiatives at educational institutions, workplaces, and within communities can play a pivotal role in promoting awareness and implementing policies that prioritize sleep health (Hershner & Chervin, 2014).

Ultimately, the recognition of the complex web linking smartphones, sleep, and overall well-being should serve as a catalyst for positive change. By embracing a holistic perspective that considers the psychological, physiological, and societal dimensions of this issue, we can pave the way for a future where technology coexists harmoniously with our innate need for restful and restorative sleep. In conclusion, the intricate relationship between smartphones and sleep quality unveils a multifaceted landscape of challenges that extend beyond personal well-being. From disruptions in REM sleep to societal implications, the repercussions are profound. As technology continues to advance, our understanding of its impact on our fundamental need for sleep must evolve in tandem. Navigating this ever-changing landscape requires a holistic approach, where scientific insights, public awareness, and individual responsibility converge to foster a society that values, protects, and nurtures the essential commodity of restful sleep.

In conclusion, the intricate relationship between smartphones and sleep quality unveils a multifaceted array of challenges that extend beyond personal well-being. From disruptions in REM sleep to societal implications, the repercussions are profound. As technology continues to advance, our understanding of its impact on our fundamental need for sleep must evolve in tandem. Navigating this ever-changing landscape requires a holistic approach, where scientific insights, public awareness, and individual responsibility converge to foster a society that values, protects, and nurtures the essential commodity of restful sleep.

Materials

1. 10 Girls
   - Ten 16-year old girls residing in the United States of America to serve as participants
2. 10 Apple Watch:
   - Each participant utilized an Apple watch series 3 model for this experiment.
3. 10 iPhone 13's:
   - Ten iPhone 13's, one for each participant
4. 10 Timers:
   - Ten timers, one for each participant to set an 8-hour sleep duration.
5. 10 Beds:
   - Ten beds, one for each participant to rest in during the course of the experiment
6. 10 Bedside Tables:
   - Ten bedside tables, one for each participant for placing the iPhone at varying distances.
7. 10 Consent Forms:
   - Informed consent documents for participants and legal guardians.
8. Technological Software:
   - Software applications to create graphs and charts.
9. The Health App on every iPhone:
   - The Health App to relay the data analysis from the Apple Watch's findings regarding percentages of REM sleep.

Procedure

Participants:

Ten 16-year-old female participants in the United States of America were recruited for the experiment.

1. **Preparation Phase:**
   - Participants were instructed to set a timer for 8 hours of sleep.
   - Each participant wore an Apple Watch on their wrist.
2. **Baseline Measurement:**
   - In the first trial, each of the ten participants were asked to place their iPhone approximately 7.62 decimeters away from their head on a bedside table.
   - Participants activated the timer on their phone, put on the Apple Watch, and went to sleep.
3. **Intermediate Measurement:**
   - In the second trial, participants repeated the process, placing their iPhone at a distance of 15.24 dm from their head on a bedside table.
   - Participants followed the same procedure, activating the timer, wearing the Apple Watch, and going to sleep.
4. **Far Distance Measurement:**
   - In the third trial, participants placed their iPhone at a distance of 30.48 decimeters from their head and then again at 60.96 decimeters from their head.
   - The timer was activated, the Apple Watch was worn, and participants proceeded to sleep.
5. **Data Collection:**
   - Throughout each trial, the Apple Watch recorded sleep data, including heart rate, movement, and sleep stages.
   - The experiment was conducted on separate nights for each participant to avoid potential carryover effects.
6. **Data Analysis:**
   - The recorded sleep data, particularly focusing on REM sleep percentages, was collected and analyzed for each trial.
   - Statistical methods were employed to identify patterns and potential correlations between iPhone proximity and REM sleep.
7. **Ethical Considerations:**
   - Participants were fully informed about the nature and purpose of the experiment.
   - Informed consent was obtained from both participants and their legal guardians.
   - The experiment adhered to ethical guidelines, ensuring the well-being and confidentiality of the participants.
8. **Control Measures:**
   - To control for external factors, such as room temperature and ambient noise, the experiment was conducted in a controlled environment.
   - The same type of bed was used for each participant to minimize variations in sleep comfort.
9. **Randomization:**
   - The order of the trials (7.62 dm, 15.24 dm, 30.48 dm, 60.96 dm) was randomized across participants to account for any potential order effects.

Data & Results

Table 1.

Percentage of REM Sleep Relative to iPhone Proximity for Participants (dm) (trial 1)										
Distance	Girl 1	Girl 2	Girl 3	Girl 4	Girl 5	Girl 6	Girl 7	Girl 8	Girl 9	Girl 10
7.62	8%	6%	9%	7%	8%	10%	6%	7%	9%	11%
15.24	18%	16%	14%	13%	17%	16%	12%	17%	18%	20%
30.48	24%	21%	22%	20%	22%	21%	23%	24%	20%	25%
60.96	28%	26%	24%	23%	28%	26%	25%	29%	27%	28%

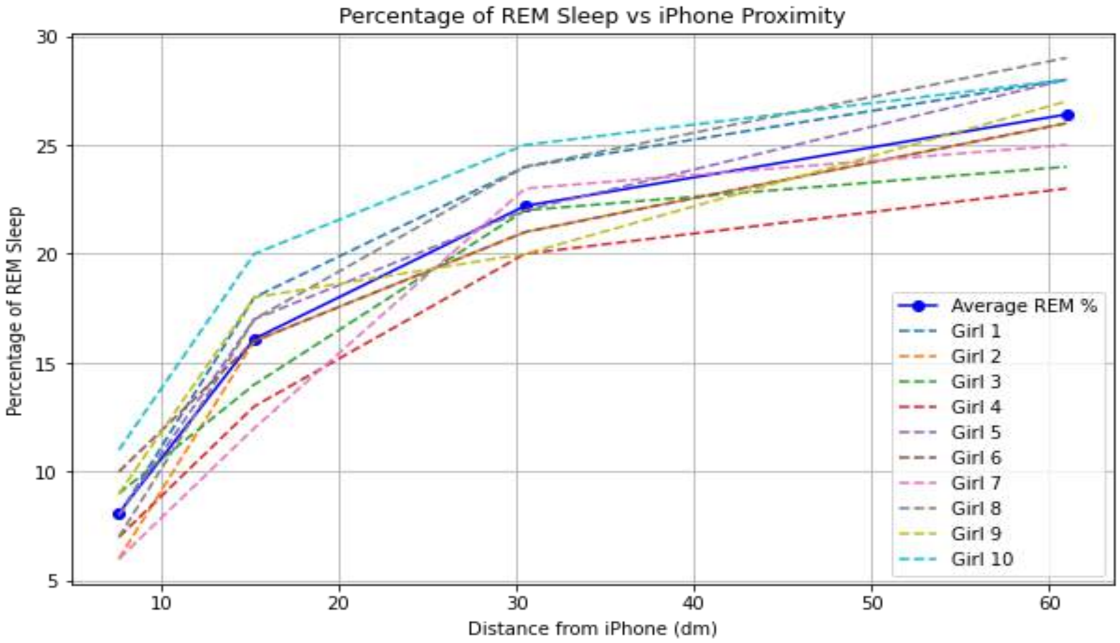
Table 2.

Percentage of REM Sleep Relative to iPhone Proximity for Participants (dm) (trial 2)										
Distance	Girl 1	Girl 2	Girl 3	Girl 4	Girl 5	Girl 6	Girl 7	Girl 8	Girl 9	Girl 10
7.62	8%	14%	13%	12%	10%	11%	12%	7%	15%	9%
15.24	19%	17%	18%	20%	18%	15%	15%	19%	14%	16%
30.48	24%	22%	25%	23%	21%	20%	24%	25%	24%	23%
60.96	26%	24%	27%	29%	25%	23%	26%	28%	23%	24%

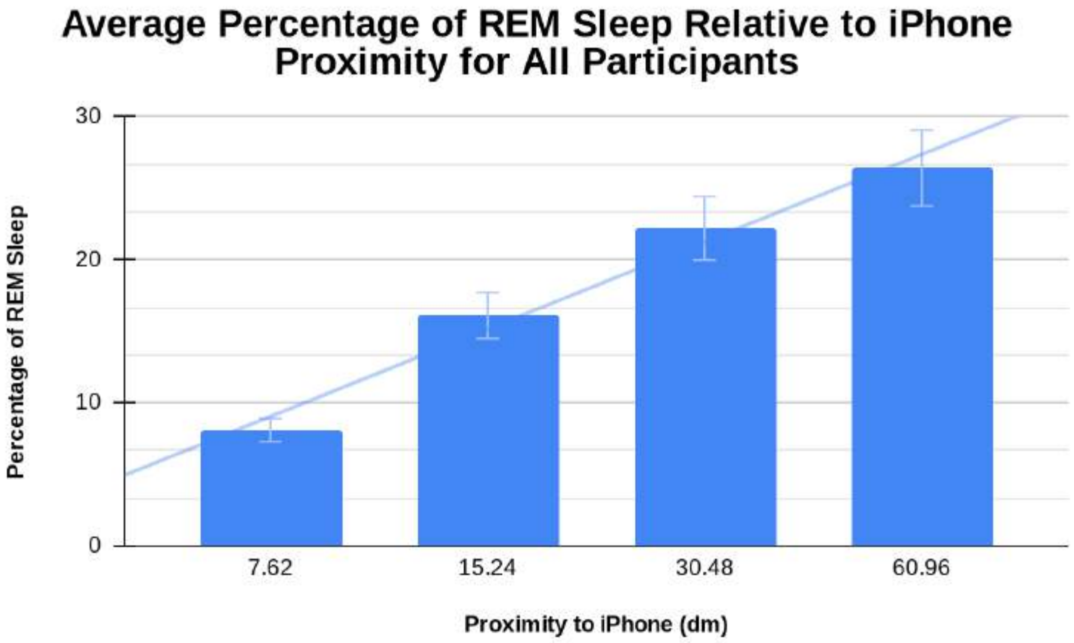
Table 3.

Percentage of REM Sleep Relative to iPhone Proximity for Participants (dm) (trial 3)										
Distance	Girl 1	Girl 2	Girl 3	Girl 4	Girl 5	Girl 6	Girl 7	Girl 8	Girl 9	Girl 10
7.62	12%	14%	11%	13%	10%	9%	12%	14%	8%	15%
15.24	18%	21%	19%	17%	19%	16%	20%	18%	15%	24%
30.48	19%	28%	25%	25%	21%	22%	26%	24%	23%	24%
60.96	23%	28%	26%	27%	24%	27%	24%	26%	28%	25%

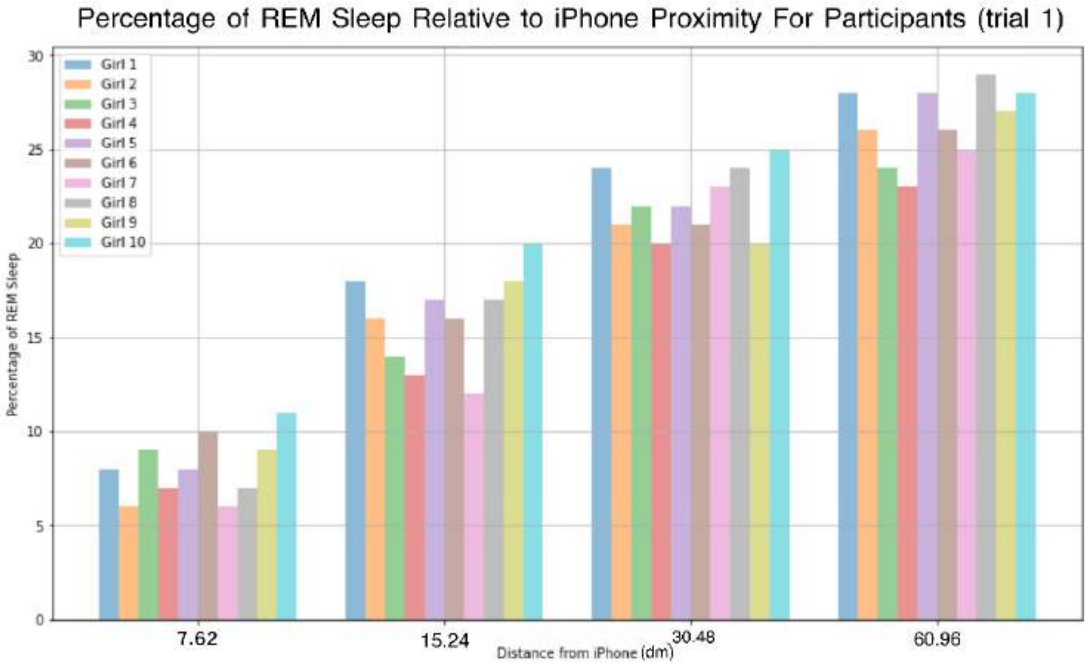
Graph 1.



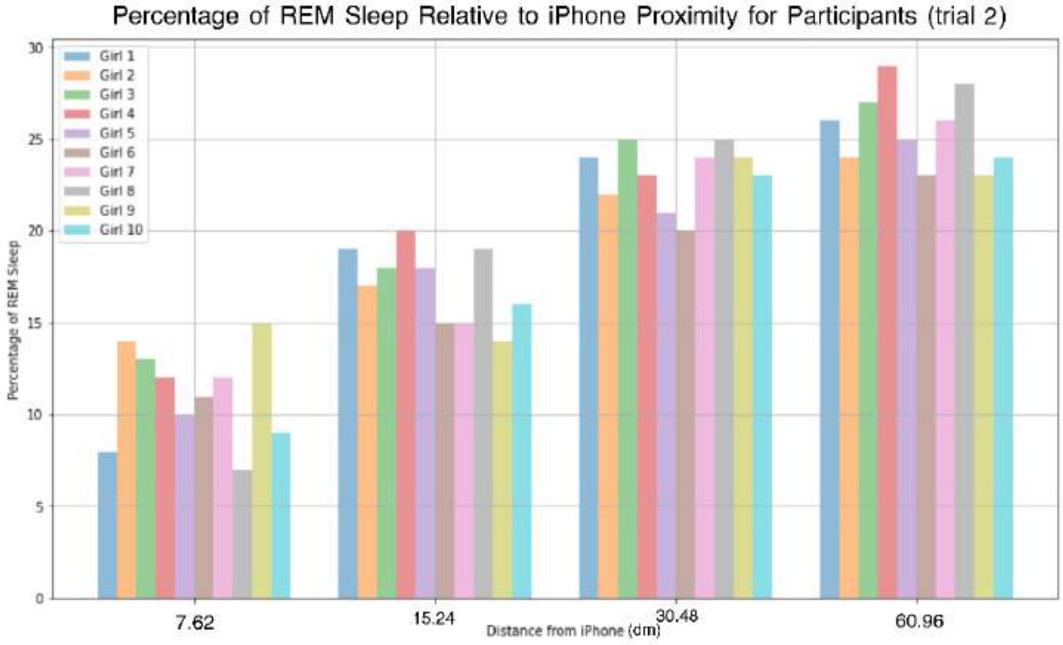
Graph 2.



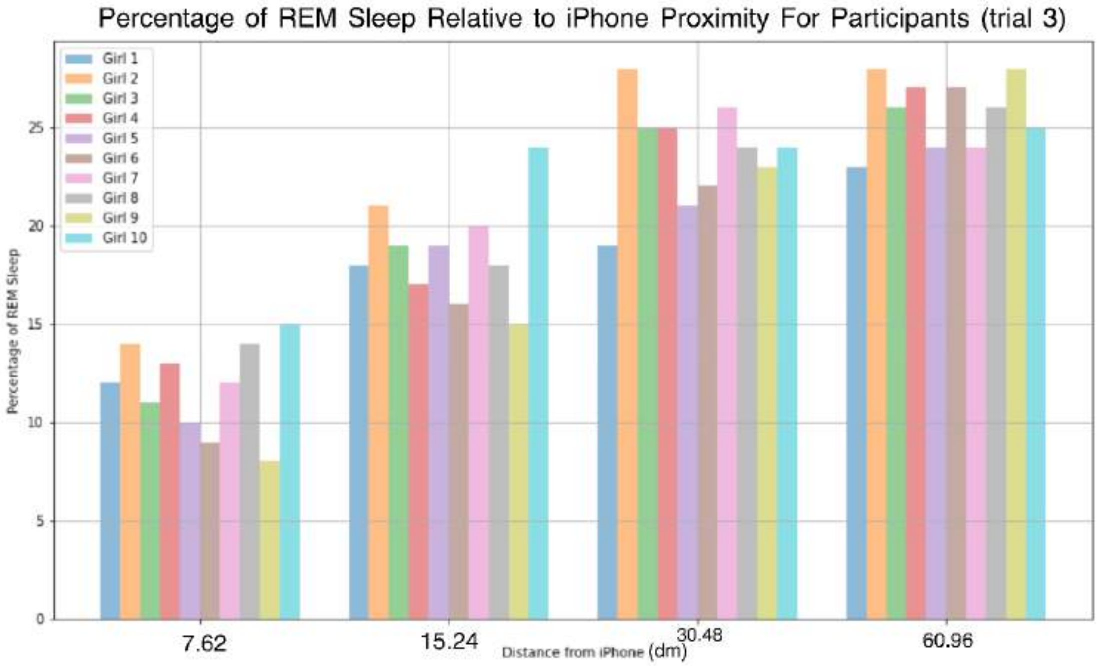
Graph 3.



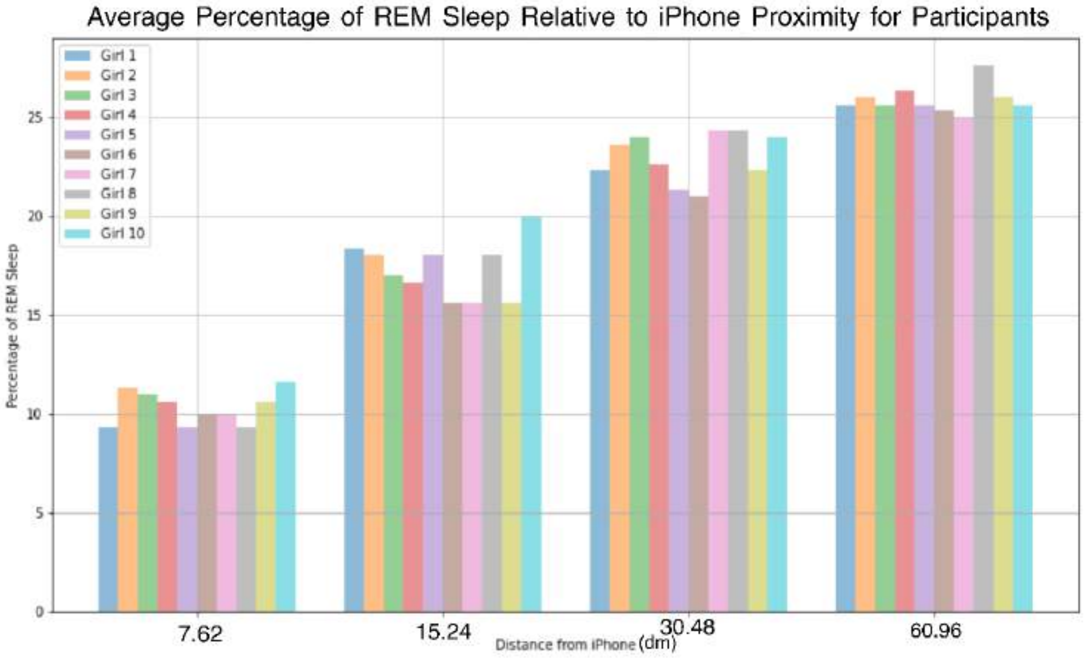
Graph 4.



Graph 5.



Graph 6.



Discussion

While testing the effect of iPhone proximity on REM sleep, there is a consistent upward trend in the graphs. The farther the proximity, the higher the percentage of REM sleep. This is evident in graph 2, as at 7.62 decimeters away, the REM sleep percentage was at 8.1%. However, when the iPhone was 15.24 dm away from the participant, the REM sleep percentage increased to 16.1%, illustrating that the farther away the iPhone is, the higher the amount of REM sleep. The same pattern is displayed when the iPhone is 30.48 dm away from each girl, as the REM sleep percentage rises to about 22.2%, which is already a 14.1% increase from the initial 8.1%. Lastly, at 60.96 dm away, the percentage of REM sleep rises to 26.4%, which displays an immense increase from the beginning percentage of only 8.1% The findings of the data further supports the hypothesis that the farther the iPhone is from each participant during their sleep, the higher their REM sleep percentage is, ultimately contributing each person's overall mood, memory, and stress!

Conclusion

In this experiment, the investigation into the relationship between iPhone proximity and REM sleep in 16-year-old participants yielded insightful results. The data consistently revealed a direct correlation: the farther the iPhone was placed from each participant during sleep, the higher the percentage of REM sleep. This observation aligns with the initial hypothesis and emphasizes the significance of considering the physical distance of electronic devices in sleep environments. The graphical representation vividly showcases this trend, with REM sleep percentages escalating as the iPhone distance increases. The most beneficial option for increasing REM sleep is placing your phone outside of your room (equivalent to 60.96 dm). This finding not only affirms the experiment's hypothesis but also highlights a critical link between iPhone proximity and sleep quality. The importance of this correlation lies in the crucial role REM sleep plays in memory consolidation, emotional regulation, and cognitive performance. The positive association between increased REM sleep percentage and greater iPhone distance suggests potential implications for mental well-being. As society grapples with the pervasive use of technology, specifically in teenagers, these findings underscore the need for a mindful reevaluation of sleep environments, especially for adolescents who may be more susceptible to the impact of electronic devices on their sleep. Ultimately, understanding and mitigating the influence of iPhone proximity on sleep can contribute to fostering healthier sleep habits and, by extension, promoting overall well-being.

Works Cited

Alvaro, Philip K., Robert M. Roberts, and Jodi K. Harris. 2013. "A Systematic Review Assessing Bidirectionality between Sleep Disturbances, Anxiety, and Depression." *Sleep* 36 (7): 1059–1068.

Chang, Anne M., D. Aeschbach, Jeanne F. Duffy, and Charles A. Czeisler. 2015. "Evening Use of Light-Emitting eReaders Negatively Affects Sleep, Circadian Timing, and Next-Morning Alertness." *Proceedings of the National Academy of Sciences* 112 (4): 1232–1237.

Giménez, M. Carmen, M. Hessels, M. van de Werken, and B. de Vries. 2014. "Effects of Artificial Dawn on Sleep Inertia, Skin Temperature, and the Awakening Cortisol Response." *Journal of Sleep Research* 23 (4): 500–506.

Gradisar, Michael, Amy R. Wolfson, Allison G. Harvey, Lauren Hale, Russell Rosenberg, and Charles A. Czeisler. 2013. "The Sleep and Technology Use of Americans: Findings from the National Sleep Foundation's 2011 Sleep in America Poll." *Journal of Clinical Sleep Medicine* 9 (12): 1291–1299.

Hatori, Megumi, Claude Gronfier, Russell N. Van Gelder, Paul S. Bernstein, Jorge Carreras, Satchidananda Panda, and Fred W. Turek. 2017. "Global Rise of Potential Health Hazards Caused by Blue Light-Induced Circadian Disruption in Modern Aging Societies." *NPJ Aging and Mechanisms of Disease* 3 (1): 1–10.

Hershner, Shelley D., and Ronald D. Chervin. 2014. "Causes and Consequences of Sleepiness among College Students." *Nature and Science of Sleep* 6: 73–84.

National Sleep Foundation. 2022. "Stages of Sleep."

Sivertsen, Børge, Allison G. Harvey, Ståle Pallesen, and Mari Hysing. 2021. "Mental Health Problems in Adolescents with Delayed Sleep Phase: Results from a Large Population-Based Study in Norway." *Journal of Sleep Research* 30 (3): e13165.

Twenge, Jean M., and W. Keith Campbell. 2018. "Associations between Screen Time and Sleep Duration Are Primarily Driven by Portable Electronic Devices: Evidence from a Population-Based Study of US Children." *Sleep Medicine* 56: 211–218.

Walker, Matthew. 2017. *Why We Sleep: Unlocking the Power of Sleep and Dreams*. New York: Simon & Schuster.

Exploring the Effects of Frequency Composition and Timbre of Noise on Active Noise Cancellation in Over-Ear Headphones

"How is the audio attenuation performance of Active Noise Cancellation in over-ear headphones affected by the frequency and timbre of the noise?"

"How is the audio attenuation performance of active noise cancellation in over-ear headphones affected by the frequency and timbre of the noise?"

Contents

1. Introduction & Background

Active Noise Cancellation (ANC) refers to the method of attenuating unwanted noise by overlaying it with an anti-noise of equal and opposite amplitude (Benoit et al.). This technique relies on the principle of wave superposition to achieve phase cancellation (Milosevic and Schaufelberger). Since German physician Paul Lueg discovered phase cancellation in 1933, ANC has most effectively been implemented in headphones/earphones due to the close proximity operating environment and ability to work alongside passive noise isolation (Dhar). Generally, ANC is effective in attenuating low-frequency noise below 1,000Hz with optimal systems working up to 1,500Hz (Elliot) whereas passive isolation, the usage of sound dampening materials, is primarily effective solely at high frequencies (Milosevic and Schaufelberger). Therefore, by working in conjunction with passive isolation, ANC headphones can achieve even attenuation across the entire human hearing spectrum (20Hz to 20,000Hz). However, since various noise control application require a dynamic approach where passive isolation is inadequate, such as slight suppression or attenuation at a specific frequency range, understanding the effects of noise characteristics like frequency and timbre is crucial for advancements in fields such as aviation, construction, dentistry, entertainment, and the military where heavy machinery such as dental drills, firearms, and aircrafts regularly expose personnel to diverse assortments of noise at harmful levels (Kaymak and Atherton).

2. Paper Overview

The purpose of this paper is to answer, **"How is the audio attenuation performance of Active Noise Cancellation in over-ear headphones affected by the frequency and timbre of the noise?"**. In context, this paper examines how pitch and types of noise (i.e. human speech, aircraft cabin noise) affect the ANC performance in terms of decibel reduction in over-ear headphones.

As such, the study first investigates the underlying theories and phenomenon that facilitate ANC, and propose the theoretical effects of frequency and timbre on the performance of an ANC system. Subsequently, an experiment is conducted to comparatively measure the attenuation of two pairs of hybrid-ANC headphones using different test signals. One pair represents an industry-leading ANC system while the other represents an intermediate system. The test signals used are: a pure tone frequency sweep, white noise, speech babble, and airplane cabin noise. Subsequently, the collected data from the ANC systems are processed to highlight key features and evaluate predictions. All successful theoretical predictions and empirical conclusions are then combined to develop a mathematical function to predict and model ANC. Specifically, one that reflects all significant features and predicts maximum attenuated frequency based on group delay value.

3. Underlying Theory & Phenomenon

ANC operates on the principle of wave superposition to achieve phase cancellation (Benoit et al.). Wave superposition states that when multiple waves overlap, the resulting wave is equal to the algebraic sum of each wave's amplitude (Russell). Essentially, sound waves are longitudinal vibrations being transferred between a series of particles in a medium, causing them to oscillate back-and-forth (Milosevic and Schaufelberger). The maximum displacement of a particle from its equilibrium position is known as the amplitude of the wave. Therefore, when two sound waves of identical but opposite amplitude matches perfectly, their superposition sums up to zero as the wave pushing the particles in a particular direction matches with an equal wave pushing said particles in the opposite direction. Thus, achieving absolute phase cancellation and eliminating the noise.

Fundamentally, ANC performs this operation using a system of microphones, electronic circuitry, and a speaker. These components work together to capture the incoming noise, process for its anti-noise, and output the anti-noise matching the original.

4. Theoretical Prediction for Frequency-Dependent Performance

4.1 Frequency of Sound

Phase refers to the amount of offset between two waves (Comeau). When two waves offsets such that the peak of one aligns with the trough of the other (see Figure 1), these waves are 180° out-of-phase (π radians). Hence, the effects of frequency can be analyzed.

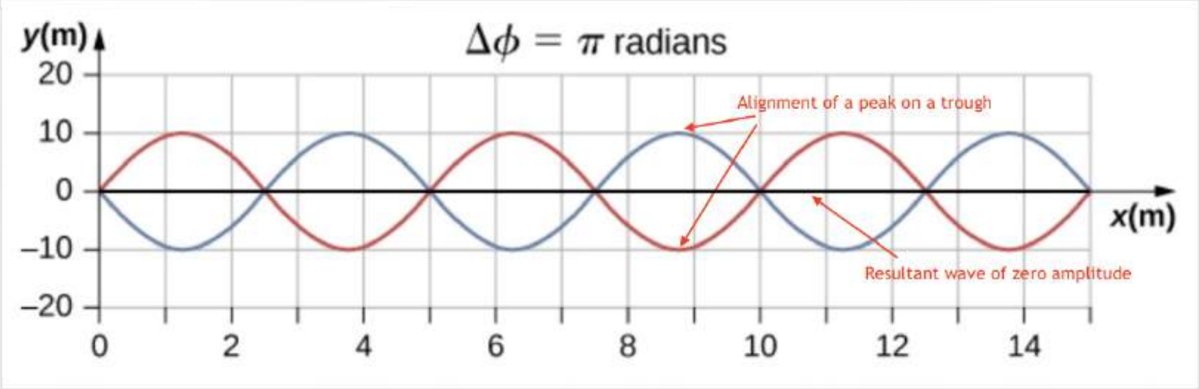


Figure 1: Annotated illustration of the superposition of two waves with identical amplitudes and frequency, but at a 180° phase shift (CNX).

Frequency refers to the number of wave cycles that pass a given point per second (Comeau). It is measured in Hertz (Hz), and it dictates the pitch of the noise. Frequency of sound is related to its wavelength (λ), which refers to the distance between 2 correlating points on the wave, through the wave equation that states:

$$c = f \cdot \lambda,$$

where c = speed of wave (m/s), f = frequency of sound (Hz), and λ = wavelength (m). As wave speed is constant in a fixed medium, isolating λ shows an inverse proportionality between wavelength (λ) and frequency (f) with a scale factor of 343:

$$\lambda = c \cdot \frac{1}{f}$$

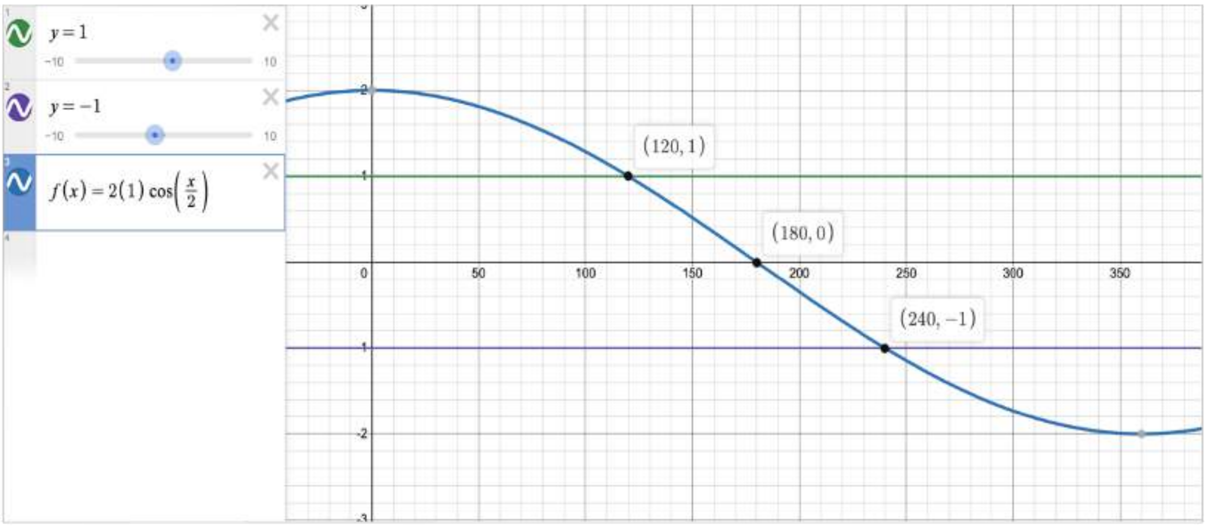
Substituting c = 343 ms^{-1}

$$\lambda = 343 \cdot \frac{1}{f}$$

While absolute phase cancellation necessitates 180° of phase difference, destructive interference, which results in noise reduction, happens anywhere between 120° to 240° of phase difference based on calculations with the following equation:

$$A_r = 2Acos\frac{\Phi}{2}$$

where A_r = the resultant amplitude, A = original amplitude, and Φ = phase difference between two waves. Therefore, the range of Φ where the resultant amplitude is reduced is found by solving the equation $|2Acos\frac{\Phi}{2}| < A$ which simplifies to $-1 < 2cos\frac{\Phi}{2} < 1$. By finding the range of x (denoting ϕ) where function $f(x) = 2(1)cos\frac{x}{2}$ lies between ± 1, shows that reduction in resultant amplitude occurs at $120^{\circ} < \phi < 240^{\circ}$.



$y = 1$

$y = -1$

$f(x) = 2(1)\cos\left(\frac{x}{2}\right)$

Figure 2: The intersection of the functions illustrates ϕ between 120° and 180° (denoted on the x-axis) gives reduced resultant amplitude (denoted by y-axis between +1 and −1).

Therefore, by converting this range of phase difference into the wave's physical path difference, the maximum anti-noise distance offset so that noise reduction occurs can be calculated. Thus, this effectively quantifies how frequency affect the degree of accuracy an ANC needs to produce an effective anti-noise.

Individual phase difference is converted to physical path difference using the following equation where $\Delta\Phi$ = phase difference in (rad), λ = wavelength of the wave (m), and Δx = path difference (m):

$$\Delta\Phi = \frac{2\pi}{\lambda}\Delta x$$

Sample path to physical difference calculations for f = 20Hz where $\lambda = 17.15m$:

1. Calculating for $\Phi = \frac{2\pi}{3}$ radians (or 120º)

$$\Delta x = \frac{\lambda}{2\pi} \cdot \Delta\Phi$$

$$\Delta x = \frac{17.15}{2\pi} \cdot \frac{2\pi}{3}$$

$$\Delta x = 5.72m$$

2. Calculating for $\Phi = \frac{4\pi}{3}$ radians (or 240º)

$$\Delta x = \frac{17.15}{2\pi} \cdot \frac{4\pi}{3}$$

$$\Delta x = 11.43m$$

This illustrates that for two 20Hz waves to be 120° phase shifted, their physical distance difference offset must be 5.72 meters, and for a phase difference of 240°, they must be 11.43 meters offset. Therefore, for a 20Hz wave, generated anti-noise can be anywhere within 11.43 - 5.72 = 5.71 meters of the original wave to create destructive interference that reduce the noise (see Figure 3). To enhance clarity, this number will be referred to as "anti-noise permissible offset".

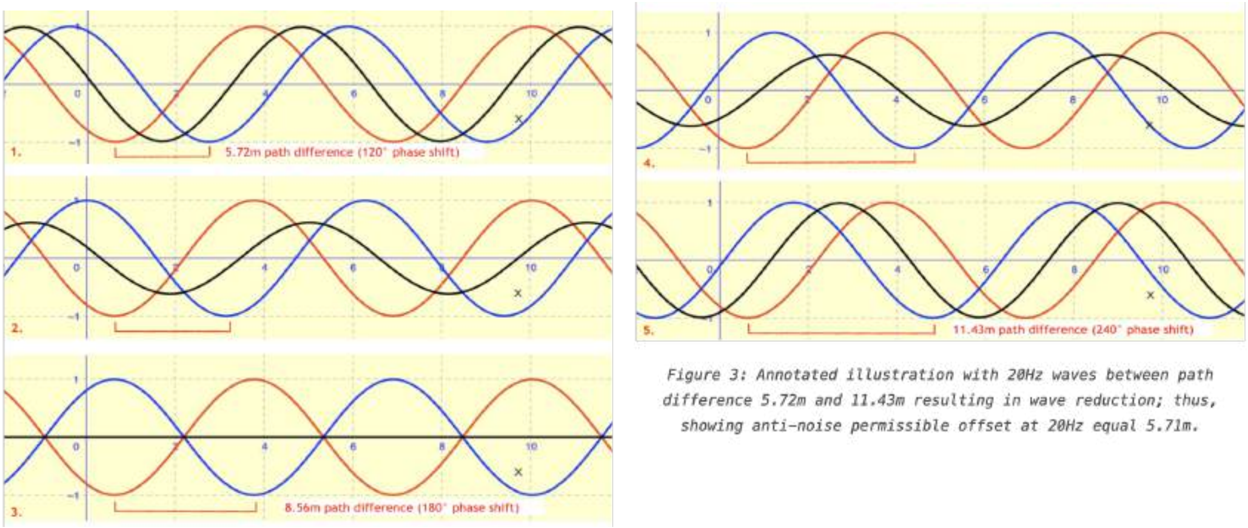


Figure 3: Annotated illustration with 20Hz waves between path difference 5.72m and 11.43m resulting in wave reduction; thus, showing anti-noise permissible offset at 20Hz equal 5.71m.

The calculations demonstrated can be derived into a mathematical function, $r(f)$, as follows:

$$r(f) = \Delta x_1 - \Delta x_2$$

Where:

$$\Delta x = \frac{\lambda}{2\pi} \cdot \Delta\Phi$$

$$\Delta x_1 = \frac{\lambda}{2\pi} \cdot \frac{4\pi}{3}$$

$$\Delta x_2 = \frac{\lambda}{2\pi} \cdot \frac{2\pi}{3}$$

Therefore:

$$r(f) = \left(\frac{\lambda}{2\pi} \cdot \frac{4\pi}{3}\right) - \left(\frac{\lambda}{2\pi} \cdot \frac{2\pi}{3}\right)$$

Where:

$$\lambda = \frac{343}{f}$$

Therefore, substituting λ:

$$r(f) = \left(\frac{\frac{343}{f}}{2\pi} \cdot \frac{4\pi}{3}\right) - \left(\frac{\frac{343}{f}}{2\pi} \cdot \frac{2\pi}{3}\right)$$

$$r(f) = \left(\frac{\frac{343}{f}}{2\pi} \cdot \frac{4\pi}{3}\right) - \left(\frac{\frac{343}{f}}{2\pi} \cdot \frac{2\pi}{3}\right)$$

$$= \left(\frac{2 \cdot \frac{343}{f}}{3}\right) - \left(\frac{\frac{343}{f}}{3}\right)$$

$$= \left(\frac{2(343)}{3f}\right) - \left(\frac{343}{3f}\right)$$

So:

$$r(f) = \frac{343}{3f}$$

Thus, this function illustrates an inverse proportionality relationship between anti-noise permissible offset range (y-axis in meters) and the frequency of sound (x-axis) with a sale factor of 343/3. This is illustrated on the following graph (see Figure 4) where the x-axis shows the frequency and the y-axis shows the anti-noise permissible offset:

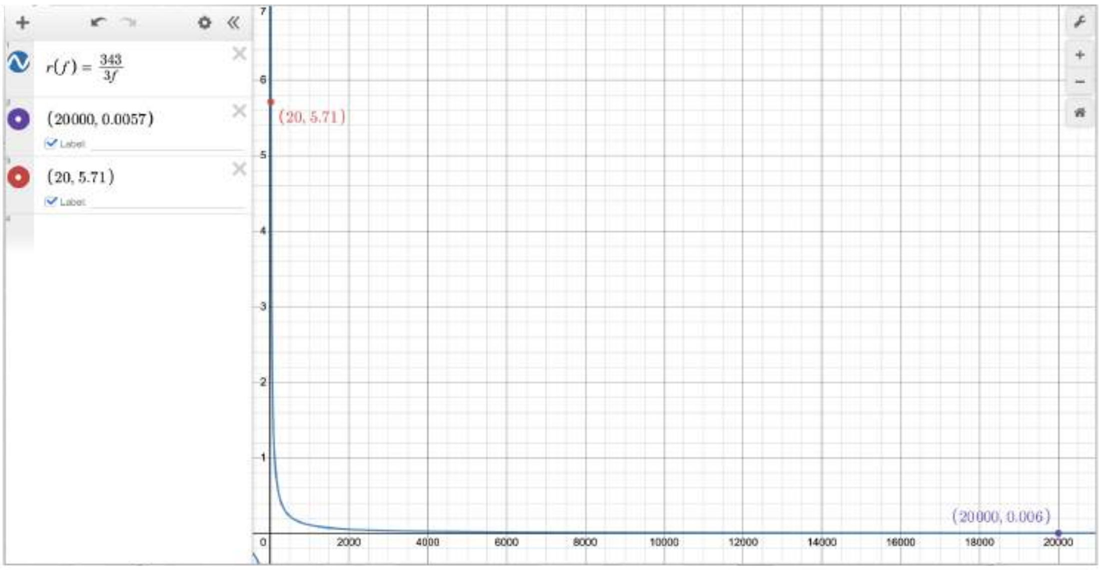


Figure 4: Graph of function r(f) showing inversely proportional relationship between anti-noise permissible offset in meters (on the y-axis) and frequency of sound (on the x-axis)

4.2 Attenuable Frequency Prediction using Group Delay

The function derived showed that permissible offset range is significantly lower at high frequencies, resulting in an ANC system needing better timing accuracy in generating and delivering the anti-noise. This timing accuracy can be measured by a variable known as group delay.

Group delay (GD) refers to the frequency-dependent time delay for an audio signal to reach the system output. In context, it measures the time taken for each audio frequency to physically reach output; thus, it directly affects the ANC system's accuracy in producing anti-noise. The delay is caused by the components of the system like its microphone, speaker, processors, and as a result, is specific to individual audio systems. According to data from rtings.com, an independent platform that provides systematic measurement of consumer headphones, GD in headphones mostly follow the following pattern represented test headphones of this study (see Figure 5):

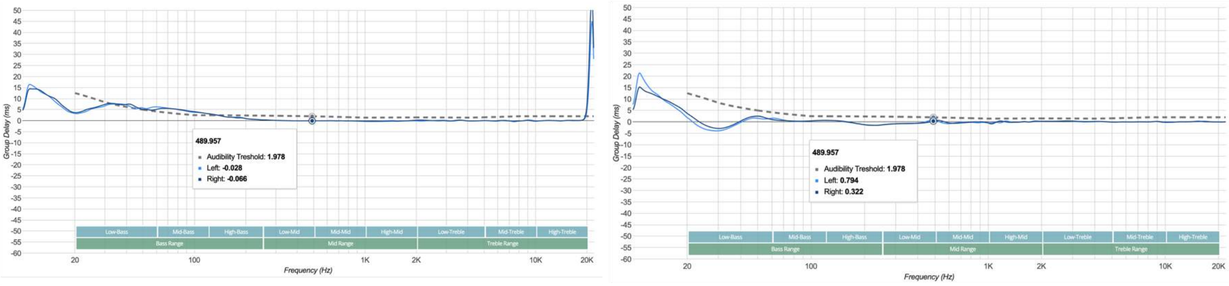


Figure 5: The group delay measurement of the industry-leading Bose QC45s (Kettling et al.) and intermediate Anker Q30s (Cuaig et al.), representing the general headphones group delay trend

The general data showed inconsistent and high GD values in frequencies from 20Hz which gradually diminishes until 490Hz where it becomes relatively linear at under 0.5ms for intermediate headphones and 0.3ms for industry-leading pairs. Owing to this factor, the following calculations for maximum attenuable frequency was calculated using the maximum GD value exhibited above 490Hz.

By multiplying the GD value with 343 ms^{-1} (typical speed of sound in air), the distance that sound waves traveled within the time delay can be calculated. Neglecting other adverse factors, this results in the distance delay that the anti-noise produced would have at maximum - essentially a measure of anti-noise distance accuracy at each audio frequency. Therefore, locating this value on the y-axis of function $r(f)$, which shows the permissible anti-noise distance offset, would correlate to the highest audio frequency that destructive interference could be reliably created.

The calculations using 0.5ms GD value shows a permissible offset of 0.1715m while the GD value of 0.3ms shows 0.1029m. Thus, matching to the x-axis of function $r(f)$ predicts that an intermediate ANC system could attenuate noise up to 666Hz and high-end systems up to 1,029Hz (see Figure 6). In addition, although the GD values for audio under 490Hz are calculated to be sufficient in creating destructive interference reliably, their high inconsistency and unreliability leads to prediction of inconsistent and uneven ANC attenuation at these ranges. Lastly, as the function $r(f)$ has y-value (representing the permissible anti-noise offset) approaching zero, the author also predicts a gradual decrease in performance from 490Hz to the system's upper limit.

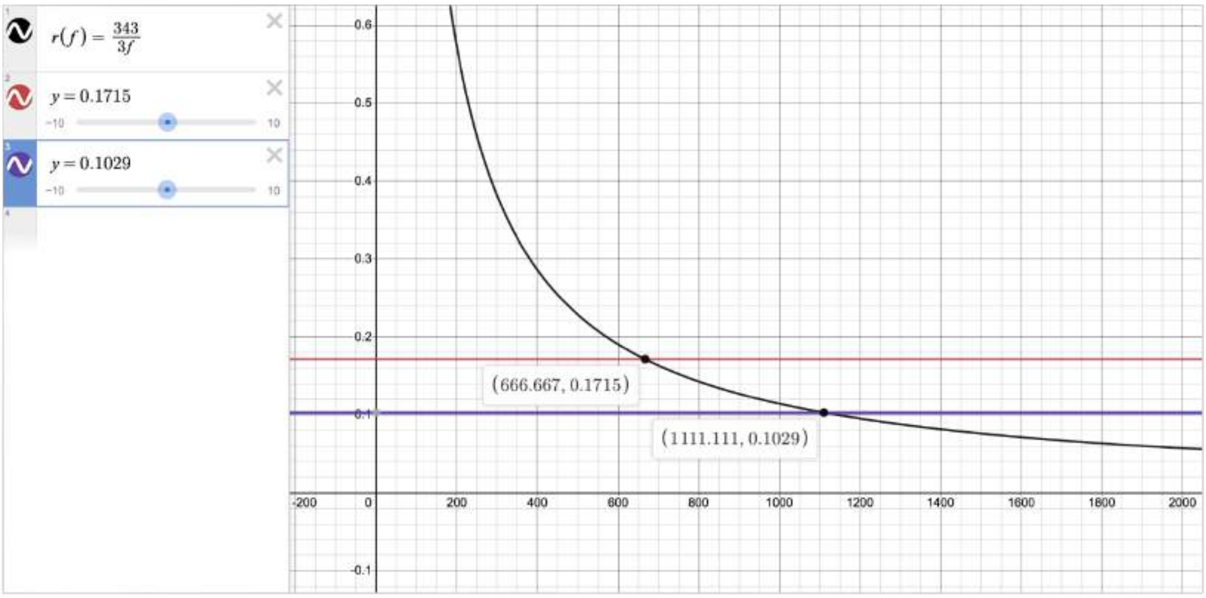


Figure 6: Matching the offset distance offset created by GD (y-axis) to its corresponding frequency (x-axis). Therefore, showing the attenuable frequency due to the system's GD value

5. Theoretical Prediction for Timbre-Dependent Performance

5.1 Timbre of a Sound

Audio timbre refers to the auditory sensation created by a sound wave (Mcadams and Goodchild). Essentially, it is the unique "color" or "flavor" of a sound that makes a note from the piano different from that of a guitar. Mathematically, timbre can be represented by their waveform - a graphical representation for the shape of wave as a function of time (see Figure 7) -, meaning different audio timbre can be seen as unique shapes on a amplitude (vertical-axis) against time (horizontal-axis) graph:

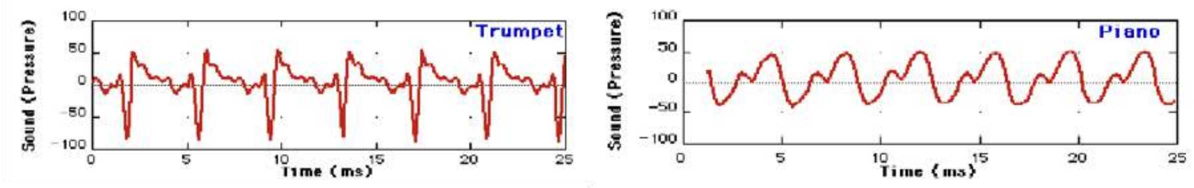


Figure 7: Different waveform of the same musical note (a middle C) on the trumpet and piano (Paiva)

5.2 Effects of Timbre Prediction using Fourier Transform

The simplest type of sound timbre is known as a pure tone, and its waveform - a sine wave - can be modeled by the following general form where A is the amplitude of the wave, ω is the angular frequency equal to $2\pi f$, and ϕ is the phase shift of the wave:

$$f(t) = Asin(\omega t + \phi)$$

It was discovered that any complex waves - sounds that are not sinusoidal - are multiple superimposed pure sine waves of various amplitude, frequency, and phase. Essentially, this indicates that any audio timbre is composed of many pure tones of different amplitudes, frequency, and phase summed together (Carillo). The theorem was first stated by Monsieur Joseph Fourier, and has been developed into a mathematical technique to decompose any signal into its sinusoidal components called the Fourier Analysis (Carillo).

Primarily, there are 2 types of Fourier Analysis: the Fourier Series, applicable for noise that repeats at a regular interval, and the Fourier Transform, which is applicable for all types of noise. The Fourier Analysis also forms the basis of computer algorithms, such as Fast Fourier transform (FFT), for many applications of signal processing including ANC. However, as their purpose is identical, the Fourier Series works as follows:

$$f(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos(nt) + \sum_{n=1}^{\infty} b_n \sin(nt)$$

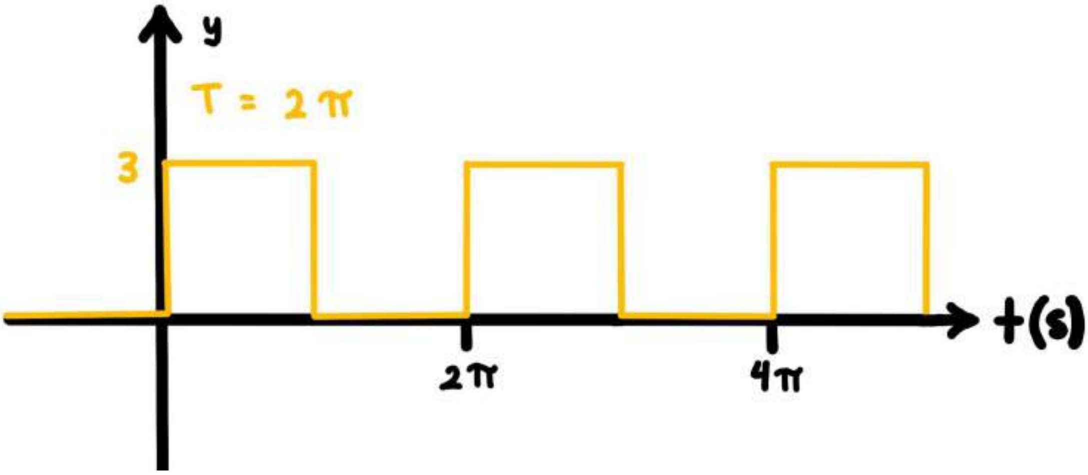
$$a_0 = \frac{1}{2L}\int_{-L}^{L} f(t)\,dt$$

$$a_n = \frac{1}{L}\int_{-L}^{L} f(t)\cos\left(\frac{nt\pi}{L}\right)dt$$

$$b_n = \frac{1}{L}\int_{-L}^{L} f(t)\sin\left(\frac{nt\pi}{L}\right)dt$$

where function $f(t)$ represents a timbre's waveform, L = half the wave's period, and a_0, a_n, and b_n are coefficients to be calculated.

For the following sample square wave (Academy):



1. Calculating for a_0:

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)dt$$

Splitting the integral:

$$= \frac{1}{2\pi}\left(\int_{-\pi}^{0} f(t)dt + \int_{0}^{\pi} f(t)dt\right)$$

Substituting $f(t)$ and integrating:

$$= \frac{1}{2\pi}\left(\int_{-\pi}^{0} 0dt + \int_{0}^{\pi} 3dt\right)$$

$$= \frac{1}{2\pi}(3t)\Big|_{0}^{\pi}$$

So:

$$a_0 = \frac{3\pi}{2\pi} = \frac{3}{2}$$

2. A similar process is done to calculate a_n:

$$a_n = \frac{1}{\pi}\left(\int_{-\pi}^{0} f(t)cos(nt)dt + \int_{0}^{\pi} f(t)cos(nt)dt\right)$$

$$= \frac{1}{\pi}\left(\int_{-\pi}^{0} 0dt \ + \int_{0}^{\pi} 3cos(nt)dt\right)$$

$$= \frac{3}{\pi} \cdot \frac{1}{n}(sin(nt))\Big|_{0}^{\pi}$$

Since $n = Z$,

$$= \frac{3}{n\pi}\left(sin(n\pi) \ - \ sin\left(a_n\right)\right)$$

$$a_n = \frac{3}{n\pi}(0 \ - \ 0) \ = \ 0$$

3. A similar process is done to calculate b_n:

$$b_n = \frac{1}{\pi}\left(\int_{-\pi}^{0} f(t)sin(nt)dt \ + \int_{0}^{\pi} f(t)sin(nt)dt\right)$$

$$= \frac{1}{\pi}\left(\int_{-\pi}^{0} 0\sin(nt)\,dt + \int_{0}^{\pi} 3\sin(nt)\,dt\right)$$

$$= \frac{1}{\pi} \cdot 3 \cdot \frac{1}{n}\cos(nt)\Big|_{0}^{\pi}$$

$$= \frac{-3}{n\pi}(\cos(n\pi) - \cos(0n)$$

$$= \frac{-3}{n\pi}(\cos(n\pi) - 1)$$

When n *is even:*

$$b_n = \frac{-3}{n\pi}(1 \ - \ 1)$$

$$= \ 0$$

When n *is odd:*

$$b_n = \frac{-3}{n\pi}(- \ 1 \ - \ 1)$$

$$= \frac{6}{n\pi}$$

Therefore, substituting:

$$a_0 = \frac{3}{2},$$

$$a_n = \ 0,$$

$$b_n = \ 0 \ (where \ n \ is \ even)$$

$$b_n = \frac{6}{n\pi} \ (where \ n \ is \ odd)$$

The Fourier series for this wave states:

$$f(t) = \frac{3}{2} + \frac{6}{\pi} sin(t) + \frac{6}{3\pi} sin(3t) + \frac{6}{5\pi} sin(5t) + ...$$

Each terms of this series represents each constituent component, and by comparing to the general form of a pure tone, $f(t) = Asin(\omega t + \phi)$, the amplitude, phase, and frequency of individual wave can be determined:

For the first 4 terms of the series, since $\omega = 2\pi f$:

2nd Term:

$$f(t) = \frac{6}{5\pi} sin(t)$$
$$2\pi(f) = 1$$
$$f = \frac{1}{2\pi} \approx 1.5Hz$$

3rd term:

$$f(t) = \frac{6}{5\pi} sin(3t)$$
$$2\pi(f) = 3$$
$$f = \frac{3}{2\pi} \approx 5Hz$$

4th term:

$$f(t) = \frac{6}{5\pi} sin(5t)$$
$$2\pi(f) = 5$$
$$f = \frac{5}{2\pi} \approx 8Hz$$

Thus, this particular square wave timbre is just pure tones of: 1.5Hz, 5Hz, 8Hz, and more combined (see Figure 8).

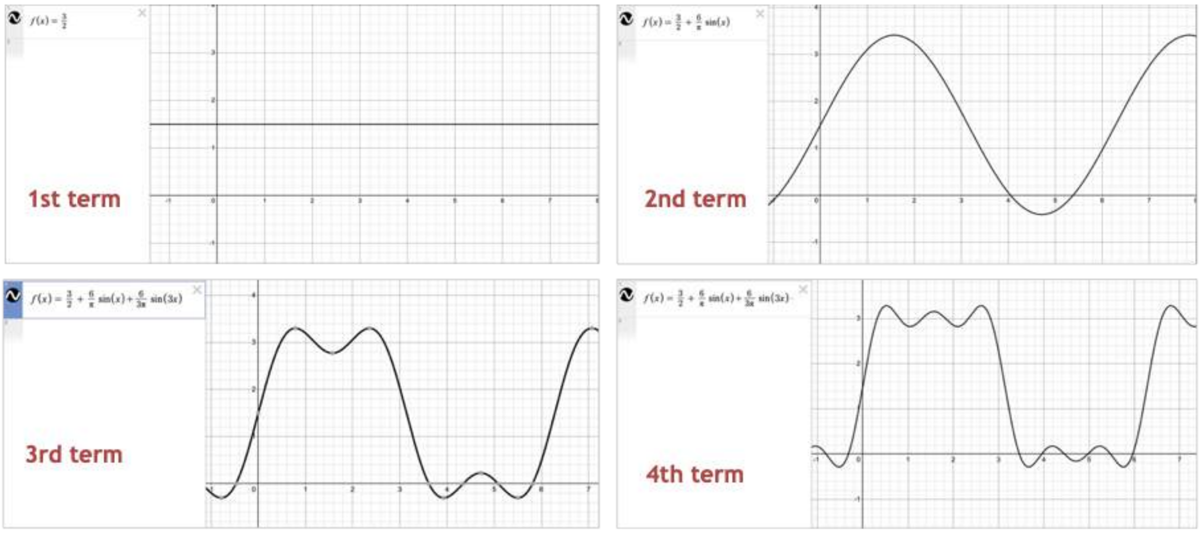


Figure 8: Combining the first 4 terms of the series

As timbres are just numerous sinusoidal frequencies, there should be no inherent physics limitations causing variations in ANC performance. The FFT processes all noise timbre into sinusoidal components, making the physical process of attenuating them practically the same. Hence, given no adverse engineering constraints, ANC should perform similarly and align with anticipated frequency trend across all noise timbres. However, certain engineering constraints may influence the outcome, as outlined below:

1. Rapidly changing (i.e. human speech) and high-impulse noise (i.e. gunshots) are too erratic and unpredictable for ANC systems to process reliably (Li and Yu). This results in reduced attenuation performance or anti-noise misalignment which could result in accidental amplification of noise.

2. The angle of incoming sound, the reflection and diffraction property of different frequencies, and the imperfect frequency response of the microphones and speakers results in amplification or reduction of particular frequencies. Thus, this distorts the input data which has to be compensated with software - a calibration process that introduces inaccuracies to the system.

6. Empirical Experiment and Collected Data

6.*1 Experiment Setup*

The subsequent experiment was designed to measure ANC audio attenuation across the frequency spectrum and at different noise timbres.

Apparatus comprises of an empty room (4 m x 2.5 m x 2.4 m), two pairs of hybrid-ANC over-ear headphones (Bose QC45s denoting industry-leading system & Anker Q30s denoting an intermediate system), a headphone measurement rig (MiniDSP EARs), a full-range speaker (Roland KC 350 Amplifier), and a computer running Room EQ Wizard software (REW) to generate test signals and collect data. The room is treated with sound diffusers on the walls and ceiling to reduce the echo, reverberation, and standing waves (Dominic). Thus, it allows for a balanced and natural sound field representing a realistic operating environment. Furthermore, the test headphones both use a hybrid-ANC system with microphones both inside and outside the ear cups to limit engineering constraints of feedforward (microphone outside) and feedback (microphone inside) systems. These limitations include: inability to compensate for inconsistent fit and an insufficient system response time. Additionally, the MiniDSP EARs are a headphones measurement rig that simulates the acoustic properties of a human head (see Appendix 1). Their microphones, positioned inside silicone ears, are calibrated to be neutral across all audio frequencies. This setup allows for recordings and measurements of noise level inside the headphones ear cup.

The EARs are located in the room-center and elevated 1.5 meters above the floor to achieve a balanced mix of direct and diffused sound. The full-range speaker system is placed at one end of the room, 2 meters away facing the EARs, and is elevated to the same height.

Both the EARs and the speaker are connected to REW which generates test signals and measure noise data inside the headphones ear cups (see Figure 9).

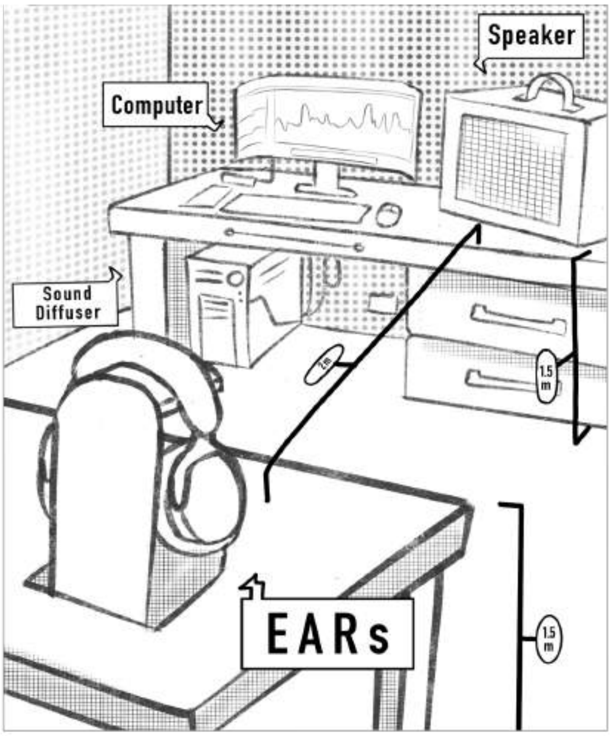


Figure 9: Illustration of the experiment setup

6.2 Experiment Procedures

For each test signal, three measurements are conducted resulting in three frequency spectrum graph of that signal. The measurements collected are: the reference measurement, the ANC-enabled measurement, and the ANC-disabled measurement. The reference measurement is the audio "heard" by the EARs without any headphones; thus, it shows the normal audio level and serves as a reference point to compare the following two measurements. The second and third measurements are both measured from the EARs "wearing" a pair of ANC headphones. The second is measured with ANC-disabled to represent audio level of passive attenuation, and the third is measured with ANC-enabled to represent audio level with headphones and ANC on. As such, the change in audio level between the second and third measurements denotes the attenuation performance of ANC. Furthermore, the noise level is collected using the Real Time Analyzer (RTA) tool which decomposes the audio signal heard into individual frequencies and their corresponding amplitude. By plotting the results on an Amplitude (db) - Frequency (Hz) graph (called a frequency response curve), the volume difference between each audio frequency can be presented. Therefore, the data shows how the effectiveness of ANC is influenced by the audio frequency.

To measure the effect noise timbre, the measurements mentioned above are repeated for the six specifically chosen test signals including: a pure tone sweep, pink noise, aircraft noise, and speech chatter noise. The pure tone sweep is a sinusoidal wave at a single frequency (a narrowband noise) that increases over time; thus, it simulates the simplest condition of noise. Alternatively, pink noise contains overlapping waves covering the entire human hearing spectrum at equal intensity (loudness); as such, it simulates the most complex noise timbre. Furthermore, aircraft cabin noise represents a predictable low-impulse noise optimal for ANC while speech chatter represents the irregular high-impulse noise challenging for ANC.

For each test signal, the measurements are taken using the following procedure: a computer generates a test signal through the full range speaker, the MiniDSP EARs record the audio signal, and the RTA decomposes the signal to obtain the frequency response curves. Furthermore, for each measurement, the average of 3 trials are taken with re-seating of the headphones in-between to reduce data alterations due to fit and placement of the headphones on the EARs. The same procedure is then repeated for 2 pairs of ANC headphones, one denoting an intermediate system while the other one denoting an industry-leading system, to provide an inclusive view of different ANC quality. This results in three groups of three frequency response curves for each test signal (for one pair of headphones), and the average of each group will be taken for evaluation.

6.3 Collected Data

The collected data of all four test signals can be plotted onto the following eight Sound Pressure Level (SPL) - Frequency graphs. The averaged and smoothened lines are color-coded as: reference measurement in teal, passive isolation in orange, ANC in white.

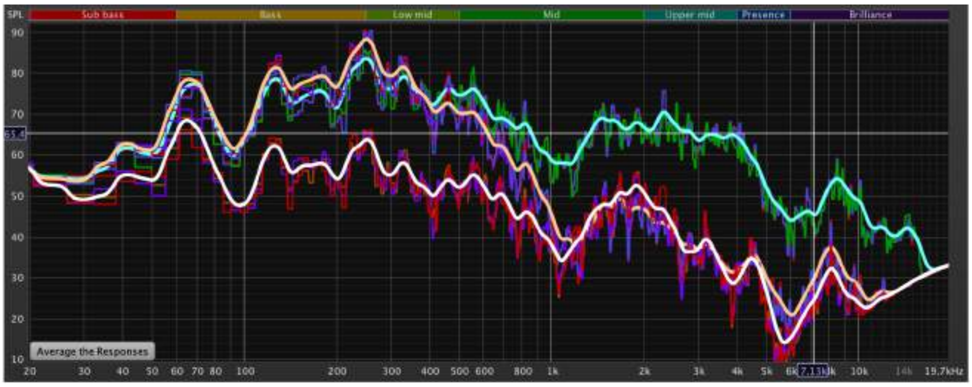
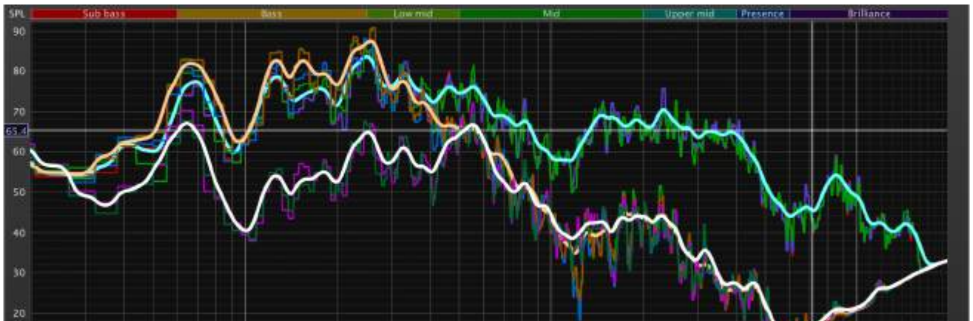


Figure 10: Graph Speech Chatter Noise | Industry-leading ANC system



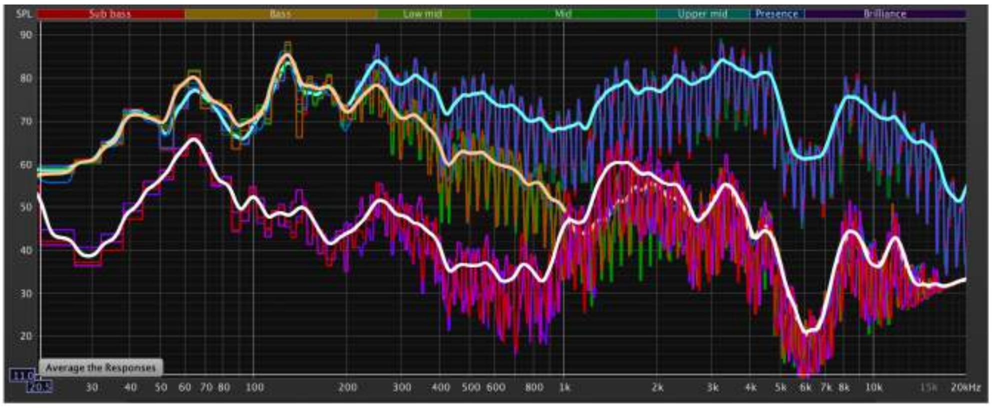


Figure 12: Graph of Sine Sweep Noise | Industry-leading ANC system

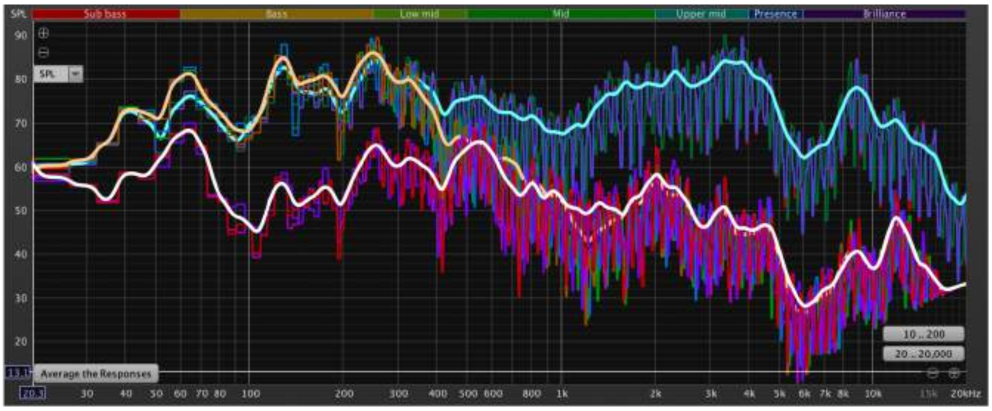


Figure 13: Graph of Sine Sweep Noise | Intermediate ANC system

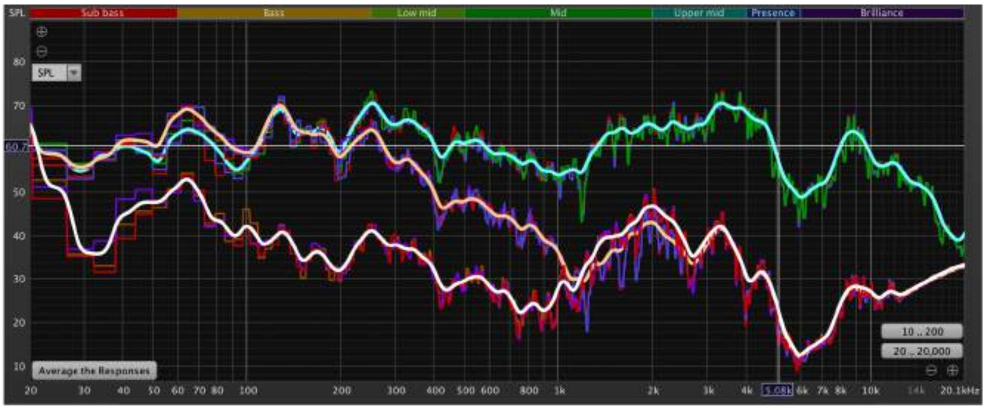
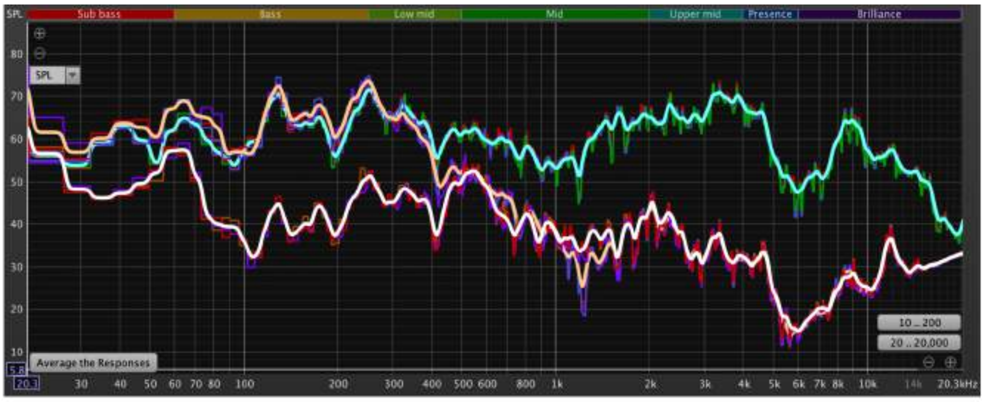


Figure 14: Graph of Pink Noise | Industry-leading ANC system



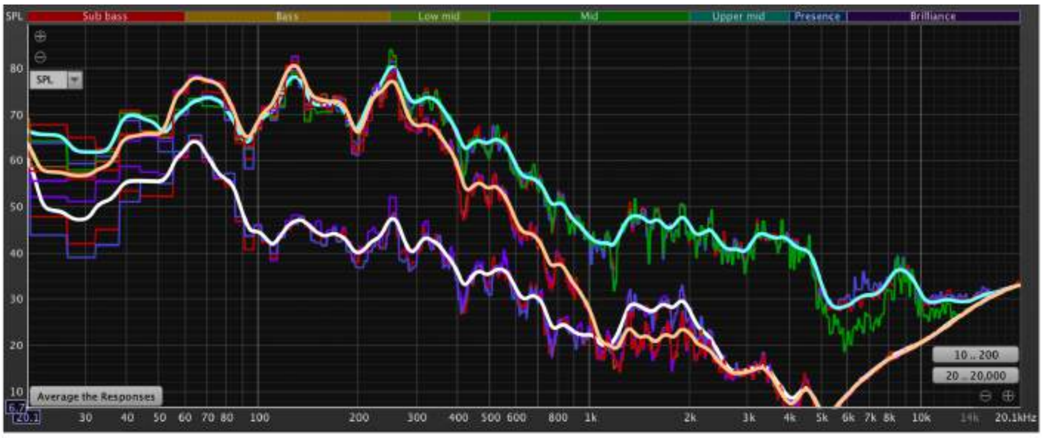


Figure 16: Graph of Airplane Cabin Noise | Industry-leading ANC system

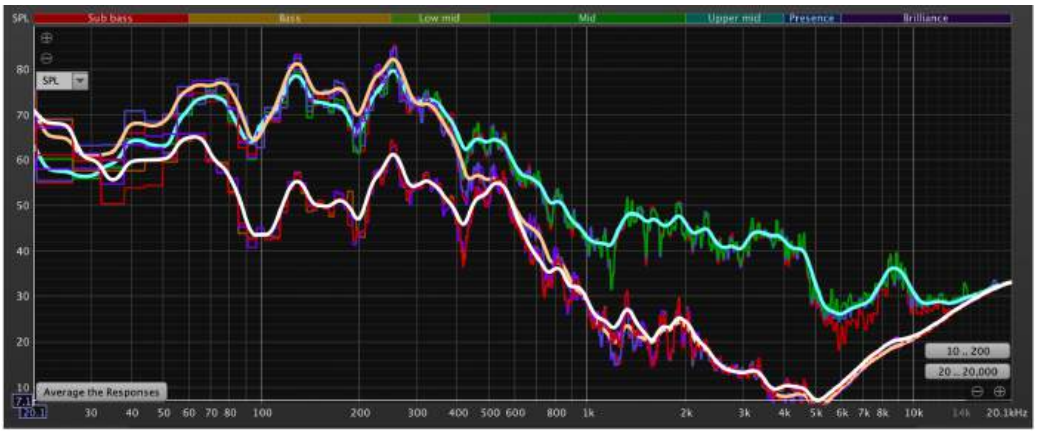


Figure 17: Graph of Airplane Cabin Noise | Intermediate ANC system

7. Discussion & Evaluation of Results

7.1 Processing of Experiment Results

To calculate the ANC attenuation for each of the graph shown, ANC noise level is subtracted from the passive isolation noise level, such as:

Frequency - Hz	Passive Isolation SPL - dB	ANC SPL - dB	*ΔSPL* *(Passive isolation minus ANC) - dB*
19.98148	58.44	61,312	*-2,872*
20.126274	58.62	61,105	*-2,485*
20.272118	58,868	60,787	*-1,919*
20.419018	59,168	60,337	*-1,169*
20.566982	59,479	59,777	*-298*

20.716019	59,742	59,201	*541*

Table 1: Intermediate ANC system raw data sample for Sine Sweep Noise

By repeating the same process for all test tones, the attenuation data for each pair of headphones can be shown, such as:

Frequency - Hz	Δ SPL Sweep - dB	Δ SPL Pink - dB	Δ SPL Plane - dB	Δ SPL Speech - dB
19.98148	-2,872	-3,452	-166	-3,053
20.126274	-2,485	-3,307	-255	-2,995
20.272118	-1,919	-3,128	-365	-2,925
20.419018	-1,169	-2,912	-0.5	-2,841
20.566982	-298	-2,659	-661	-2,746
20.716019	541	-2,372	-846	-2,641
20.866135	1,237	-2,058	-1,052	-2,529

Table 2: Intermediate ANC system processed data sample for all test noise

Frequency - Hz	Δ SPL Sweep - dB	Δ SPL Pink - dB	Δ SPL Plane - dB	Δ SPL Speech - dB
19.98148	4,255	-2,398	3,308	-0.175
20.126274	4,629	-2,284	3,369	-0.101
20.272118	5,076	-2,137	3,447	-0.013
20.419018	5,602	-1,947	3,549	0.093
20.566982	6,202	-1,708	3.68	0.214
20.716019	6,868	-1,409	3,843	0.348
20.866135	7,585	-1,047	4,043	0.491

Table 3: Industry-leading ANC system processed data sample for all test noise

7.2 Summary of Experiment Results

By plotting the data table above, ANC attenuation of each ANC systems for all timbre is shown on the following graphs:

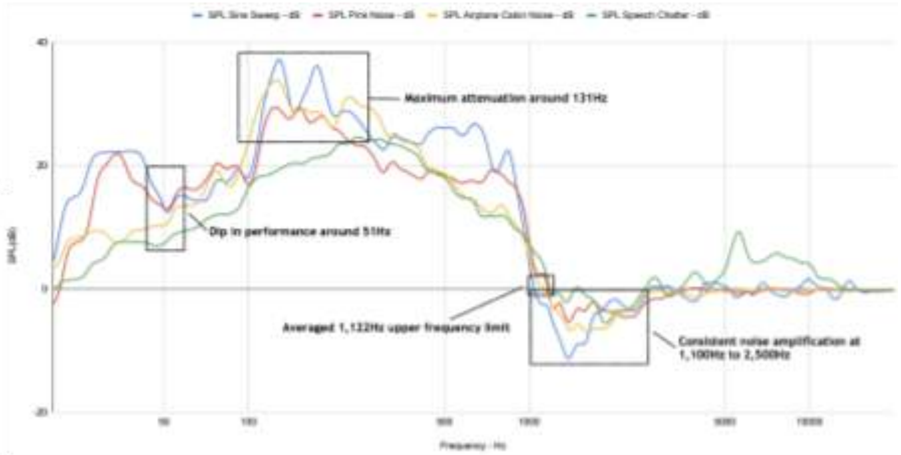


Figure 19: Annotated ANC attenuation - Frequency graph of the industry-leading system (from table 3)

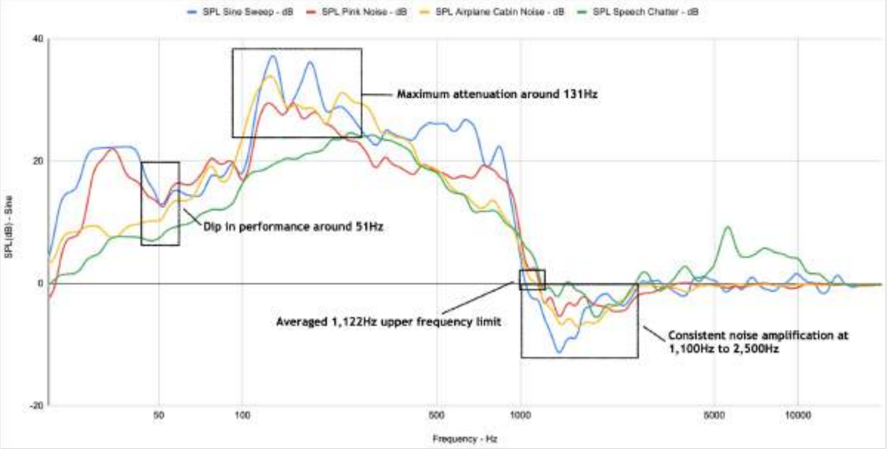


Figure 19: Annotated ANC attenuation – Frequency graph of the industry-leading system (from table 3)

For all test signals, the industry-leading system (see Figure 18) effectively attenuated noise up to around 1,122Hz while the intermediate system (see Figure 17) attenuated noise up to around 550 Hz. Both systems exhibit a similar pattern of: peak attenuation at 131Hz, a performance dip around 51Hz, and gradual performance reduction approaching the system's maximum and minimum. Furthermore, a consistent amplification of noise at 1,100Hz to 2,500Hz was observed in both systems. As amplification is caused by misalignment of anti-noise, it suggests the limit of ANC lies at 1,100Hz with current level of timing accuracy.

Analyzing test timbres, the trends observed in frequency analysis are largely present; however the attenuation level varies to a perceivable extent and there are major discrepancies between the systems. For the intermediate system, attenuation of pink noise (most complex timbre) performed similarly to that of sine sweep (simplest timbre); however, it is marginally lower on the industry-leading system. Furthermore, attenuation of speech chatter, a high-impulse noise challenging for ANC, is suboptimal across all frequencies on industry-leading systems, but is highly effective on the intermediate system. Lastly, attenuation of airplane noise, a low-impulse noise optimal for ANC, is effective at most frequencies which is the only point of agreement for both systems .

7.3 Verifying Theoretical Predictions with Experiment Results

For frequency-dependent performance, it was hypothesized based on GD characteristics that an industry-leading ANC system could attenuate noise up to ~1,029Hz, while an intermediate system could attenuate up to ~666Hz. This is largely supported by the experiment where the Bose Qc45s reached an average of 1,122Hz and the Anker Q30s reached an average of 790Hz. The slightly lower predicted value is likely due to GD being assumed as constant starting from 490Hz while actual GDP tends to decrease in a concave trajectory. Regardless, as there is strong correlation between GD characteristics with the maximum attenuated frequency, it suggests GD as a primary constrain at higher frequencies and validates GD-based prediction method for maximum attenuated frequency.

However, various features observed such as gradual decrease in performance when approaching system minimum, a performance dip around 51Hz, and peak attenuation at 131Hz were not predicted; thus, it suggests that GD does not influence attenuation levels under 131Hz and invalidates any other GD-based prediction aside from maximum attenuable frequency mentioned above.

For timbre-dependent performance, the predictions made are supported to a small extent and there are crucial discrepancies. Timbre is predicted to not affect ANC physically but engineering limitations would be involved. Specifically, this suggests pink noise would perform similarly to the sine sweep, and airplane noise should perform better than speech chatter; however, these features are only exhibited on one system at a time while the other system mostly exhibited contrary results. As such, inconsistent findings could indicate human error or engineering limitations beyond initial assumptions. Contradicting experiment results strongly suggests the effects of timbre are likely due to engineering limitations which are specific to individual manufacturers. Thus, determining their specific cause and effect yields inconclusive results at current level.

7.4 Comparative Analysis with Prior Studies

ANC effectiveness has previously been measured. Notably, Rtings.com (Kettling et al.; Cuaig et al.) provides systematically measured data using methodology comparable to this study. However, no measurements found assesses different timbres, so only sine sweep tone is compared.

Rtings.com's data shares general trends with our study; however, the specific features are indistinct and only broad observations are observed. Specifically, Rtings.com's findings confirms the upper and lower frequency limit of ANC, and the gradual attenuation decrease as the system approaches its upper and lower limits. Thus, validating this studies' GD-based method in predicting maximum attenuable frequency. However, features such as maximum attenuation at 131Hz and a sudden drop in performance at 51Hz were either unclear or absent despite the identical methodology involved. This lack of correlations may be attributed to excessive smoothing and different measuring devices as graphical shifts cause features to be obscured when averaged.

8. Predictive Modeling of Frequency-Dependent Performance

8.1 Formulating a Function using the Fourier Series

The data collected and the validated maximum attenuable frequency prediction using GD can be integrated to formulate a mathematical function modeling ANC performance. A function as such allows for theoretical predictions of maximum attenuable frequency while also illustrating empirical characteristics; thus, resulting in a framework that benefits future studies and technological advancements. The function is modeled based on the pink noise performance of industry-leading Bose QC45s up to 1,400Hz as they would represent an ideal benchmark that can be scaled down to represent less sophisticated systems.

The author opted to apply the Fourier Series demonstrated prior as it could replicate data with resolution beyond typical functions, enabling the model to demonstrate both trends and specific features such as peaks and dips. However, to reduce the number of terms in the series, the sampling rate was scaled down from 5,521 data points to 98, taking one measurement every 12.5Hz from 20Hz up to 1,400Hz (see Table 4).

Frequency - Hz	Passive isolation SPL - dB	ANC SPL - dB	ΔSPL (Passive isolation minus ANC) - dB
20	70,342	72,957	-2,615
32.5	60,217	38,694	21,523
45	63.46	49.16	14.3
57.5	67,952	51,271	16,681

Table 4: Sample of data used to calculate for the Fourier constants to model Frequency - ΔSPL graph

Owing to the extensive nature of data, manual calculations of the Fourier Series coefficients as previously demonstrated is no longer feasible, so a computer program was used ("Fourier Analysis of Real Data Sets"). As such, the input values and calculated coefficients is represented on the following table:

Frequency - Hz	ΔSPL (Passive isolation minus ANC) - dB	n	c_n	s_n	Δt
20	-2,615	0	15.905082	-1.26E-14	*12.5s (for all value)*
32.5	21,523	1	-3.7444838	5.396902	
45	14.3	2	-2.2598035	6.2530107	
57.5	16,681	3	-1.0802796	3.1917983	
70	17,858	4	-1.7002484	1.0168825	
82.5	19,025	5	-1.244693	-0.11164409	
95	19,552	6	-1.6754339	0.29108381	

Table 5: Sample of input values (Frequency and ΔSPL) and calculated Fourier constants

Therefore, substituting these values into the Fourier series equation:

$$x(t) = \sum_{n=0}^{n<N/2} \left[c_n \cos\left(\frac{2\pi n t}{N\Delta t}\right) + s_n \sin\left(\frac{2\pi n t}{N\Delta t}\right) \right]$$

results in the following 98 terms function that models the general industry-leading ANC performance:

$$f(t) = 11.849694 \cdot \cos(0) + (-9.7773029 \cdot 10^{-15}) \cdot \sin(0) + (-4.6526909) \cdot \cos(0.25132741(t-21)) + \ldots + 0.052020779 \cdot \sin(15.079643(t-21)) \,|\, 15 < t < 1413$$

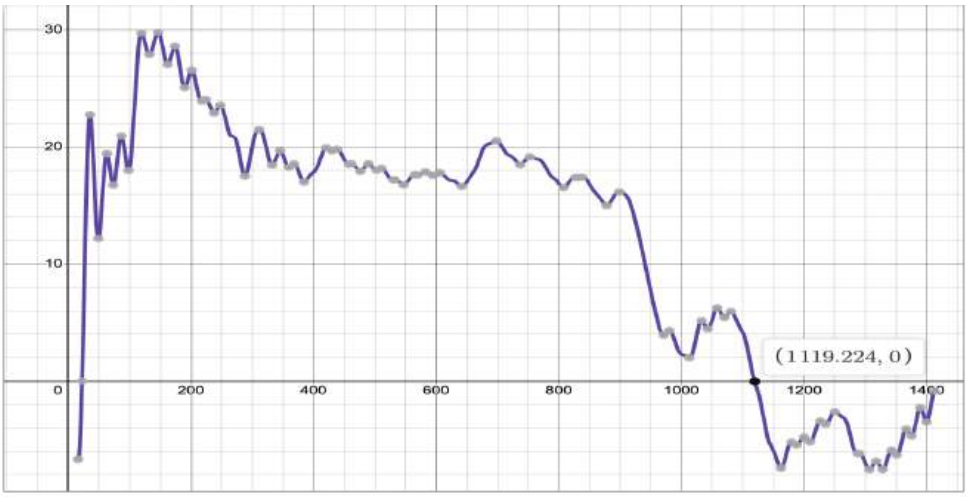


Figure 20: Graph of function f(t) limited to $15 < t < 1413$ showing all graphical features and trends

8.2 Creating a Predictive Function based on Group Delay

To enable the function to predict maximum attenuable frequency based on GD as a variable, the function developed can be multiplied to logarithmic function $g(t)$ to enable movement of its right-side x-intercept based on GD input:

$$g(t) = d \cdot log\left(- t + \left(\frac{u}{w^r}\right)\right)$$

where w is the headphone's highest GD value above 490Hz in milliseconds, and u, d, r are calibrating constants. Through trial-and-error using experiment data from the secondary (intermediate) ANC system and the proposed method of predicting maximum attenuated frequency, the constants are calculated to be $d = \frac{337}{100}$, $u = 340$, and $r = \frac{97}{100}$. Thus, effectively resulting the following final function capable of predicting specific ANC system performance based just on their maximum GD value above 490Hz:

$$g(t) = \frac{337}{1000} \cdot \log\left(-t + \left(\frac{340}{w^{\frac{97}{100}}}\right)\right) \cdot [11.84 \cdot \cos(0) + (-9.77 \cdot 10^{-15}) \cdot \sin(0) + (-4.65) \cdot \cos(0.25(t-21)) + \ldots + 0.05 \cdot \sin(15.07(t-21))] \,|\, 15 < t < 1413$$

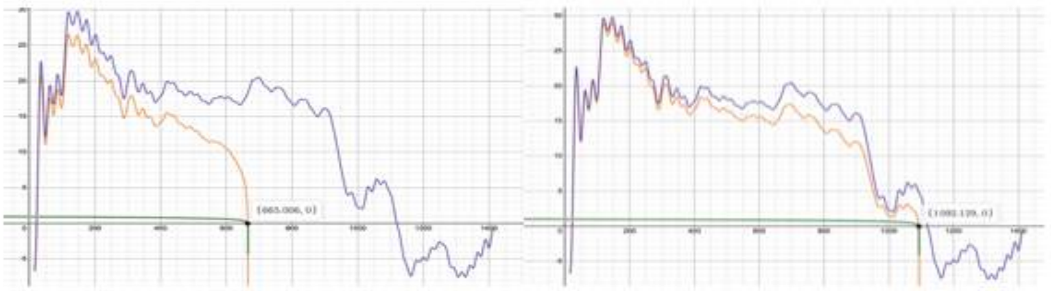


Figure 21: Graphs showing orange function g(t) predicting maximum attenuable frequency with GD value of 0.5ms and 0.3 ms with satisfactory accuracy. Original function f(t) is in purple and logarithmic function used to control x-intercept is in green.

However, this method is limited to only predict ANC over-ear headphones having GD values greater than 0.3ms. As 0.3ms was the lowest value (giving highest performance) measured in the experiment, this study lacks empirical data beyond this level. It was worth considering stretching the current function using a scale factor since maximum attenuated frequency could be predicted using theoretical data; however, it altered features such as maximums and minimum beyond desired outcome. Therefore, w must be limited to be equal to or above 0.3ms.

9. Conclusion

This study investigated how timbre and frequency of noise affect the performance of ANC. Concerning frequency, empirical data revealed that ANC is most effective at approximately 131Hz, abruptly ineffective at 51Hz, and gradually declines towards upper and lower limits. While the lower limit of ANC appears to be around 20Hz, it was not theoretically predicted nor explicitly verified by empirical data. Furthermore, a method for predicting frequency-dependent performance, including a system's upper limit, based on its group delay characteristics was proposed and largely supported by experimental data. However, as only upper limit predictions matched, this suggest GD is a constraining factor for maximum attenuable frequency, but is irrelevant at frequencies below 131Hz.

Concerning timbre, the study tested six noise timbres on two pairs of ANC headphones, one industry-leading and one of intermediate quality. However, there were major data contradictions between the two headphones, leading to the inability to verify theoretical predictions. This suggests a high influence of manufacturer-specific engineering limitations beyond initial assumption

Despite the challenges, all successful deductions from the study were combined to create a mathematical function that serves as a framework contributing to future studies and technological advancements. The resulting function is a 98 terms Fourier Series with GD as a variable that models frequency-based performance and predicts maximum attenuable frequency using the proposed GD prediction method. However, this method is limited to only predict GD values of equal or above 0.3ms as this study lacked empirical data to evaluate performance beyond this threshold.

10. Bibliography

Academy, Khan . “Signals and Systems | Electrical Engineering | Science.” *Khan Academy*, www.khanacademy.org/science/electrical-engineering/ee-signals.

Benoit, Michael, et al. *Engineering Silence: Active Noise Cancellation*. Oct. 2012, rsmith.math.ncsu.edu/MA574_S19/REFERENCES/silence.pdf. Accessed 27 Aug. 2023.

Carillo, Sandra. “Fourier Series.” *Encyclopedia of Thermal Stresses*, Jan. 2014, pp. 1736–1742, https://doi.org/10.1007/978-94-007-2739-7_23. Accessed 18 Nov. 2023.

CNX, OpenStax . “16.5 Interference of Waves | University Physics Volume 1.” *Courses.lumenlearning.com*, courses.lumenlearning.com/suny-osuniversityphysics/chapter/16-5-interference-of-waves/.

Comeau, Josh. “Let’s Learn about Waveforms.” *The Pudding*, 2019, pudding.cool/2018/02/waveforms/.

Cuaig, Vanessa, et al. “Anker Soundcore Life Q30 Wireless Review.” *RTINGS.com*, 10 Oct. 2023, www.rtings.com/headphones/reviews/anker/soundcore-life-q30-wireless. Accessed 20 Nov. 2023.

David L. Bowen. “Active Noise Control to the Rescue?” *Acentech*, 30 Oct. 2018, www.acentech.com/resources/active-noise-control-to-the-rescue/. Accessed 7 May 2023.

Dhar, Payal. “Noise-Canceling Headphones without the Headphones - IEEE Spectrum.” *Spectrum.ieee.org*, 21 Dec. 2020, spectrum.ieee.org/active-noise-cancellation-using-ldvs.

Dominic, Author. “Sound Diffuser vs. Absorber - What’s the Difference?” *Sound Proof Central*, 10 Feb. 2021, soundproofcentral.com/sound-diffuser-vs-absorber/.

Elliot, Stephen. *Signal Processing for Active Control*. San Diego, Calif. ; London, Academic, 26 Sept. 2000, books.google.com.vn/books?hl=en&lr=&id=GklDOTI6ZLIC&oi=fnd&pg=PP1&dq=Signal+Processing+for+Active+Control&ots=mo7Gaw9Iw5&sig=RecfO2Rnx01MoF9iPdLUfaheSuA&redir_esc=y#v=onepage&q=Signal%20Processing%20for%20Active%20Control&f=false. Accessed 14 Nov. 2023.

“Fourier Analysis of Real Data Sets.” *Lampx.tugraz.at*, lampx.tugraz.at/~hadley/num/ch3/3.3a.php. Accessed 18 Nov. 2023.

Kaymak, Erkan, and Mark Atherton. *Active Noise Control at High Frequencies Blood Flow Modelling View Project Patent Knowledge Design Tool (PKDT) View Project*. July 2006.

Kettling, Theresa , et al. “Bose QuietComfort 45/QC45 Wireless Review.” *RTINGS.com*, 21 June 2023, www.rtings.com/headphones/reviews/bose/quietcomfort-45-qc45-wireless.

Li, Peng, and Xun Yu. “Active Noise Cancellation Algorithms for Impulsive Noise.” *Mechanical Systems and Signal Processing*, vol. 36, no. 2, Apr. 2013, pp. 630–635, https://doi.org/10.1016/j.ymssp.2012.10.017. Accessed 18 Nov. 2023.

Mcadams, Stephen, and Meghan Goodchild. *MUSICAL STRUCTURE Sound and Timbre*. 2017.

Milosevic, Aleksandar, and Urs Schaufelberger. *Active Noise Control*. 14 Dec. 2005.

Paiva, Rui. *Content-Based Classification and Retrieval of Music: Overview and Research Trends “Content-Based Classification and Retrieval of Music: Overview and Research Trends.”* Oct. 2002.

Russell, Daniel A. "Superposition of Waves." *Psu.edu*, 2014, www.acs.psu.edu/drussell/Demos/superposition/superposition.html.

Alternate Treatment for Autistic Children using Generative AI with Virtual Reality

By:
Sashvathkumar Krishnakumar
11th Grade, Cox Mill High School, North Carolina

2024

Mentored by:
Dr. Bross, UNC Charlotte, NC
Mrs. Kellie Pierrot, Cox Mill High School, NC

Alternate Treatment for Autistic Children using Generative AI with Virtual Reality

Abstract

This paper delves into the expanding field of virtual reality (VR) with Generative AI as a therapeutic intervention for individuals with Autism Spectrum Disorder (ASD). It reviews emerging research demonstrating the efficacy of VR-based programs designed to address core challenges associated with ASD such as social skills deficits, communication difficulties, anxiety, and sensory sensitivities.

The current research focused on a Personal Avatar which introduced 2 autistic children to a virtual storyteller with whom the children had a one-on-one experience with. The storyteller was preloaded with stories generated by a Generative AI about fictional characters. While the avatar was narrating the story to the child, the number of times the child lost eye contact with the avatar and the duration of attention were measured through eye-tracking capabilities in the VR headset. This data was collected over 7 sessions conducted every other day for one child and over 4 sessions for another child.

After analyzing the results, a 47% improvement in eye contact and 32% improvement in the focus of the first child with a 45% improvement in eye contact and 7% improvement in focus in the second child was observed. This led to the conclusion that with the use of technologies such as Generative AI and Virtual Reality, ASD symptoms can be alleviated in a safer, economical and nimble way. The Social Avatar (stage 2) will introduce the child to a Metaverse Autistic Community (MAC).

Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by persistent deficits in social communication and interaction, as well as restricted interests and repetitive patterns of behavior, interests, or activities. Here's a breakdown of its core characteristics:

1. **Social Communication Impairments**: Individuals with ASD often struggle with social interactions and communication skills. This can manifest as difficulties in understanding and using nonverbal cues such as eye contact, facial expressions, and body language. They may also have challenges in initiating and maintaining conversations, understanding social norms, and developing friendships.
2. **Restricted Interests**: People with ASD often display intense interests in specific topics or activities. These interests can be highly focused and may dominate their thoughts and conversations. They might spend a significant amount of time learning intricate details about their interests, sometimes to the exclusion of other activities.
3. **Repetitive Behaviors**: Repetitive behaviors are a hallmark feature of ASD. These can include repetitive movements such as hand flapping or body rocking, insistence on

sameness or routines, and highly specific rituals or routines. Individuals with ASD may become distressed or upset when these routines are disrupted.

Prevalence rates of ASD have been increasing over the past few decades, with recent estimates indicating that around 1 in 54 children in the United States are diagnosed with ASD. However, it's essential to note that ASD traits can vary significantly among individuals. This variability can manifest in several ways:

1. **Severity of Symptoms**: ASD is a spectrum disorder, meaning that individuals can experience a wide range of symptoms, from mild to severe. Some individuals may have relatively mild challenges and may be able to function well in mainstream society with appropriate support, while others may require more intensive interventions and support services.
2. **Co-occurring Conditions**: Many individuals with ASD have co-occurring conditions such as intellectual disabilities, epilepsy, anxiety disorders, or ADHD. The presence of these additional conditions can further influence the expression of ASD traits and impact an individual's overall functioning and quality of life.
3. **Strengths and Abilities**: While individuals with ASD often face challenges in certain areas, they may also possess unique strengths and abilities. Some individuals with ASD have exceptional talents in areas such as music, art, mathematics, or computer programming. Recognizing and nurturing these strengths can be instrumental in supporting individuals with ASD to thrive.
4. **Developmental Trajectory**: The expression of ASD traits can evolve over time as individuals grow and develop. Early intervention and appropriate support services can play a crucial role in helping individuals with ASD to develop skills, cope with challenges, and achieve their full potential.

In summary, ASD is a complex and heterogeneous neurodevelopmental disorder characterized by deficits in social communication, restricted interests, and repetitive behaviors. While there are common core characteristics, the expression of ASD traits can vary widely among individuals, highlighting the importance of individualized support and interventions.

Traditional treatment approaches for Autism Spectrum Disorder (ASD) typically involve a combination of behavioral, educational, and therapeutic interventions.
These traditional treatment approaches are often implemented in a multidisciplinary and individualized manner, taking into account the unique strengths, challenges, and needs of each individual with ASD. Early intervention and ongoing support are key components of effective treatment for ASD.

Virtual Reality (VR) is an immersive technology that simulates a realistic three-dimensional environment, allowing users to interact with and explore virtual worlds. At its core, VR works by creating a computer-generated environment that is presented to the user via specialized hardware, such as headsets or goggles, and input devices, such as motion controllers or gloves.

- Controlled and safe environment where social and behavioral skills can be practiced repeatedly without real-world consequences.

- Customizable simulations that can cater to the specific needs and sensitivities of individuals with ASD.
- VR's ability to "gamify" learning to increase engagement and motivation.
- Objective progress tracking capabilities enabled by the VR environment.

Generative AI represents a transformative paradigm within artificial intelligence research, enabling machines to create novel content autonomously. At its core, generative AI encompasses a diverse set of algorithms and models designed to generate data, images, text, audio, or other types of content that closely resemble human-created output. Generative AI has demonstrated remarkable capabilities across various domains, including image generation, natural language processing, music composition, and even drug discovery. Generative AI can work interactively by leveraging real-time feedback loops between the user and the AI system, enabling dynamic collaboration and co-creation. Generative AI has the potential to provide several benefits to autistic individuals by addressing various challenges they may face and enhancing their quality of life in different ways. Overall, generative AI holds great promise for improving the lives of individuals with autism spectrum disorder by addressing their unique needs and challenges, enhancing communication, social skills, sensory integration, education, and daily functioning.

Ultimately, Generative AI combined with VR might become a more accessible, less time-consuming, and more effective treatment than currently available solutions. This research paper intends to explore the potential role of these technologies in the treatment of Autism compared to the conventional therapy that is commonly used today.

Virtual reality (VR) environments powered by generative AI can create simulated social scenarios where individuals with ASD can practice and develop their social skills in a safe and controlled setting. These VR simulations can simulate real-world social interactions, such as conversations, group activities, and nonverbal communication cues, allowing individuals with ASD to learn and practice social skills in a supportive environment.

Autism – Causes & Treatments

No research has uncovered a ‘characteristic’ brain structure for autism, meaning that no single pattern of changes appears in every autistic person. Studies of brain structure often turn up dissimilar results — there is great variety across individuals in general. But some trends have begun to emerge for subsets of autistic people. These differences might one day provide some insight into how some autistic people’s brains function. They may also point to bespoke treatments for particular subtypes of autism.

Studies that make use of a brain-scanning technique called magnetic resonance imaging (MRI) have highlighted a few brain regions that are structurally distinct in people with autism.

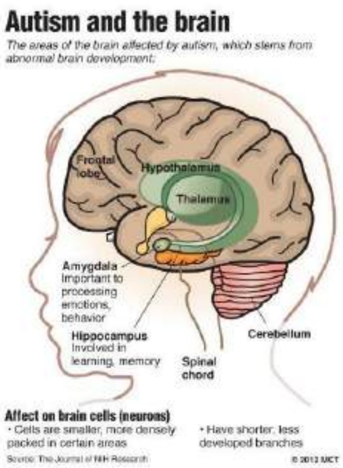

Autism and the brain
The areas of the brain affected by autism, which stems from abnormal brain development:
Frontal lobe
Hypothalamus
Thalamus
Amygdala Important to processing emotions, behavior
Hippocampus Involved in learning, memory
Spinal chord
Cerebellum
Affect on brain cells (neurons)
• Cells are smaller, more densely packed in certain areas
• Have shorter, less developed branches
Source: The Journal of NIH Research
© 2012 MCT

Figure 1 . Shows the areas of brain affected by Autism, Obtained from https://maximind.ca/brain-connectivity-and-autism

- Children and adolescents with autism often have an enlarged hippocampus, the area of the brain responsible for forming and storing memories, several studies suggest, but it is unclear if that difference persists into adolescence and adulthood.
- The size of the amygdala also seems to differ between people with and without autism, although researchers from different labs have turned up conflicting results. Some find that people with autism have smaller amygdalae than people without autism, or that their amygdalae are only smaller if they also have anxiety[3]. Others have found that autistic children have enlarged amygdalae early in development and that the difference levels off over time.
- Autistic people have decreased amounts of brain tissue in parts of the cerebellum, the brain structure at the base of the skull, according to a meta-analysis of 17 imaging studies. Scientists long thought the cerebellum mostly coordinates movements, but they now understand it plays a role in cognition and social interaction as well.
- On a more global level, the cortex — the brain's outer layer — seems to have a different pattern of thickness in people with and without autism. This difference tracks with alterations to a single type of neuron during development, a 2020 study suggests.

Traditional Treatments:

Traditional treatments for autism spectrum disorder (ASD) encompass a range of interventions aimed at addressing core symptoms, promoting skill development, and improving overall quality of life for individuals with ASD. Some of the most commonly used traditional treatments include:

1. **Applied Behavior Analysis (ABA)**: ABA is a widely used behavioral therapy that focuses on increasing desired behaviors and reducing problematic behaviors by breaking down tasks into smaller steps and providing positive reinforcement for progress.
2. **Speech and Language Therapy**: Many individuals with ASD experience difficulties with communication skills. Speech and language therapy aims to improve language development, social communication, and pragmatic language skills.

3. **Occupational Therapy (OT):** OT helps individuals with ASD develop fine motor skills, sensory processing abilities, and adaptive behaviors necessary for daily living activities. It may also address sensory sensitivities and promote sensory integration.
4. **Social Skills Training**: Social skills training programs teach individuals with ASD social interaction skills, such as initiating and maintaining conversations, interpreting social cues, and understanding social norms.
5. **Special Education Services**: Individuals with ASD often benefit from specialized educational programs tailored to their specific needs. These programs may include individualized education plans (IEPs), accommodations, and modifications to support academic, social, and behavioral goals.
6. **Medication**: While there is no medication that can treat the core symptoms of ASD, certain medications may be prescribed to manage associated conditions such as anxiety, depression, ADHD, or aggression.
7. **Parent Training and Support**: Providing parents and caregivers with education, training, and support is crucial for effectively managing the challenges associated with raising a child with ASD. Parent training programs often focus on behavior management strategies, communication techniques, and accessing community resources.
8. **Structured and Predictable Environment**: Consistency and predictability in the environment can help individuals with ASD feel more comfortable and secure. Creating structured routines and visual supports can aid in reducing anxiety and supporting learning and communication.

Limitations of Traditional Treatments:

Various therapies, such as Cognitive Behavior Therapy (CBT), early intervention, educational and school-based therapies, joint attention therapy, medication treatment, nutritional therapy, and occupational therapy, exist. However, each comes with its set of challenges and limitations.

Therapies	Challenges
Cognitive Behavior Therapy (CBT)	Requires Active Participation from both patient and therapist. CBT primarily focuses on addressing current problems and developing coping strategies for the future. It may not delve deeply into past experiences or traumas. CBT often requires time and consistent effort for individuals to see significant improvements.
Early Intervention.	Delays in diagnosis may postpone the initiation of early intervention, missing a critical period of developmental plasticity. Limited access can result in delayed or inadequate services. Limited availability of trained professionals can result in long waiting lists and reduced access to timely and consistent intervention services.
Educational and School-Based Therapies.	Schools may struggle to provide the necessary staff, training, and materials required to meet the diverse needs of students with autism. Insufficient collaboration and communication among educators, therapists, and parents can hinder the coordination of interventions.
Joint Attention Therapy.	Individuals with autism may have sensory sensitivities or challenges that impact their ability to engage in joint attention activities. Miscommunication or difficulty interpreting social cues may hinder the development of joint attention skills
Medication Treatment.	Pharmacological treatments for ASD are limited due to the heterogeneity of the disorder and the lack of real understanding of its pathology. Medical treatments are mainly used to control the secondary symptoms associated with ASD
Nutritional Therapy.	Implementing restrictive diets, such as the Gluten-Free Casein-Free (GFCF) diet, may pose challenges in meeting nutritional requirements. Restrictive diets may lead to nutritional deficiencies, especially if not properly monitored or supplemented.
Occupational Therapy.	Co-occurring behavioral issues, such as aggression or self-stimulatory behaviors, may interfere with the therapeutic process. Access to occupational therapy services may be limited, especially in certain geographical areas or for individuals with financial constraints

Common Problems:

Along with the above-mentioned specific limitations, there are few common problems with Traditional treatments such as

- Lack of certified & qualified Therapists
  - Out of 73 million children, approximately 1 in 36 children in US have ASD.
  - Over 5.4 million adults in US have ASD.
  - Total BCBA Certified therapists = 66,339.

- An autistic person has to wait for 4 to 5 weeks for a therapist appointment.
- Of 3,108 counties, about 54 percent do not have a single BCBA

- Cost of Therapy is High
  - Estimates range from $17,000 - $21,000. For intensive ABA therapy, yearly costs can easily exceed $60,000.

How does Generative AI work?

Generative AI chatbots have the potential to provide valuable support and assistance to autistic individuals by facilitating social communication, offering emotional support, providing personalized assistance, promoting daily living skills, and facilitating access to resources and information. By leveraging the capabilities of generative AI technology, chatbots can empower autistic individuals to navigate challenges, develop skills, and lead fulfilling lives.

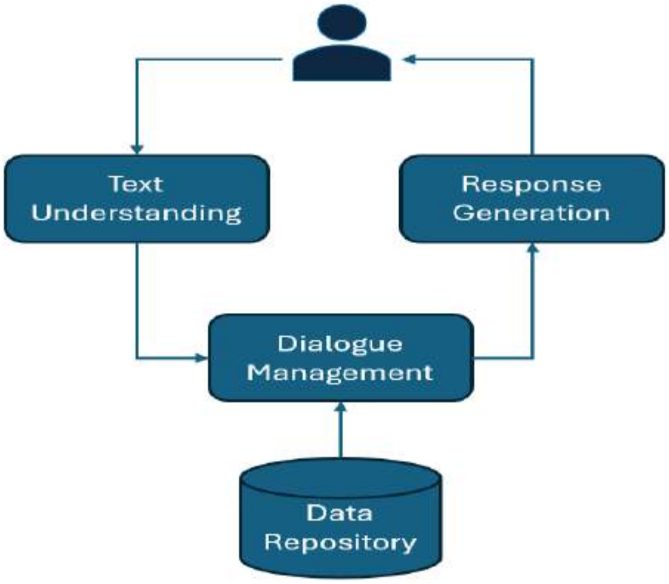


Figure 2. Shows how Generative AI (chatbot) works. Created and Copyright by Sashvathkumar.

Generative AI operates through a combination of natural language processing (NLP), machine learning algorithms, and predefined rules. Here's a breakdown of how it works:

1. **Natural Language Processing (NLP):** NLP is a branch of artificial intelligence that enables computers to understand and interpret human language. When a user interacts with a chatbot by typing or speaking, NLP algorithms analyze the input to extract meaning, intent, and context. This involves tasks such as tokenization (breaking text into words or phrases), syntactic analysis (identifying sentence structure), semantic analysis (understanding the meaning of words and phrases), and entity recognition (identifying relevant entities such as names, dates, or locations).
2. **Intent Recognition**: Once the user's input is processed, the chatbot's algorithms identify the user's intent or purpose behind the message. For example, if a user asks a question about a product, the chatbot recognizes the intent as "product inquiry." Intent recognition is crucial for understanding what action the chatbot should take in response to the user's query.

3. **Dialogue Management**: Based on the recognized intent and context, the chatbot determines the appropriate response or action to take. This involves accessing a knowledge base or database of predefined responses, rules, or algorithms to generate a relevant and meaningful reply. In some cases, the chatbot may use machine learning techniques to dynamically generate responses based on patterns learned from previous interactions.
4. **Response Generation**: Once the appropriate response is determined, the chatbot generates a text or speech output to communicate with the user. This may involve selecting a predefined response from a list of options, generating a response using natural language generation (NLG) techniques, or accessing external data sources to provide real-time information.
5. **Feedback Loop and Learning**: After delivering the response, the chatbot may collect feedback from the user to assess the quality and relevance of the interaction. This feedback can be used to improve the chatbot's performance over time through a process known as machine learning. By analyzing user interactions and feedback data, the chatbot's algorithms can learn and adapt to user preferences, improve response accuracy, and expand its knowledge base.
6. **Integration with Platforms**: Chatbots can be deployed on various platforms and communication channels, such as websites, messaging apps, voice assistants, and social media platforms. Integration with these platforms involves connecting the chatbot's backend systems with the user interface and communication channels, enabling seamless interaction between the user and the chatbot across different devices and platforms.

How does Virtual Reality work?

Virtual Reality is an up-and-coming form of treatment for Schizophrenia, but what is virtual reality? Virtual reality is a computer-generated environment that a user is placed into and can interact with through specialized equipment. An example of this equipment is an Oculus Quest and a handheld controller. Virtual Reality can be integrated with a patient’s treatment while they continue with standard treatment carried out by a psychiatrist. With the use of virtual reality, patients can be placed into specific scenarios chosen by a psychologist, helping them get over fears they may experience in public places or situations. It would be possible to collect data from the patient through conversation time, facial emotion recognition, and social anxiety exhibited by the patient while they are in a virtual environment. An adaptation algorithm that could change the virtual environment and tailor it towards the patient’s needs can also be implemented. It also provides a patient with a drug-free treatment option if they prefer not to consume medications. Virtual Reality works by creating the illusion that the reality the patient experiences within virtual reality is the only reality because it feels so real. “‘Even if you know intellectually that you’re not at the beach, your brain can’t live in two realities at once. Instead, the brain accepts [the input] it’s given’”, making it harder for patients to focus on other stimuli from the outside world such as pain or anxiety (Virtual Reality in Medicine, 2021). This is what causes intense immersion when you enter Virtual Reality and why many psychologists find Virtual Reality to be a viable option.

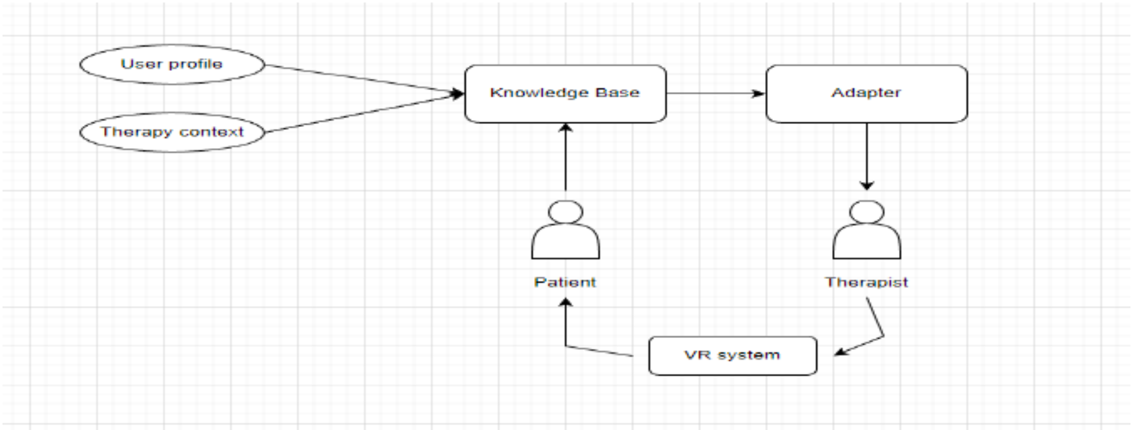


Figure 3. Shows how VR system works. Created and Copyright by Sashvathkumar

Alternate Treatments

How Generative AI Can Enhance VR for Autism Therapy

1. **Ultra-Personalized Content:**
   - Generative AI can tailor VR scenarios to the specific needs and interests of individuals with ASD. Imagine an AI system that analyzes a user's profile and generates social simulations involving their favorite characters or topics, increasing engagement and relevance.
   - For anxiety treatment, generative AI could adapt exposure therapy environments in real-time. If a user shows stress, the AI can subtly alter the simulation to be less overwhelming.
2. **AI-Powered Virtual Therapists or Companions:**
   - Generative AI can create virtual characters that act as therapists, practice partners, or simply supportive companions in the VR world.
   - These AI characters could provide immediate feedback on social interactions, model appropriate behaviors, and offer guidance for skill development.
3. **Dynamic and Adaptive Skill Building:**
   - Generative AI can make skill-building exercises far more flexible and adaptable. For example, during a virtual conversation, the AI could generate various dialogue paths and responses, making the interaction less scripted and more like a real-life encounter.
   - AI can analyze performance data and adjust the difficulty and content of simulations in real-time, optimizing learning.
4. **Data Analysis and Progress Tracking:**
   - VR combined with AI allows for deep analysis of a user's interactions within the simulations. This data can identify strengths, weaknesses, and patterns that may not be obvious to a human therapist.
   - AI can assist in detecting subtle cues related to emotions, eye tracking, and even physiological data, providing a comprehensive and objective evaluation

Generative AI + Virtual Reality Treatment:

- **Social Skills Training:** Discuss studies using VR for practicing social interaction, role-playing various scenarios, interpreting social cues, and understanding emotions. Highlight research that demonstrates improvements in recognizing facial expressions, maintaining eye contact, and initiating conversations within VR environments.
- **Communication Development:** Examine how VR can be used for enhancing both verbal and nonverbal communication skills. Review studies focused on turn-taking in conversations, practicing public speaking, and developing appropriate nonverbal cues.
- **Managing Anxiety and Sensory Sensitivities:** Review studies that explore VR's use in exposure therapy for managing anxieties and specific phobias. Discuss how VR creates safe spaces to confront triggers in a controlled manner. Additionally, cover research on VR's ability to create calming and controlled sensory environments for those with sensitivities.
- **Improving Motor Skills and Daily Living Skills:** Explore studies that examine VR simulations for developing motor coordination and practicing the sequences of tasks involved in activities of daily living. Provide examples such as VR programs for practicing dressing, hygiene routines, or cooking tasks.

Personal Avatar – Introduction & Setup

- **Tech setup**: VR Avatar with real time Generative AI in a controlled environment
- **Description**: This stage will introduce the autistic child to a VR avatar to have an engaging relationship and one-on-one conversation.
- **Summary**: Build a storytelling VR Avatar which will ask the autistic child (who wears the VR headset) to listen and focus on the avatar using eye tracking capability. If the child is not looking at the avatar while the avatar is narrating the story, the child will be asked to focus on the avatar. Stories will be preloaded in the device and grouped by common characters. For eg: Dora, Cinderella, Spiderman, Ironman, etc.
- **Expected Outcome**:
  - ❑ Improved eye-to-eye contact.
  - ❑ Improved focus and attention during a conversation.

<u>Pre-setup for StoryTeller</u>: The VR headset will be preloaded with a couple of stories for each cartoon character that are commonly liked by the autistic child. For eg: Dora, Cinderella, Spiderman, Ironman, etc.

1. The autistic child will wear the VR headset and launch the application.
2. Application loaded and AI Therapist Avatar will initiate the conversation by welcoming the autistic child with their name.
3. The autistic child who wears the VR headset will select their favorite character for stories.
4. The VR Avatar will start narrating the story.
5. The VR headset will track the eyeball movement of the autistic child. If the child is not looking at the Avatar, then the Avatar will ask the autistic child to keep looking at the

Avatar. For eg: “*Hi John, Look at me, if you are not interested in this story then go back to the home page and change to a different character*“

6. The autistic child can go back to the home page and listen to another story by selecting a different character.

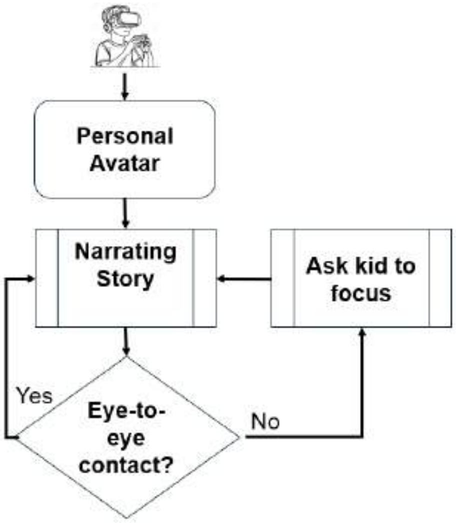


Figure 4. Shows Personal Avatar process flow. Created and Copyright by Sashvathkumar

Sequence diagram for StoryTeller flow:

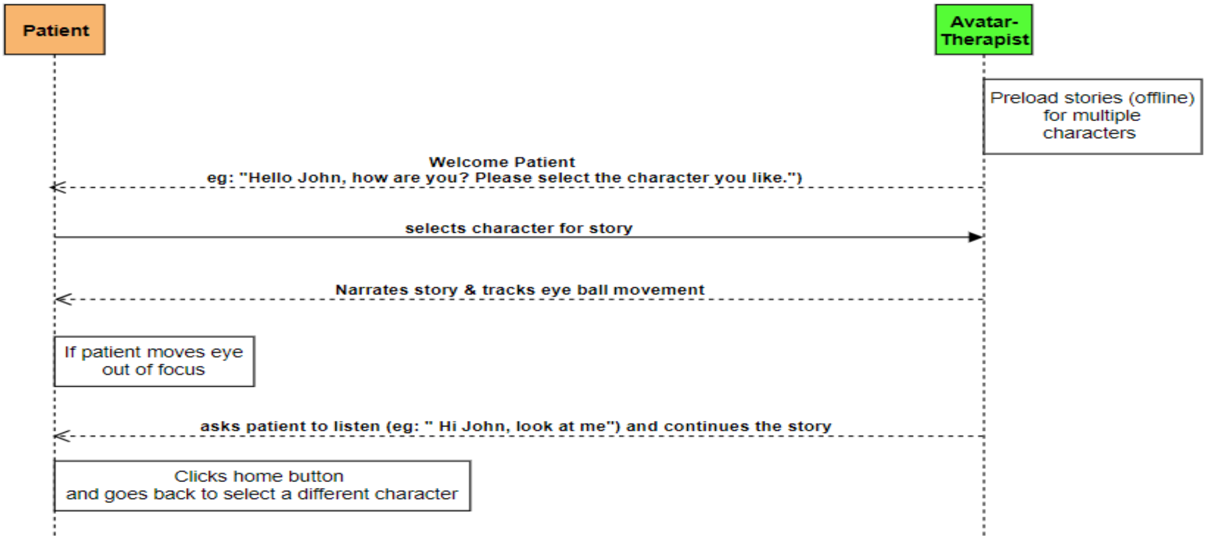

Patient
Avatar-Therapist
Preload stories (offline) for multiple characters
Welcome Patient
eg: "Hello John, how are you? Please select the character you like.")
selects character for story
Narrates story & tracks eye ball movement
If patient moves eye out of focus
asks patient to listen (eg: " Hi John, look at me") and continues the story
Clicks home button
and goes back to select a different character

Figure 5. Shows Sequence Diagram of conversation between Patient and Avatar. Created and Copyright by Sashvathkumar

Pre-Step for Conversational flow: An admin will feed the autistic child’s profile such as name, DOB, interests, and grade/employment to the database via an Admin portal.

1. The application will connect with Generative AI and request a list of possible questions based on the profile provided.
2. Generative AI will respond with a list of possible questions back to the Application.
3. Avatar will ask one of the questions to the autistic child in order to initiate the conversation, For eg: “*Hi John, it looks like you like baseball very much, who is your favorite baseball player?*”
4. Then wait for the autistic child to respond. Once they have responded, it will take the answer back to the Generative AI and get another related question based on the previous answer.
5. Avatar will ask this question back to the autistic child.
6. Likewise, the conversation will continue to happen, back and forth between the autistic child and the avatar.
7. In between, if the eyeball position is changing then the avatar will instruct the autistic child to look at it. For eg: “*Hi John, Look at me* “
8. Also, the Generative AI will suggest & correct any grammatical or sentence formation errors.
9. Avatar will take the autistic child to a safe zone in Metaverse to communicate and socialize with other people (Avatars) without any fear of embarrassment. This will improve the social communication and interaction ability of the autistic child.

The integrated Generative AI will ask questions about the autistic child’s interests and favorite players/cartoon characters/food. This will improve the child’s ability to express their feelings.

Sequence Diagram for Conversational Flow

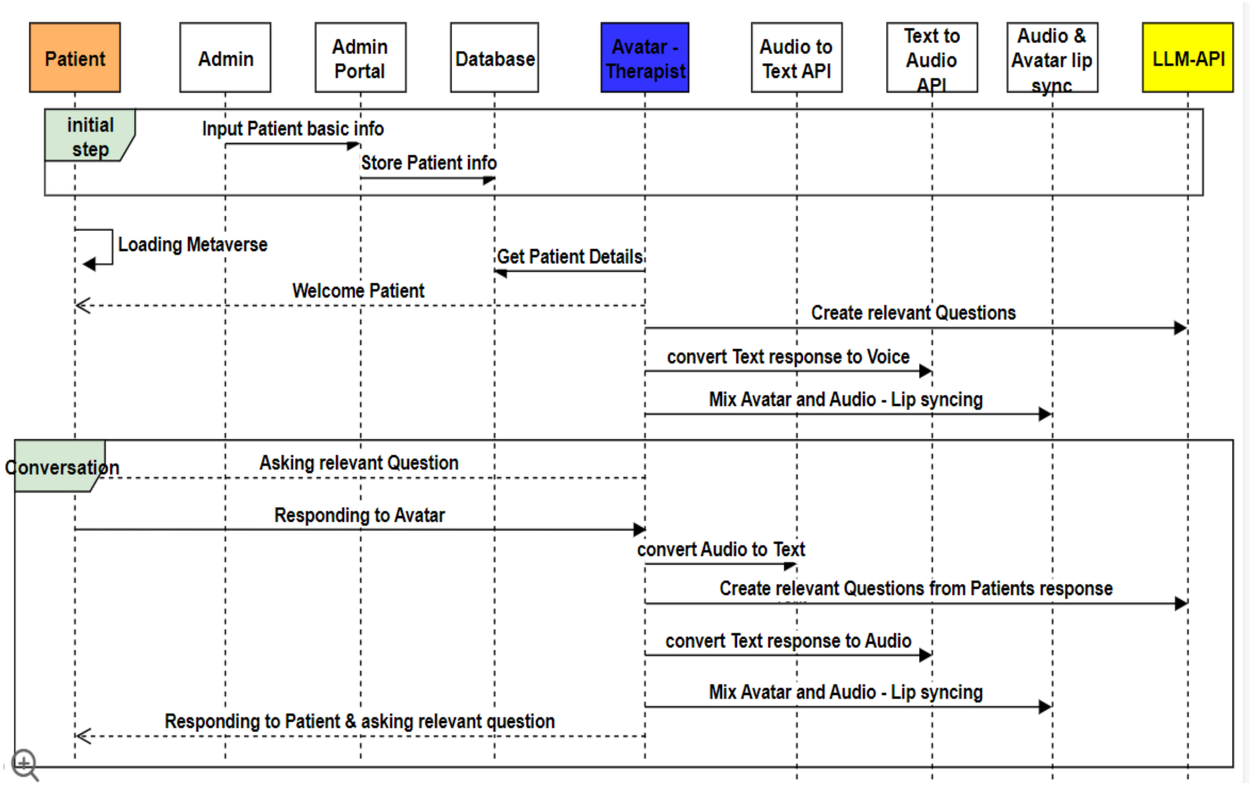
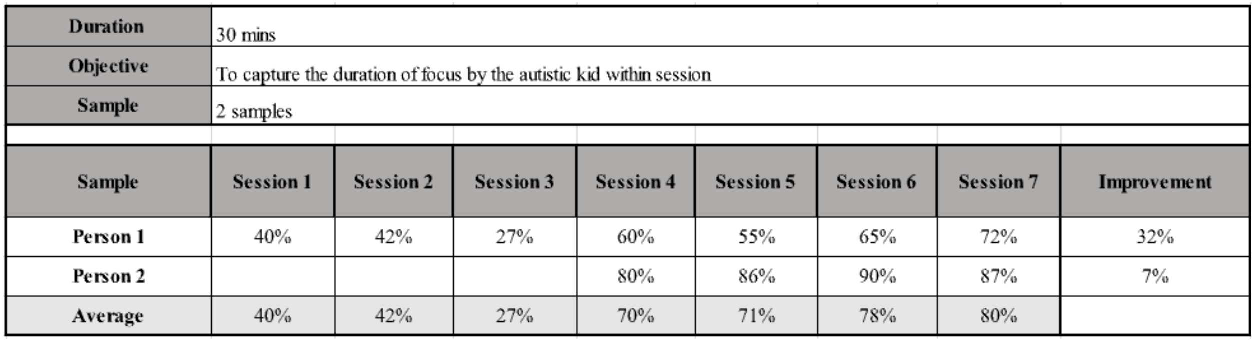


Figure 6. Shows Sequence Diagram of real time conversations between Patient and Avatar. Created and Copyright by Sashvathkumar

Research Data



Duration	30 mins							
Objective	To capture the duration of focus by the autistic kid within session							
Sample	2 samples							
Sample	**Session 1**	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**Session 6**	**Session 7**	**Improvement**
Person 1	40%	42%	27%	60%	55%	65%	72%	32%
Person 2				80%	86%	90%	87%	7%
Average	40%	42%	27%	70%	71%	78%	80%	

Figure 7. Shows Data captured for focus of the subjects. Created and Copyright by Sashvathkumar

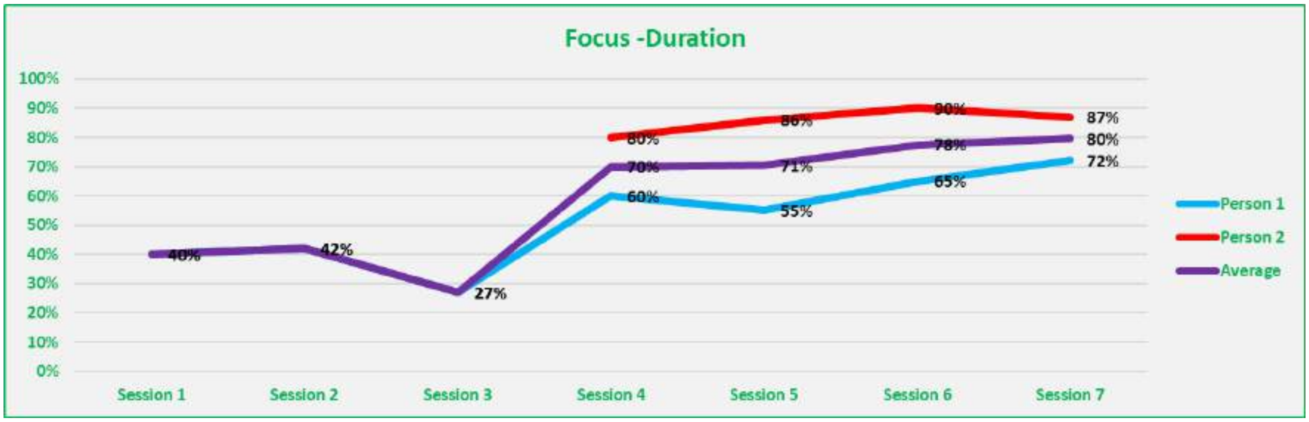
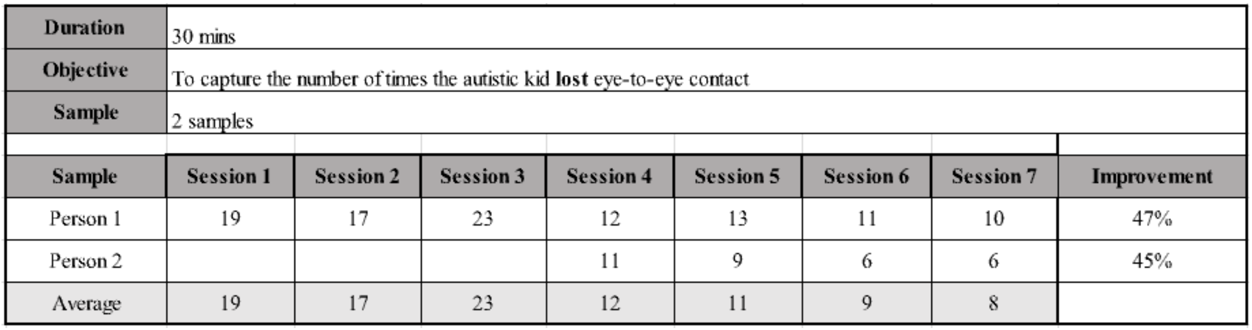


Figure 8. Shows graph for focus of the subjects. Created and Copyright by Sashvathkumar



Duration	30 mins							
Objective	To capture the number of times the autistic kid **lost** eye-to-eye contact							
Sample	2 samples							
Sample	**Session 1**	**Session 2**	**Session 3**	**Session 4**	**Session 5**	**Session 6**	**Session 7**	**Improvement**
Person 1	19	17	23	12	13	11	10	47%
Person 2				11	9	6	6	45%
Average	19	17	23	12	11	9	8	

Figure 9. Shows Data captured for eye to eye contact of the subjects. Created and Copyright by Sashvathkumar

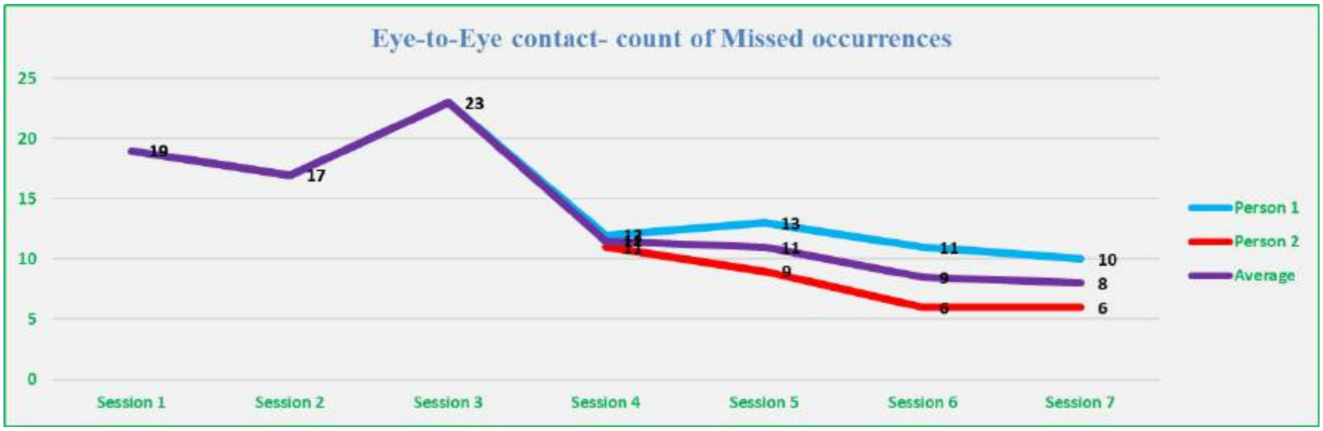


Figure 10. Shows graph captured for eye-to-eye contact of the subjects. Created and Copyright by Sashvathkumar

The data will be captured as mentioned below in every session and compare the results to see the progress of the autistic child.

- Capture how many times the autistic child made eye-to-eye contact with the Avatar.
- Capture the duration the autistic child continuously focuses on the Avatar.

- Capture the number of times disengaged from conversation with other avatars

Benefits of Generative AI + Virtual Reality

Combining Virtual Reality (VR) and Generative AI offers a powerful and synergistic approach to addressing the unique needs of autistic children, leading to improvements across various key parameters:

- **Improved Eye-to-Eye Contact**: VR simulations can create realistic social scenarios where children can practice making eye contact with virtual characters generated by AI. By gradually exposing children to these scenarios and providing positive reinforcement, the combination of VR and Generative AI can help desensitize them to the discomfort associated with eye contact and encourage more frequent and natural engagement.
- **Improved Focus and Attention During Conversation**: VR environments powered by Generative AI can create interactive and engaging experiences that captivate children's attention and maintain their focus during conversations. By tailoring the content and interaction dynamics to the individual child's interests and preferences, the combined approach can enhance attentional control and improve sustained engagement in social interactions.
- **Improved Social Interaction**: Virtual scenarios populated with AI-generated characters can provide children with opportunities to practice and develop social interaction skills in a controlled and supportive environment. Through repeated exposure to social situations and guided feedback from AI characters, children can learn and generalize social skills such as turn-taking, active listening, and perspective-taking, leading to improved social interaction abilities in real-world settings.
- **Improved Linguistic Abilities**: Generative AI can assist children in developing linguistic abilities by generating contextually relevant and grammatically correct responses during simulated conversations. By engaging children in dialogue and providing opportunities for language practice and reinforcement, the combined approach can facilitate language acquisition, vocabulary expansion, and syntax comprehension.
- **Improved Conversational Skills**: VR scenarios can simulate realistic conversational contexts where children can practice initiating and maintaining conversations with AI-generated characters. By incorporating conversational prompts, role-playing exercises, and feedback mechanisms, the combined approach can help children build confidence, fluency, and social reciprocity in their interactions with peers and adults.
- **Reduced Fear of Embarrassment**: The virtual nature of VR environments can create a safe and non-threatening space where children feel more comfortable experimenting with social interactions and expressing themselves without fear of judgment or embarrassment. By gradually exposing children to social challenges and providing positive reinforcement for their efforts, the combined approach can help reduce social anxiety and increase self-confidence in social settings.

Overall, the integration of Virtual Reality and Generative AI offers a promising avenue for addressing the core symptoms of autism spectrum disorder and promoting positive social, linguistic, and emotional outcomes in autistic children. Through immersive and interactive experiences tailored to individual needs, this combined approach has the potential to significantly enhance the effectiveness of intervention programs and improve the quality of life for children with autism.

Conclusion:

Results clearly show.

- Improved eye-to-eye contact on average by 46%.
- Improved focus on average by 20%
- Parents noticed an improvement in a normal conversational skill after these sessions.

The limitations of current social cognitive therapies may be overcome by combining VR and generative AI since it provides more affordable access, dependable use, and an enjoyable experience. The demand for patient motivation and the shortage of licensed therapists have rendered today's therapy unsuccessful. One way to deal with these issues is through virtual reality. This research further emphasizes the potential of the metaverse to treat autism because it can give people with autism more flexibility and social interaction. To sum up, virtual reality and the metaverse have a lot of promise for the development of virtual healthcare in the future and are an excellent way to treat autism.

Future Work: Social Avatar: Introducing to Metaverse Autistic Community (MAC)

Once an autistic child is comfortable and shows improvements with Personal Avatar, they are slowly introduced to Social Avatar for a more real-time conversational phase in the MAC environment.

<u>Social Avatar –MAC</u>

Creating a restricted Metaverse Autistic Community (MAC) in Metaverse and entrance to this community will be managed by an admin. This community will have other autistic child avatars, virtual volunteers, therapists and parents for more socialization.

References

- Kourtesis, Panagiotis, et al. “Virtual Reality Training of Social Skills in Adults with Autism Spectrum Disorder: An Examination of Acceptability, Usability, User Experience, Social Skills, and Executive Functions.” Behavioral Sciences, vol. 13, no. 4, 17 Apr. 2023, p. 336, https://doi.org/10.3390/bs13040336. Accessed 4 May 2023.

- Wang, Michelle, and Denise Reid. “Virtual Reality in Pediatric Neurorehabilitation: Attention Deficit Hyperactivity Disorder, Autism and Cerebral Palsy.” Neuroepidemiology, vol. 36, no. 1, 2011, pp. 2–18, https://doi.org/10.1159/000320847.

- Kandalaft, Michelle R., et al. “Virtual Reality Social Cognition Training for Young Adults with High-Functioning Autism.” Journal of Autism and Developmental Disorders, vol. 43, no. 1, 9 May 2012, pp. 34–44, https://doi.org/10.1007/s10803-012-1544-6.

- Parsons, Sarah, and Sue Cobb. “State-of-The-Art of Virtual Reality Technologies for Children on the Autism Spectrum.” European Journal of Special Needs Education, vol. 26, no. 3, Aug. 2011, pp. 355–366, https://doi.org/10.1080/08856257.2011.593831.

- Bellani, M., et al. “Virtual Reality in Autism: State of the Art.” Epidemiology and Psychiatric Sciences, vol. 20, no. 3, 4 May 2011, pp. 235–238, https://doi.org/10.1017/s2045796011000448.

- Maskey, Morag, et al. “Reducing Specific Phobia/Fear in Young People with Autism Spectrum Disorders (ASDs) through a Virtual Reality Environment Intervention.” PLoS ONE, vol. 9, no. 7, 2 July 2014, www.ncbi.nlm.nih.gov/pmc/articles/PMC4079659/, https://doi.org/10.1371/journal.pone.0100374.

- Shapiro, Gideon, and Dorothy G. Flood. Treatment of Autism Spectrum Disorders, Obsessive-Compulsive Disorder and Anxiety Disorders. patents.google.com/patent/US20210220365A1/en. Accessed 9 July 2024.

- Centers for Disease Control and Prevention. “Treatment and Intervention Services for Autism Spectrum Disorder.” Centers for Disease Control and Prevention, 9 Mar. 2022, www.cdc.gov/ncbddd/autism/treatment.html.

- Lotufo Denucci, Bruna, et al. “Current Knowledge, Challenges, New Perspectives of the Study, and Treatments of Autism Spectrum Disorder.” Reproductive Toxicology, vol. 106, 1 Dec. 2021, pp. 82–93, www.sciencedirect.com/science/article/abs/pii/S0890623821001660, https://doi.org/10.1016/j.reprotox.2021.10.010.

- Centers for Disease Control and Prevention. “Data & Statistics on Autism Spectrum Disorder.” Centers for Disease Control and Prevention, 2023, www.cdc.gov/ncbddd/autism/data.html.

- Tatom, Carol. “How Much Does ABA Therapy for Autism Cost?” Autism Parenting Magazine, 17 Nov. 2020, www.autismparentingmagazine.com/aba-therapy-autism-cost/.

- “How Many Hours of ABA Therapy Are Needed?” Www.crossrivertherapy.com, www.crossrivertherapy.com/how-many-aba-hours-is-needed.

- “Google Patents.” Patents.google.com, patents.google.com/?inventor=Ute+Geigenmuller&peid=60efee4cfe5c0%3Ac%3A1ab51eb6. Accessed 9 July 2024.

- Glock, Melanie. “Treating the Phobias of Individuals with Autism with VR.” Autism Research Institute, 3 June 2019, autism.org/virtual-reality-overcoming-phobias/.

- Lotufo Denucci, Bruna, et al. “Current Knowledge, Challenges, New Perspectives of the Study, and Treatments of Autism Spectrum Disorder.” Reproductive Toxicology, vol. 106, 1 Dec. 2021, pp. 82–93, www.sciencedirect.com/science/article/abs/pii/S0890623821001660, https://doi.org/10.1016/j.reprotox.2021.10.010.

Viability of Low-Cost Infrared Sensors for Short Range Tracking

Noah Haeske
noah.haeske@gmail.com

ABSTRACT

A classic task in robotics is tracking a target in the external environment. There are several well-documented approaches to this problem. This paper presents a novel approach to this problem using infrared time of flight sensors. The use of infrared time of flight sensors is not common as a tracking approach, typically used for simple motion detectors. However, with the approach highlighted in this paper they can be used to accurately track the position of a moving subject. Traditional approaches to the tracking problem often include cameras, or ultrasonic sensors. These approaches can be expensive and overcompensating in some use cases. The method focused on in this paper can be superior in terms of cost and simplicity.

Introduction

The sensor highlighted in this paper is the **VL53L7CX** by ST**.** This sensor was selected for its wide field of view (see methods) but is comparable to most other Infrared TOF sensors. This sensor works by emitting a 940nm wavelength light in the direction it is facing. The on-board microcontroller then begins a clock. The light is reflected off the target as well as other surroundings, and a fraction of it is returned to the sensor. The sensor contains an array of sixty-four single photon avalanche diodes (SPADs). Each of these diodes, when triggered by the reflected infrared light, stops the clock, and returns a time value. This time value can then be multiplied by the speed of light in order to calculate the distance the light traveled. This distance is then divided by two in order to find the distance from the sensor to the target.

$$\frac{1}{2}(t \times c) = d$$

Sensor Comparison

While not a traditional method of detection, in certain short-range applications infrared TOF sensors are an appealing option. In comparison to cameras and computer vision, infrared TOF sensors are significantly cheaper and require less space. In comparison to ultrasonic sensors, infrared TOF sensors offer a wider field of view (Adarsh, et. al). This is significant for small builds, where four or more ultrasonic sensors could be replaced by two infrared sensors.

Ambient Lighting

A known potential drawback of the approach used in this paper is the effects of ambient lighting. As previously mentioned, the ultrasonic sensor relies on an array of sixty-four SPADs to measure distance. These diodes are specifically designed to pick up on the 940nm wavelength infrared light emitted by the sensor. However in conditions with strong light coming from the external environment, the SPADs can detect light that did not originate from the sensor. This is a common source of error for infrared sensors.

Triangulation

As previously discussed, infrared TOF sensors return an array of distance values. However, these distance values on their own are not enough to derive position. While there are multiple ways to combine sensors in order to derive position, the method highlighted in this paper is a form of triangulation. In order to calculate position, two sensors placed in different positions along the same axis must scan the same area. The lowest distance value returned by each sensor is then stored in a program. The known distance between the two sensors is also stored in the program. These three distance values allow for the mathematical formulation of a triangle, with the distance between the two sensors as the base. The following equations are then used to find the (x, y) coordinates of the target.

$$\theta = \frac{\left(A^2+C^2-B^2\right)}{(2\times A\times C)}$$

$$x_2 = x_1 + (A \times \cos cos\,\theta)$$

$$y_2 = y_1 + (A \times \sin sin\,\theta)$$

In this these equations:

A: The lowest returned distance from sensor one
B: The lowest returned distance from sensor two
C: The known distance from sensor one to sensor two
θ: The angle between the axis the sensors are on and the target
x_1: The x coordinate of sensor one
y_1: The y coordinate of sensor one
x_2: The x coordinate of the target
y_2: The y coordinate of the target

Methods

In order to perform repeatable tests using this sensor, a test frame was constructed. To simulate the close range tracking this paper is focused on, the test frame was built in a 1m x 1m square, with one side missing to allow a subject to move in and out of the frame. Furthermore, to ensure the sensors remained in the same position through all the tests, a specialized holding bracket was modelled, and 3D printed. (see Fig. 1) The bracket allows the sensor to sit at a constant angle. It also incorporates holes for wiring and screws. Two of these brackets were installed in adjacent corners of the frame. (see Fig. 2)

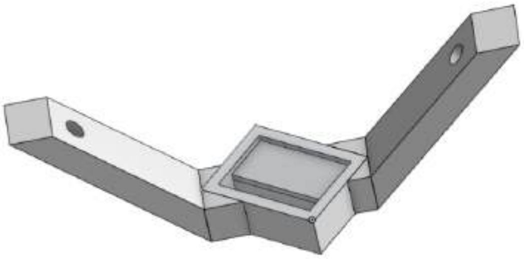


Figure 1

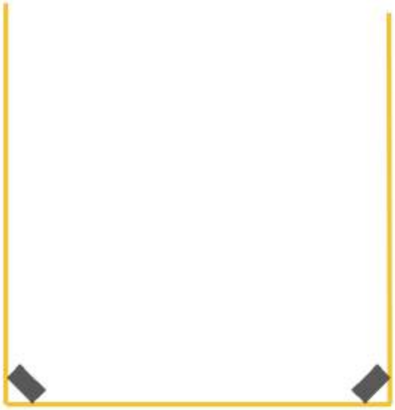


Figure 2

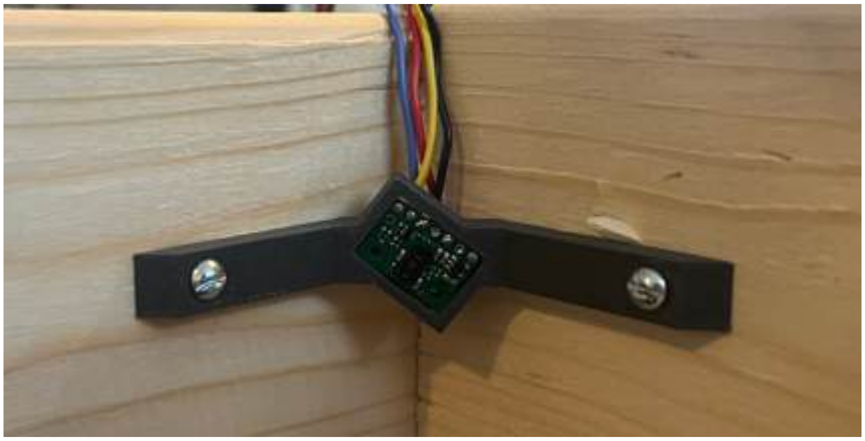


Figure 3

Figure 1. CAD model of sensor mounting bracket.

Figure 2. Drawing of test frame. The outer lines represent 1-meter lengths of wood. Gray regions represent mounting brackets.

Figure 3. The mounted sensor in the bracket.

A program was written in order to drive the sensors and interpret data. The script uses python as well as ST's library for the sensor. The program was run on a raspberry pi, to which the sensors were connected via I2C connections. A graphical interface was also created in order to debug the program. The graphical interface included a four by four heatmap from each sensor, as well as a marker indicating estimated position of the target. Code is available at https://github.com/noah-haeske/research

In order to determine accuracy of the system, two experiments were conducted. In each experiment four locations were measured out and marked within the frame. A measuring target was created from a tripod attached to a 10 cm diameter foam cylinder. The stand was covered in fabric to simulate a clothed human limb. The actual measured position of the stand was recorded. The system was then run for 10 trials (~10 seconds) and the reported position of the stand was recorded. The stand was then moved to the next marked location and all steps were repeated. This experiment was performed under two different ambient lighting conditions: no ambient light, and artificial lighting.

Results

After conducting the previously mentioned experiments, the results were as follows.

Experiment One

Experiment one was performed in artificial ambient lighting. Ten trials were performed in each of the target's four set positions. The recorded measurements can be found below. All values are in millimeters.

Table 1.

Trial	**Position 1 x**	**Position 1 y**	**Position 2 x**	**Position 2 y**	**Position 3 x**	**Position 3 y**	**Position 4 x**	**Position 4 y**
1	304	298	647	267	659	715	300	751
2	329	266	640	263	654	769	242	767
3	305	278	636	261	804	666	204	787
4	307	269	637	266	640	734	152	781
5	295	297	637	270	636	728	203	785
6	302	296	634	247	661	745	226	774
7	317	258	643	273	652	763	247	768
8	319	289	625	238	630	664	225	770
9	315	276	628	243	645	667	250	772
10	292	294	637	253	526	725	210	770
Actual	330	330	660	330	330	660	660	660

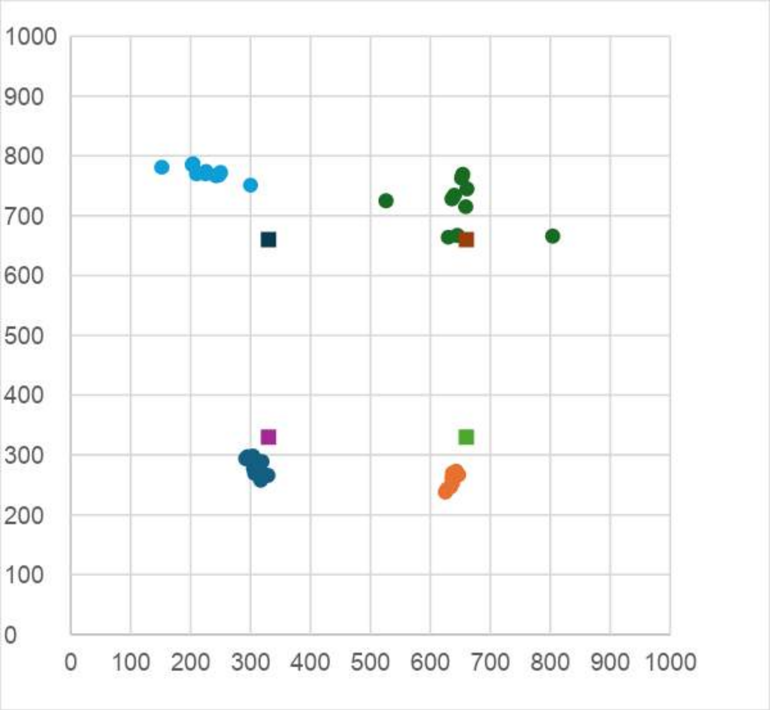


Figure 3

Figure 3. A scatterplot of the measured values from experiment 1. Each color represents a trial with the target placed in a different position. Each dot represents one trial of the program. Each square represents an actual position of the target. All measurements are in millimeters.

Experiment Two

Experiment two was performed with no ambient lighting. Ten trials were performed in each of the targets four set positions. The recorded measurements can be found below. All values are in millimeters

Table 2.

Trial	**Position 1 x**	**Position 1 y**	**Position 2 x**	**Position 2 y**	**Position 3 x**	**Position 3 y**	**Position 4 x**	**Position 4 y**
1	314	324	648	294	645	675	300	669
2	315	304	638	296	686	680	342	661
3	315	295	646	295	662	666	335	680
4	317	289	636	292	670	673	311	682
5	299	332	647	301	647	682	305	673
6	317	299	636	296	655	690	347	649
7	327	288	640	303	652	687	343	648
8	335	326	635	298	657	683	347	671
9	317	306	631	293	671	684	323	682
10	312	309	647	305	649	675	301	689
Actual	330	330	660	330	660	660	330	660

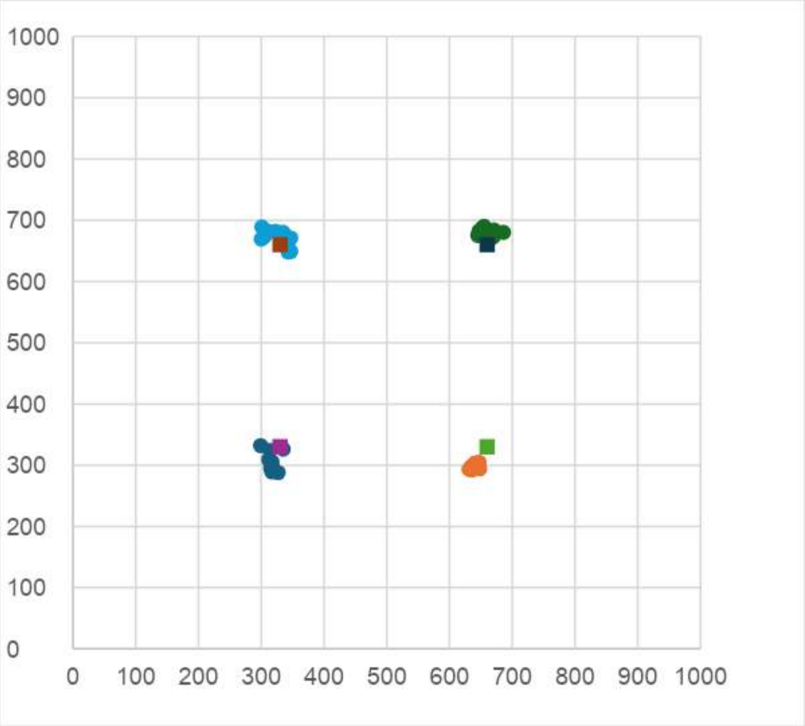


Figure 4

Figure 4. A scatterplot of the measured values from experiment two. Each color represents a trial with the target placed in a different position. Each dot represents one trial of the program. Each square represents an actual position of the target. All measurements are in millimeters.

Further Calculations

Further calculations were done in order to enhance understanding of the systems precision and accuracy. The standard deviation between each of the ten trials performed in each position were calculated. The standard deviation is based on the grouping of the trials, and can be used to track precision across the experiments. The data is shown below.

Table 3.

Standard Deviation	Position 1 x	Position 1 y	Position 2 x	Position 2 y	Position 3 x	Position 3 y	Position 4 x	Position 4 y	Average
Exp. 1	10.85	13.87	6.14	11.48	63.27	37.40	36.74	9.83	23.70
Exp. 2	8.89	14.77	5.82	4.15	12.29	6.86	18.66	13.28	10.59

Furthermore, the percentage error for each position in each experiment was calculated. The percentage error is based on the actual position of the target and can be used to track accuracy across the experiments. The data is shown below

Table 4.

Percent Error	Position 1 x	Position 1 y	Position 2 x	Position 2 y	Position 3 x	Position 3 y	Position 4 x	Position 4 y	Average

Exp. 1	6.52%	14.52%	3.58%	21.79%	97.18%	8.73%	65.77%	17.05%	29.39%
Exp. 2	4.30%	7.03%	2.97%	9.91%	1.58%	2.95%	5.27%	2.27%	4.54%

Discussion

The data from experiments one and two lead to interesting discoveries about the system. Overall, the system performed well. Visually, there is a noticeable difference between the experiment with ambient lighting and no ambient lighting. This is further emphasized in the calculations section. In the experiment with no ambient light, the system performed with over two times the precision of the experiment with ambient light. This is shown in the grouping of the scatterplots, as well as in the standard deviation figures. Additionally, the experiment with no ambient light performed with approximately six times the accuracy of the experiment with ambient light. Once again, this is shown both in the scatter plot, and the percentage error figures. In both experiments, a decrease in both precision and accuracy can be observed as the range of the target increased. This effect is especially prevalent in experiment one, as the ambient light appears to decrease effective range. Another trend to notice is the grouping of some measurements not over the target. In certain instances, such as experiment two's position two, the dots on the scatterplot are tightly grouped indicating high precision. However, this group of dots is shifted slightly away from the actual position of the target. This could indicate problems with the program, or the need for further calibration.

Conclusion

Throughout both experiments, infrared TOF sensors have proved to be a viable alternative to traditional tracking means. This study has revealed not only the viability of these sensors, but also shown strengths and weaknesses of this approach, illuminating possible use cases. It is clearly shown in the experiments that the system functions better with no ambient lighting. This data could reveal an application in nighttime tracking for these sensors. This is especially viable when considering the difficulty with the camera and computer vision approach at night (Liu, et. al). Further experimentation with ultrasonic sensors on a similar test setup would be beneficial to make accurate comparisons. One potential limitation of this system shown in this study was the decrease in accuracy with increased distance to the target. Further tests could be conducted in order to quantify what distance from the sensor causes this decrease in accuracy. Additionally, accuracy results from this study could be improved with better software calibration. In many trials of the experiments, reported values were y-shifted away from the target. This y-shift could be measured and compensated for in software.

References

Adarsh, S, et al. "Performance comparison of infrared and ultrasonic sensors for obstacles of different materials in vehicle/ Robot Navigation Applications." *IOP Conference Series: Materials Science and Engineering*, vol. 149, Sept. 2016, p. 012141, https://doi.org/10.1088/1757-899x/149/1/012141.

Liu, Jiaying, et al. "Benchmarking low-light image enhancement and beyond." *International Journal of Computer Vision*, vol. 129, no. 4, 11 Jan. 2021, pp. 1153–1184, https://doi.org/10.1007/s11263-020-01418-8.

"VL53L7CX." *STMicroelectronics*, www.st.com/en/imaging-and-photonics-solutions/vl53l7cx.html#overview. Accessed 10 July 2024.

A simple mechanical innovation in industrial uses of Stators

Introduction:

Precision Stampings is the largest and oldest manufacturer and exporter of Electrical Stampings and Laminations in India, with an annual turnover upwards of US$225 million. Established in 1971 they make custom made stampings to exact specifications. One of the products they produce which I found an innovation for .(Precision stamping 2023).

Stators play a pivotal role in the operation of various electrical machines, particularly in electric motors and generators. They are responsible for creating the magnetic field that interacts with the rotor to produce motion, making advancements in their design and production methods essential for the engineering industry. Traditional methods of producing stators often involve straight-line configurations, which, while effective, may not optimize performance or efficiency in all applications.

In this research, we explore an innovative solution to produce skewed lines on stators using a triangular model to introduce friction, thereby creating angles on the skew. This novel approach not only enhances the performance of the stators but also offers significant economic benefits. By implementing this method, we managed to save a company approximately 50 lakhs, highlighting its potential for substantial cost savings in the industry.

This literature review aims to contextualize the importance of stators, examine current production methods, and critically assess the potential impact of the proposed innovation. We will explore the engineering and economic implications of this new method and review existing research on the use of automation, particularly robots, in the production of stators.

Thematic Review

Engineering and Physics of Stators

Stators are essential components in electric motors and generators, responsible for creating the magnetic field that interacts with the rotor to produce motion. The design and configuration of stators can significantly influence the efficiency and performance of the machines they are part of.

Traditional stator designs often feature straight-line configurations which limit their possible uses, it can be relatively simple to produce but may not always offer optimal performance .

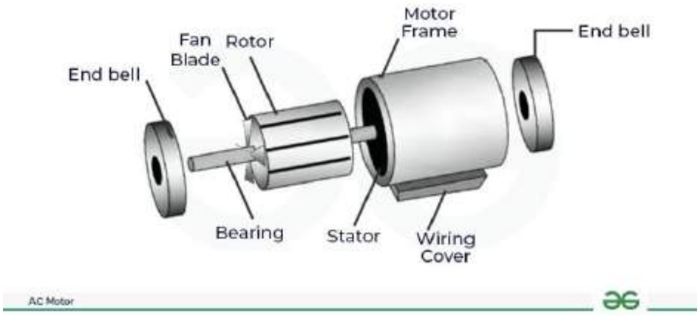
Recent advancements in engineering have introduced various methods to enhance the design and functionality of stators. One such method involves creating skewed stator lines, which can reduce harmonic distortion and improve the smoothness of the machine's operation. Skewing the stator lines can lead to a more efficient distribution of the magnetic field, reducing losses and enhancing overall performance .

The mechanics of stators

Stators are essential components in various electromechanical systems, serving a crucial role in the generation and control of electrical energy. Found predominantly in electric motors and generators, stators play a pivotal part in converting electrical energy into mechanical energy or vice versa. Unlike rotors, which rotate within the system, stators remain stationary.

In electric motors, the stator creates a magnetic field when energized, inducing the rotation of the rotor. This rotational motion is harnessed for applications ranging from industrial machinery to household appliances. Conversely, in generators, the stator is responsible for producing a magnetic field as it interacts with the rotor's motion, thus generating electrical energy.

Stators are integral to the functioning of electric motors and generators because they provide a stable and fixed magnetic field that interacts with the moving components. This interaction enables the conversion of energy between electrical and mechanical forms, making stators indispensable for powering a diverse array of devices and machinery essential to modern life. Without stators, the efficient and controlled conversion of electrical energy into mechanical work or vice versa would be unattainable, hampering the functionality of countless technological applications.



In an alternating current (AC) motor, several key components work together to convert electrical energy into mechanical energy. Understanding the role of each part is crucial for appreciating how the stator operates within the system.

End Bell

Function: The end bells are situated at both ends of the motor. They serve as structural support, housing the bearings that facilitate the smooth rotation of the rotor. Additionally, they protect the internal components from dust and other environmental contaminants.

Contribution to Stator Operation: The end bells are critical in maintaining the alignment of the rotor relative to the stator. By ensuring that the rotor spins centrally within the magnetic field generated by the stator, the end bells enable efficient interaction between the rotor and the stator's magnetic field, thereby optimizing the motor's performance.

Fan Blade

Function: The fan blade is attached to the rotor. Its primary function is to cool the motor by dissipating heat generated during operation. This is crucial for preventing overheating and ensuring the longevity of the motor components.

Contribution to Stator Operation: The cooling provided by the fan blade helps maintain the optimal temperature of the stator windings. This prevents thermal degradation of the insulation and ensures consistent performance of the stator's magnetic field generation.

Rotor

Function: The rotor is the rotating component of the motor. When the stator generates a magnetic field, it induces a current in the rotor, causing it to rotate. This rotational motion can then be used to perform mechanical work.

Contribution to Stator Operation: The rotor directly interacts with the magnetic field produced by the stator. As the stator creates a rotating magnetic field, it induces a current in the rotor, leading to its rotation. The efficiency of this interaction depends on the precise alignment and design of both the rotor and stator.

Bearing

Function: Bearings support the rotor, allowing it to spin freely with minimal friction. They are essential for the smooth and efficient operation of the motor.

Contribution to Stator Operation: By reducing friction and supporting the rotor, bearings ensure that the rotor remains properly aligned within the stator's magnetic field. This alignment is crucial for the effective transfer of electrical energy to mechanical energy.

Stator

Function: The stator is the stationary part of the motor that generates a rotating magnetic field when supplied with AC power. It consists of windings or coils that produce this magnetic field.

Contribution to Motor Operation: The stator is fundamental to the motor's operation. It creates the magnetic field that induces the rotor's rotation. The design and configuration of the stator windings significantly influence the efficiency and performance of the motor.

Motor Frame

Function: The motor frame houses all the internal components of the motor, providing structural integrity and protection. It also helps dissipate heat generated during operation.
Contribution to Stator Operation: The motor frame supports the stator, ensuring its stability and alignment. It also aids in heat dissipation, which is crucial for maintaining the performance and longevity of the stator windings.

Wiring Cover
Function: The wiring cover protects the electrical connections and wiring within the motor. It prevents accidental contact and shields the wiring from environmental factors.
Contribution to Stator Operation: By protecting the electrical connections, the wiring cover ensures that the stator receives a stable power supply. This stability is essential for the consistent generation of the magnetic field required for motor operation.

Existing Methods of Producing Skewed Stators

The production of skewed stators typically involves complex manufacturing processes that require precise control and advanced machinery. One such method is detailed by Nuzzo et al. (2017) in their study on modeling skew and its effects in salient-pole synchronous generators. This research highlights the use of mechanical adjustments and specialized tools to achieve the desired skew in stator windings, emphasizing the precision required in such processes. The techniques employed include the multislice (MS) and single-slice approaches, which rely on detailed electromagnetic analysis and the permeance function to model skewing effects accurately.

While effective in enhancing the performance of synchronous generators by reducing torque ripple and electromagnetic losses, these conventional methods are inherently costly and time-consuming. The need for advanced machinery and precise control mechanisms translates into significant production expenses. Additionally, the complexity of the processes can limit their adoption, especially in smaller manufacturing setups that may not have access to the necessary resources and expertise.

In contrast, our proposed innovation introduces a simpler and more cost-effective approach to producing skewed stator lines. By using a triangular model to create friction, the method allows for the formation of angles on the skew without the need for intricate machinery. This innovative technique leverages the natural interaction between the triangular model and the stator material to achieve the desired skew, thereby reducing the reliance on complex mechanical adjustments.

The comparative analysis of our triangular model with the conventional methods discussed by Nuzzo et al. (2017) underscores the advantages of our approach in terms of cost-effectiveness and simplicity. While traditional methods like the multislice and single-slice approaches offer precise control and improved performance, the associated expenses and complexity pose significant challenges. Our triangular model addresses these challenges by providing an alternative that maintains performance enhancements while significantly reducing production costs and complexity.

Economic Impact of Motors Using Stators

The economic impact of motors utilizing stators is profound, influencing not only individual companies but also the broader economy of many countries. Stators are critical components in electric motors, facilitating the conversion of electrical energy into mechanical energy, thereby enabling the operation of a wide range of industrial machinery and equipment. This section explores the significance of stators in the engineering sector, emphasizing their contribution to economic growth, with particular reference to the insights provided by Trevelyan (2012) on the Australian context.

Role of Stators in the Economy why it is important for countries to run

Electric motors, which rely on stators for their operation, are indispensable in numerous industrial applications, including manufacturing, transportation, and energy production. The stator's function in generating the magnetic field necessary for motor operation is crucial for the conversion of electrical energy into mechanical energy. This fundamental process drives various machinery and equipment, underscoring the importance of stators in industrial operations.

1. Manufacturing Sector:

Efficiency and Productivity**:** Advanced stator designs significantly enhance the efficiency and productivity of manufacturing processes. High-efficiency motors, which reduce energy consumption, lower operational costs and boost industrial competitiveness. This efficiency is particularly critical in high-demand sectors, where minimizing energy costs is paramount.
Automation and Robotics: The integration of electric motors with optimized stator designs is essential for automation and robotics. These motors enhance precision, reliability, and output in automated systems, enabling manufacturers to achieve higher production rates and improve product quality.

2. Transportation Sector:

Electric Vehicles (EVs): The shift towards electric vehicles is a transformative trend in the transportation sector. Stators are integral to the electric motors powering these vehicles, contributing to the development of sustainable and energy-efficient transportation solutions. The adoption of EV technology is crucial for reducing greenhouse gas emissions and promoting environmental sustainability.

Railways and Aviation: In railways and aviation, electric motors with advanced stator designs are used in propulsion systems and auxiliary equipment. These applications enhance the efficiency, reliability, and environmental performance of transportation systems, supporting advancements in both sectors.

3.Energy Sector:

Renewable Energy Generation: Stators play a pivotal role in renewable energy systems, such as wind turbines and hydroelectric generators. These systems convert mechanical energy from natural sources into electrical energy, facilitating the transition to sustainable energy solutions. The efficiency of stator designs directly impacts the performance and output of renewable energy systems.

Energy Efficiency: The development of high-efficiency motors with optimized stators contributes significantly to energy efficiency across various industrial applications. This not only reduces electricity consumption but also minimizes the environmental footprint of industrial operations.

Economic Benefits at a National Level

The widespread use of electric motors equipped with advanced stators has substantial economic implications at a national level. By enhancing industrial efficiency and productivity, these motors drive economic growth, create employment opportunities, and support technological advancements. Moreover, the adoption of energy-efficient technologies reduces energy costs, mitigates energy supply risks, and promotes sustainability.

The Australian Context

In the context of Australia, Trevelyan (2012) examines the challenges faced by engineers in articulating the commercial value of their work. His research, drawing on empirical data from interviews and field observations, reveals that many engineers and professional organizations struggle to convey the economic significance of engineering contributions. This perception issue may undermine the recognition of engineering's role in economic development and innovation.

1.Engineering Practice and Economic Value:

Commercial Value of Engineering: Trevelyan's study underscores the substantial economic impact of engineering, particularly through the development and implementation of technologies that enhance industrial efficiency. The use of advanced stator designs in electric motors exemplifies this contribution, demonstrating the tangible benefits of engineering innovation to the economy.

Public Perception and Support: Enhancing the public and employer understanding of engineering's value, as suggested by Trevelyan, is crucial for fostering support for engineering education and research. This understanding can drive investments in engineering talent and innovation, further accelerating technological advancements in sectors reliant on stator technology.

2***. Case Study of Economic Impact:***

Cost Savings and Efficiency Gains: The implementation of innovative stator designs, such as the triangular model proposed in this study, has demonstrated significant economic benefits. In our case study, this innovation resulted in cost savings of approximately 50 lakhs, underscoring its potential to enhance economic efficiency and competitiveness in the industry.

Hence this economic analysis tells us why stators are such an essential mechanical part and why its manipulation with skews can bring about such a massive influence which is one of the reasons why I find it so fascinating.

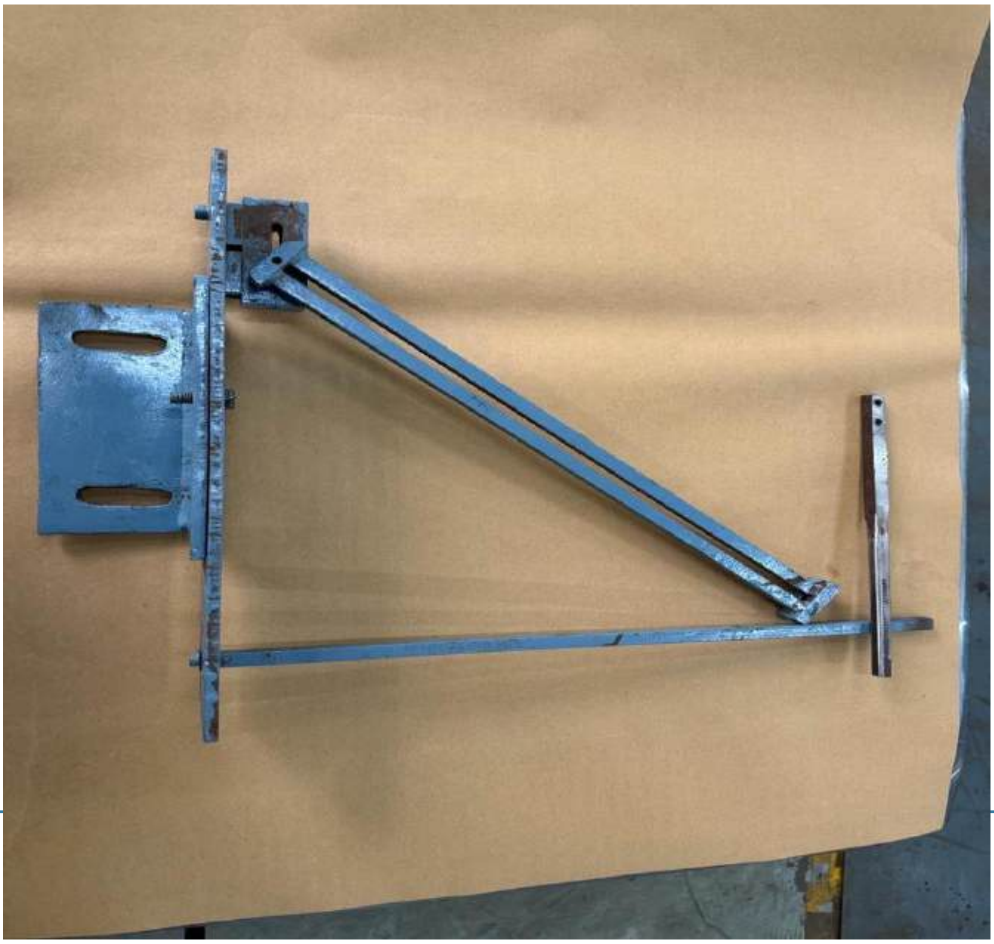
A simple and circumspect solution

In the realm of industrial innovation, The company Precision stampings, faced a challenging proposal from one of its clients: the creation of skewed stators. This concept was uncharted territory for the company, which had previously only dealt with the intricacies of straight stators. The allure of skewed stators lay in their promising advantages, including smoother power delivery at very low RPM, a reduced acoustic signature, enhanced starting torque, decreased starting current, increased slip, avoidance of magnetic locking (cogging effect), and a reduction in humming noise. However, embracing this new technology came with a hefty price tag—a £50,000 investment in specialized machinery.

Contemplating the substantial cost and the uncertain returns on the project, the company found itself at a crossroads. It leaned towards the possibility of dropping the client altogether. However, after discussing the dilemma with my friend's father, who happened to be well versed in the intricacies of wielding machinery, I envisioned a simple yet efficient alternative after visiting their factory .

The breakthrough idea involved leveraging the principles of friction and simple rigid body mechanics. The proposed solution was elegantly straightforward: create a triangular model in which you could place an extension from the wielding machine, causing the effect of friction on the machine and hence could allowing it to weld while rotating, thereby achieving the desired skewed effect. This ingenious approach not only bypassed the need for an exorbitant investment in new machinery but also showcased the power of applying fundamental physics concepts to real-world challenges.

The physics behind the solution.



$System\ Setup$

Let:
$-\ (\ B\ )\ be\ the\ base\ of\ the\ triangle$

$-\ (\ H\ )\ be\ the\ height\ of\ the\ triangle$

$-(L) be\ the\ hypotenuse, where (L = \sqrt{B^2 + H^2})$

$-(\theta) be\ the\ angle\ between\ the\ base\ and\ the\ hypotenuse, (\theta = \tan^{-1}\left(\frac{H}{B}\right))$

$-\ (\ m\ )\ be\ the\ mass\ of\ the\ welding\ machine$

$-\ (\ g\ )\ be\ the\ acceleration\ due\ to\ gravity$

$-\ (\ \mu)\ be\ the\ coefficient\ of\ friction$

2. $Force\ Components$

The components of gravitational force along the hypotenuse and perpendicular to the hypotenuse are given by:

$$F_{g\_parallel} = mg\sin(\theta)$$

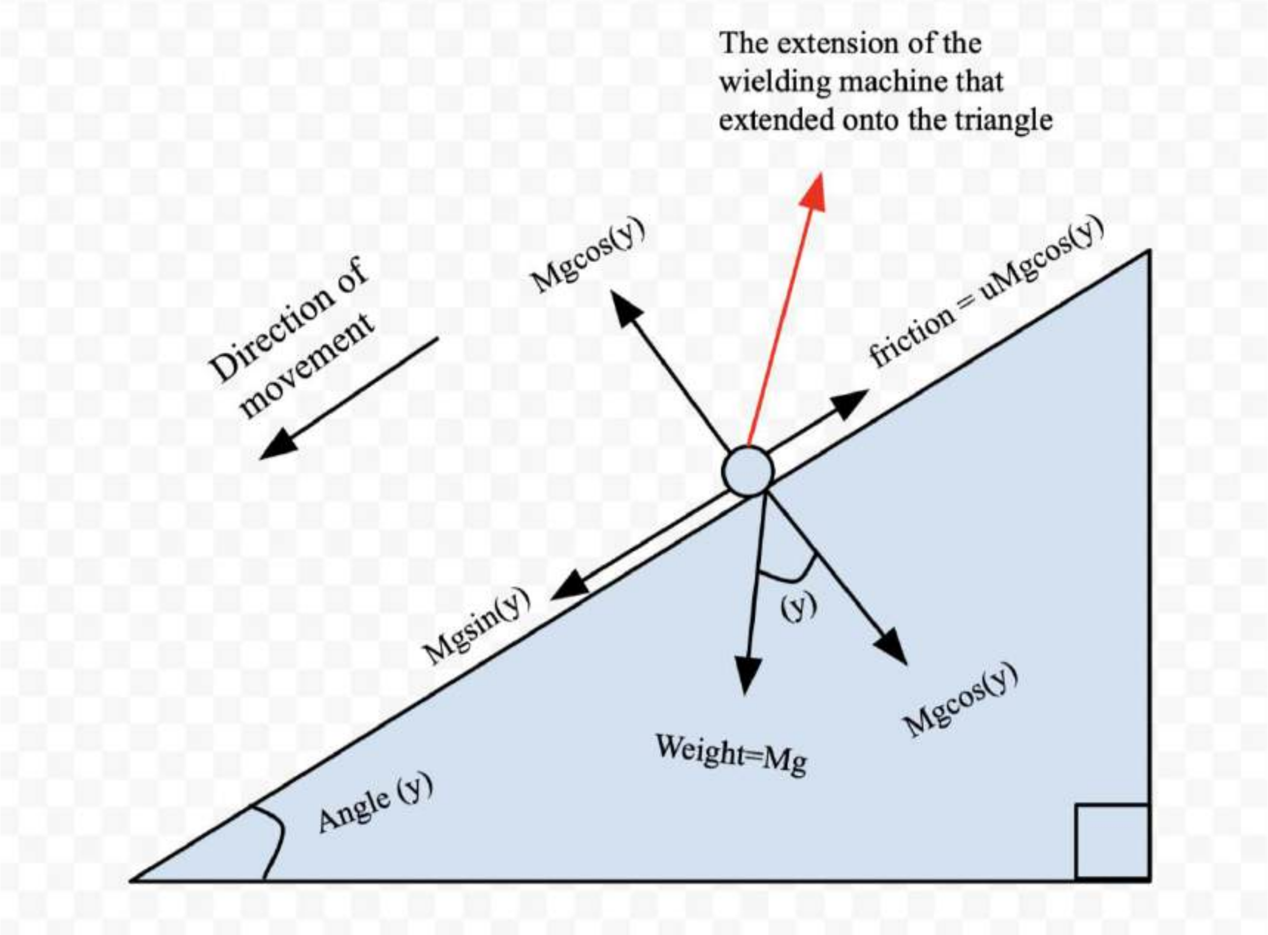
$$F_{g\_perpendicular} = mg\cos(\theta)$$

3. $Frictional\ Force$

The normal force and the resulting frictional force are:

$$F_n = F_{g\_perpendicular} = mg\cos(\theta)$$

$$F_f = \mu F_n = \mu mg\cos(\theta)$$



The diagram above shows how as the extension from the welding machine goes down it faces a frictional force in the opposite direction which leads to the rotational effect on the Stator as will be shown below.

4. $\boldsymbol{Torque\ and\ Angular\ Acceleration}$

The torque caused by the frictional force and the moment of inertia of the welding machine about the pivot point are:

$$\tau = F_f \cdot r$$

$$Assuming(r = \frac{L}{2}):$$

$$\tau = \mu mg\cos(\theta) \cdot \frac{L}{2}$$

$$I = m\left(\frac{L}{2}\right)^2 = \frac{mL^2}{4}$$

$$\mathit{The\ angular\ acceleration\ is\ then}:$$

$$\alpha = \frac{\tau}{I} = \frac{\mu mg\cos(\theta) \cdot \frac{L}{2}}{\frac{mL^2}{4}} = \frac{2\mu g\cos(\theta)}{L}$$

]

5. *Angular Velocity*

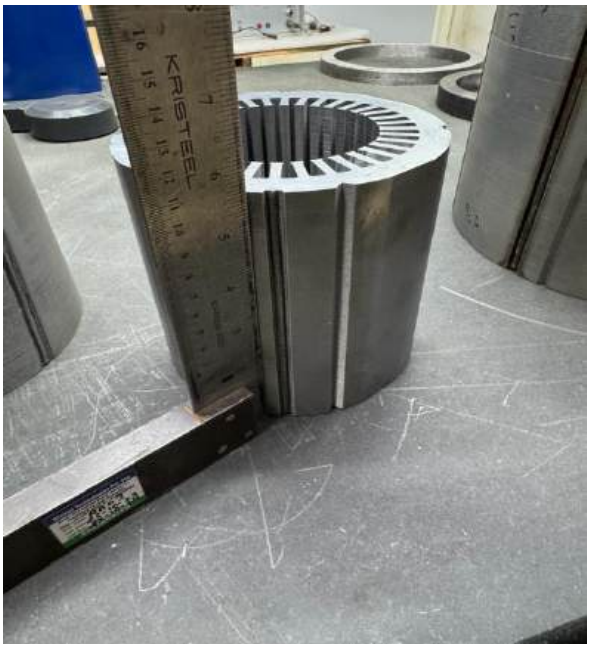
Assuming initial angular velocity is zero, the angular velocity after a time t is:

$$\omega = \alpha t = \frac{2\mu g\cos(\theta)}{L} t$$

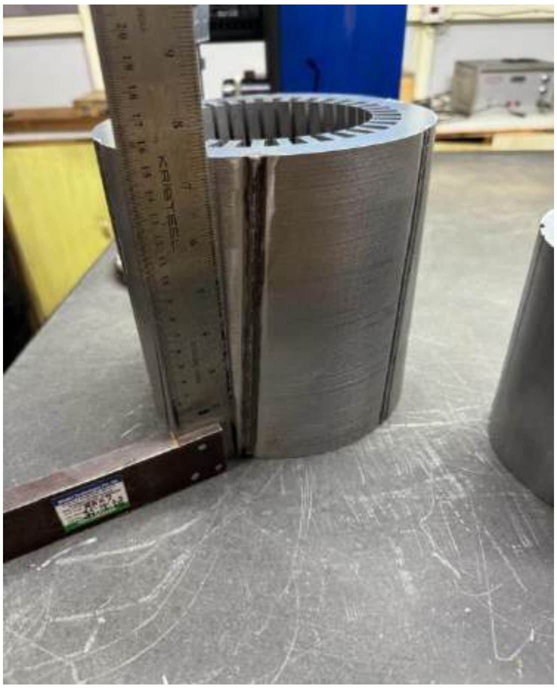
6. *Skewed Welding Path*

The welding machine's movement along the hypotenuse results in a skewed welding path due to the induced angular velocity. The resulting path of the weld head relative to the stator surface is skewed by an angle calculated as:

$$\phi = \omega t = \left(\frac{2\mu g\cos(\theta)}{L}\right) t^2$$



This Photo illustrates the Skew made on the stator with the normal wielding machine without The rig I made



This Photo shows how the skew is no longer 90 degrees and now has a slight angle to it hence the name skewed stator.

Strengths and limitations of my solution?

To evaluate the effectiveness of my proposed solution, I collaborated with the factory personnel and conducted multiple simulations. The objective was straightforward: to determine whether the adjustable triangular model consistently produced a skewed line on the stator at a precise angle.

After numerous simulations, the success of the model was evident, and the company was extremely pleased with the results.

This raises an important question: why are expensive robots typically employed to create angled skews on stators? The answer lies in two key factors: efficiency and adjustability. My solution was feasible for this particular case due to its small-scale application and the requirement for only a limited number of skewed stators. Robots, however, offer significantly faster production rates. Additionally, the triangle would necessitate regular lubrication to prevent excessive friction.

The second factor is adjustability. Although the angle of the triangular model can be modified to alter the skew angle, it is limited to this adjustment alone. In contrast, robots can vary the thickness of the skews, produce a wider range of designs, and achieve a much higher precision with an error margin of less than 0.003%. Therefore, for companies where skewed stators constitute a primary product, the investment in such robots is justified.

When designing a solution, it is crucial to consider the specific needs and applications. If the requirement had been for a rapid and highly precise device, I would not have devised a simple adjustable triangle. However, given the need for a relatively small batch of stators and the absence of a requirement for extremely precise angles, my design proved effective and resulted in substantial cost savings for the company.

The adjustable triangle model could be advantageous in contexts where cost savings are prioritized, and the production of skewed stators is not the primary focus. It is particularly suitable for facilities that need to produce thousands, rather than millions, of skewed stators, provided that a welding machine is already available. However, for operations solely focused on stator manufacturing, this approach may become laborious and less efficient compared to robotic solutions.

Concluding remarks

A significant takeaway from this experience working with the company, as well as spending time planning the implementation of the solution with skilled engineers, is the realization that simplicity is often overlooked. The engineers I collaborated with were exceptionally brilliant and knowledgeable. However, it was the straightforward thinking of a fresh perspective that illuminated the path to an effective solution. This experience underscores the beauty of design, which can manifest in various forms and often emerges from a combination of simplicity and elegance. The principle of Occam's Razor—stating that the simplest solution is usually the correct one—aptly applies here. It serves as a reminder that in the pursuit of innovation, the most effective answers are often found in the simplicity of design, harmonizing complex ideas with practical execution.

What initially appeared to be a daunting and costly endeavor transformed into a testament to the synergy between theoretical physics and practical problem-solving. Armed with this innovative solution, the company not only embraced the challenge of producing skewed stators but also established a cost-effective and ingenious method that defied conventional industry practices. This case study exemplifies how effective solutions can emerge from the integration of scientific principles and creative thinking when faced with industrial challenges.

In conclusion, Precision Stampings faced the challenge of producing skewed stators, a task traditionally requiring expensive robotic machinery. By implementing a simple mechanical solution—a triangular model that adjusted the welding machine's path—I successfully achieved the desired skew effect, resulting in significant economic savings for the company. This case highlights the value of exploring straightforward mechanical solutions to complex industrial problems. However, the approach has limitations in terms of efficiency and precision compared to advanced robotic systems, making it suitable for specific, small-scale applications rather than large-scale, high-precision manufacturing.

BIBLIOGRPAHY

https://precisionstampings.in

"Motor Characteristics." Electric Circuits, Iowa State University, https://iastate.pressbooks.pub/electriccircuits/chapter/motor-characteristics/.

Kim, K.H., Sim, D.J., and Won, J.S. "Analysis of skew effects on cogging torque and BEMF for BLDCM." IEEE IAS Conference Record, September 1991, Dearborn, MI, p. 191–197.

S. Nuzzo, M. Galea, C. Gerada and N. Brown, "A Fast Method for Modeling Skew and Its Effects in Salient-Pole Synchronous Generators," in IEEE Transactions on Industrial Electronics, vol. 64, no. 10, pp. 7679-7688, Oct. 2017, doi: 10.1109/TIE.2017.2694378.

https://www.geeksforgeeks.org/ac-motor/

J. Trevelyan, "Understandings of value in engineering practice," 2012 Frontiers in Education Conference Proceedings, Seattle, WA, USA, 2012, pp. 1-6, doi: 10.1109/FIE.2012.6462258. keywords: