

|CURIEUX|

ACADEMIC JOURNAL

January Issue

Part 2 Issue 46

Editing Staff

Chief-In-Editor

Caroline Xue

George Hasnah

Chief of Operations

Anshal Vyas

Assisting Editors

Olivia Li

Luzia Thomas

Madalyn Ramirez

Shefali Awasthi

Soumya Rai

Sarah Strick

Table Of Contents

Page 6: Therapeutic Analysis of Plerixafor on Relapsed AML Patients By Sanjay Vignesh

**Page 17: The Differentiation of Viral and Bacterial Pneumonia Using Deep Learning
By Arnav Dhar**

**Page 25: The Applications of CRISPR-Cas to Treating Genetic Diseases Caused By Repeat Expansions Such as Huntington's Disease, Friedrich Ataxia, and Fragile X Syndrome
By Gabriel Rui**

Page 40: Antigone and Thecla: Females In Martyrdom Literature By Jian Wang

Page 45: Autonomous Sensory Meridian Response (ASMR)'s Effectiveness on Lowering Stress and Tension By Kexin Lyu

Page 52: Homelessness: Working Towards Helping Poverty By Sonya Gupta

Page 60: Considering Biological Sex in The Development of Dementia By Hoa-Yen Trinh

Page 67: The Intelligence Divide: Examining Gender Roles in Modern Marriage By Katherien Lu

Page 73: Understanding Portugal's Drug Decriminalization Policy: Public Awareness and Perceptions of Law 30/2000 By Agastya Rai Gupta

Page 82: Exploring Contextual Learning Using Convolutional Neural Networks to Enhance Autonomous Driving By Soo Ban Yoon

Page 95: A Greedy Algorithm for Solving the Graph Coloring Problem By Matej Drenjancevic, Denis Ivanovic, Tomislav Rudec

Page 104: Optimizing Electric Vehicle Charging Station Placement in New York City: A Spatiotemporal Analysis of Traffic Patterns and Range Scenarios By Aaron Yin

Page 123: Exploring How Prompt Engineering Can Be Used Alongside Bloom's Taxonomy to Improve Learning Efficiency By Vardhan Madhavarapu

Page 132: Climate Change Has No Significant Negative Effects on the Suitable Habitats for Granular Poison Frog (*Oophaga granulifera*) According to Species Distribution Modeling By Juliana Careaga

Page 140: When will advancing technology make most of us unemployable? What should we do about this? By Davin Miao

Page 146: Making Classical Music Relevant: Applying Cultural Entrepreneurship to Contemporary Innovations in Classical Music Performance By Jiayi Sun

Page 157: Early Detection & Intervention in Bipolar Disorder: Role of Neuroimaging Biomarkers By Anika Denny

Page 172: Analyzing the Prominent Environmental Factors that Cause Forest Fires in the Algerian and Montesinho Forests By Sneha Narayan

Page 183: Foreign Relations In Maoist and Post-Maoist China By Winston Vance

Page 191: Model versus Defame: Chinese American Representation in American Cinema By Yushu Wu

Page 212: Challenges and Recent Advancements in Deep Space Craft Design Regarding Galactic Cosmic Radiation Protection By João Pedro O. Toledo Piza

Page 220: GenExpressivity: The impact of rights to freedom of expression on adolescents' mental health outcomes in the United States By Anya Khera

Page 230: The Impact of Artificial Intelligence and Data Centers on the Stock Prices of Utility Companies By Lucas Camacho

Page 235: A Retrospective of SPACs By Alexander Cao, Andrew Cun, Conrad Krol, Ryan Lung

Page 254: Growth Dynamics and Doubling Time of *Mycobacterium Smegmatis* Under Controlled Laboratory Conditions By Sophia Qin

Page 263: Impact of Technology on Achievement Motivation in Underserved Elementary Students During COVID-19 By Meghna Chellapilla

Page 272: Breaking Domestic Chains: A Comparative Analysis of Feminist Ideals in Wollstonecraft and Lu Xun By Xinyue Chen

Page 278: Exploring the Relationship between Land Surface Temperature and Vegetation Density in Delhi from 2014-2024 By Angad Singh Miglani

Page 289: Exploring the Relationship Between Sleep Disorders and Addiction to Alcohol or Nicotine By Gia Bhat-Roach

Page 311: The Harmonizing Role of Socially Beneficial Finance on Pakistan-India Relations and the Prospect of Peace By Sheryar Afridi

Page 320: Tesla's Use of Artificial Intelligence in Self-Driving Cars By Yitao Wang

Page 332: "Advancements in Artificial Organs: Self-Sensing Actuators for Enhanced Functionality" By Keerthi Chandran¹, Plinio Rodrigues de Oliveira Zanini²

Page 347: I've Been Freed: Understanding How Womanism Informs Black Women's Identity Development By Zoe' Bell

Page 366: A Comparison of LSTM and ARIMA For Stock Price Prediction By Muhammad Ayaan Ameen

Page 373: Is There Such A Thing as Too Much Democracy? By Michael Wu

Page 379: Case Study: Improving the efficiency and output of solar power systems of KCI Solar Array Project By Seonbeom (Joshua) Kim, Andrew Lin

Page 395: Navigating Vietnam's Legal Landscape Amidst AI and Digital Economy Growth By Dieu-Linh Do

Page 404: Small Modular Reactors: Advancing Electrical Grid Flexibility and Reliability By Janak Bhuta

Page 415: Effects of Socioeconomic Status on Food Choice and Health Written By Sonal Mohanty

Page 419: Exploring the Bidirectional Granger Causality Between Bitcoin and Nasdaq Index Volatilities By Jiajin(Buster) Xie

Page 428: AI-Driven Solutions for Reducing Healthcare Inequities: A Focus on Underrepresented Groups By Matthew Kim, R Ha

Page 438: Cultural Exchange in Nineteenth-Century Shanghai By Patrick Liu

Page 445: Protection or punishment: analyzing how policy shapes experiences of homelessness in New York City and Miami By Matthew Pollak

Page 461: Factors of Type 2 Diabetes in the 21st Century (United States and India) By Aleena Bakerywala

Page 470: Effects of Typically Developing Siblings on the Social and Emotional Development of Adolescents with Autism Spectrum Disorder By Navita Prabhakar

Page 481: Perceptions on Artificial Intelligence, Mental Health, Emotion Detection Technology By Saanvi Movva

Page 502: How Food Insecurity Affects African Children's Mental Health By Fabrizio Manganiello

Page 510: MicroRNA as a Potential Biomarker for Ovarian Cancer Diagnosis By Hoang Tue Anh Nguyen

Page 531: Strategic Financial Management for FC Barcelona: Developing Sustainable and Successful Business Strategies in a Challenging Economic Landscape By Soma Nakahori

Page 545: Navigating the Dual Edges: Social Media's Impact on Academic Performance and Psychological Well-being By Noah Kim

Page 555: The Influence of Ancient Greece and Rome on the United States Criminal Justice System By Yeonwoo Sung

Page 565: "Friendship Fallout: Teen Mental Health" By Syeda Rubab Bokhari

Page 568: Artificial Intelligence and the the Chrome Dino Game By Kevin Hsieh

Page 575: Efficient Denoising Method to Improve the Resolution of Satellite Images By: Jhanavi Hegde

Therapeutic Analysis of Plerixafor on Relapsed AML Patients By Sanjay Vignesh

Abstract

Acute myeloid leukemia (AML) is one of the most common and fatal types of leukemia, diagnosing an estimated twenty-thousand U.S. patients every year (Chen 1). Relapsed AML patient studies show that stem cell niches in combination with plerixafor and other chemotherapeutic medications can mobilize leukemia stem cells (LSCs) out of the hematopoietic stem cell niche, and can stop LSCs from being kept dormant and protected from chemotherapy (Villatoro 1). LSCs can relapse because the cancer metastasized, the cancer was unresponsive to the patient's first treatment, or the treatment couldn't eradicate all the leukemia cells. Plerixafor, which is approved by the US Food and Drug Administration is injected into patients for mobilizing LSCs out of the hematopoietic stem cell niche to make them vulnerable and unprotected to chemotherapy or other treatments (Villatoro 3). In newly diagnosed AML patients, a phase 1 clinical study was conducted, and 67% of participants achieved complete remission (Roboz 1). In elderly AML patients, a phase 1 clinical study was conducted, and 43% of participants achieved complete remission with 11 months of survival (Roboz 1). In relapsed AML patients, a phase 1/11 clinical study was conducted, and 39% of patients achieved complete remission with 2.5 fold mobilization of AML blasts (Uy 1). For this paper, research papers on the topics of plerixafor in combination with other drugs, stem cell niche, and acute myeloid leukemia were reviewed. Treatments such as chemotherapy and other drugs used in combination with stem cell niche interactions with plerixafor are able to treat and subdue acute myeloid leukemia relapse (Treatments for relapsed or refractory acute myeloid leukemia 1). Some implications of this paper will include, educating patients and students, adding on to the biomedical research done on acute myeloid leukemia, and expanding the general knowledge about therapeutics and plerixafor. This literature review goes over the background of AML, evaluates the details of the three clinical trials, and analyzes the trials and their implications.

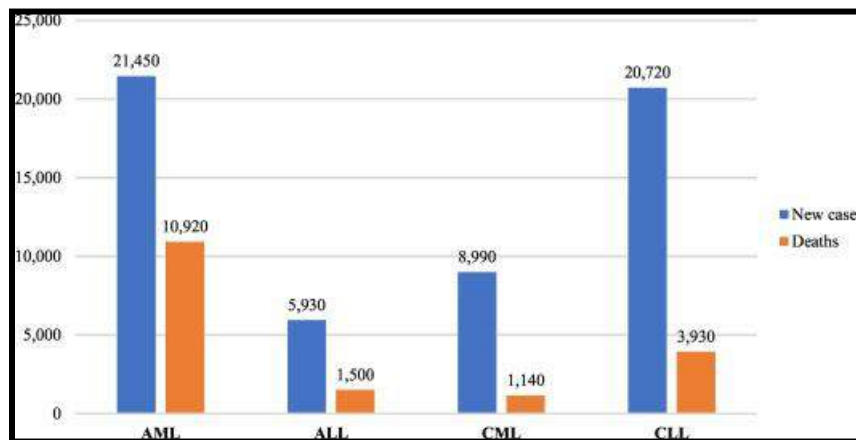


Fig 1: New cases and deaths for different types of Leukemia (Shallis 1)

Introduction

Acute myeloid leukemia (AML) is hematologic cancer that takes almost ten-thousand lives every year (*Chen 1*). AML is a type of blood cancer that will develop into different types of blood cells including red blood cells, white blood cells, and platelets (*Blood Cancers 1*). Leukemia can be categorized into four main types based on its progression and cell (*Buie 1*). Some types of leukemia are either lymphocytic meaning the cancer develops in the lymphocytes or myelogenous meaning the cancer develops in the granulocytes or monocytes (*Buie 1*). The other types of leukemia are either acute meaning immature cells called blasts increase and spread the disease quickly or chronic meaning certain blast cells are more mature which means that they grow slowly, and the disease progresses gradually (*Buie 1*). AML is different from other types of leukemias because it develops in the myeloid cells whereas other leukemias will develop in the lymphocytes (*Buie 1*). AML is also acute, meaning that it contains very immature blasts which cannot perform their functions, so the myeloblasts grow uncontrollably and the disease spreads quickly (*Buie 1*).

AML originates in the bone marrow's myeloblasts when the bone marrow cell's DNA is mutated to make the bone marrow cell divide which causes myeloblast production to come out of control (*Acute Myeloid Leukemia 1*). The bone marrow produces immature blasts that develop into leukemia white blood cells called myeloblasts (*Buie 1*). These irregular myeloblasts cannot function properly which causes them to divide in high numbers, and this makes them build up, crowd out, and interfere with healthy cells (*Buie 1*). There is no singular cause for AML, but there are a plethora of controllable and uncontrollable risk factors that contribute towards AML development (*Risk Factors for Acute Myeloid Leukemia (AML) 1*). Some of the main Risk Factors for Acute Myeloid Leukemia (AML) include having a family history of AML, a genetic syndrome, blood cancers like chronic myeloproliferative disorders, radiation exposure, exposure to chemotherapeutic drugs like alkylating agents, exposure to chemicals like benzene, smoking, being a biological male, and being in an older age range (*Risk Factors for Acute Myeloid Leukemia (AML) 1*). While some of the risk factors of AML are uncontrollable, multiple risk factors such radiation exposure, benzene exposure, and smoking are controllable. Radiation and benzene exposure can be avoided by staying away from refineries and manufacturing plants, and smoking can be avoided by substituting it for a natural dopamine-inducing activity (*Risk Factors for Acute Myeloid Leukemia (AML) 1*). Some symptoms for AML are extremely similar to common disease symptoms such as fever, bone pain, fatigue, pale skin, shortness of breath, frequent infections, easy bruising, and abnormal bleeding (*Acute Myelogenous Leukemia 2*). Traditional treatments for AML include chemotherapy, stem cell transplantation, and targeted therapy drugs (Treatments for relapsed or refractory acute myeloid 1). However, when chemotherapy is used to treat AML, it may cause the leukemia stem cells (LSCs) to relapse (Villatoro 1). In some cases, targeted therapy drugs are used with the stem cell niche to keep LSCs dormant during the relapse from chemotherapy (Villatoro 1). This stops the LSCs from growing and multiplying, which can effectively treat relapsed AML patients (*Treating Acute Myeloid Leukemia (AML) 1*).

Stem cell niches are microenvironments in a specific part of the body where stem cells are found (Ferraro 1). The first major component of stem cell niches are stromal support cells that include cell to cell adhesion molecules, they secrete soluble factors, which are found close to the stem cells (Ferraro 1). The next component is the extracellular matrix (ECM) proteins that act as an anchor for the stem cells, and consist of a mechanical scaffolding unit that transmits stem cell signals (Ferraro 1). The ECM is regulated by the stromal support cells. Blood vessels that carry nutritional support and systemic signals to the stem cell niche from other organs, and also recruit circulating stem cells to and from the niche (Ferraro 1). The last component are neural inputs that mobilize stem cells out of their niches, and integrate signals from other organ systems (Ferraro 1). Stem cells can receive input that can activate or repress genes and transcription programs (Ferraro 1). As a result of these interactions, the stem cells maintain a dormant state or are induced to self-renewal (Ferraro 1). Hematopoietic stem cell niches regulate hematopoietic stem cells (HSCs) which are immature cells that can develop into all types of blood cells (Boulais 1). Hematopoietic stem cell niches are also located in the bone marrow.

Chemotherapy is able to treat AML, but it promotes relapse in chemoresistant LSCs (Lanza 1). These LSCs may diversify and result in more aggressive forms of AML (Villatoro 1). Through in vivo models, it has been proven that the hematopoietic stem cell niche can help keep LSCs dormant (Villatoro 1). If the LSCs are dormant, then it cannot be eradicated which opens up the possibility of AML relapse. Scientists have discovered various ways of dealing with LSCs; whether it be LSCs eradication through anchoring interference or mobilization of LSCs out of the hematopoietic stem cell niche (Lanza 1). Scientists are able to use chemotherapeutic drugs, specifically plerixafor which is a selective inhibitor of CXCR4 mixed with other chemotherapeutic medications to mobilize LSCs out of the stem cell niche which allows them to be eradicated (Lanza 1). The chemokine receptor CXCR4 and its ligand CXCL12 are an essential part of the HSCs retention mechanism in the hematopoietic stem cell niche. Plerixafor consists of two cyclam rings with a methylene linker. It binds CXCR4 and CXCL12, and this interaction results in chemotaxis which is when immune cells can detect and migrate towards a site of chemical stimulus. The plerixafor gives a unique binding mode to CXCR4, and rapidly mobilizes HSCs within hours. One result of the plerixafor was in a phase 1/11 clinical study (NCT00512252), where plerixafor was used to treat relapsed AML patients in combination with mitoxantrone (mitoxantrone injection is used alone or with other medicines to treat advanced prostate cancer and acute nonlymphocytic leukemia), etoposide (medication that slows or stops the growth of cancer cells), and cytarabine (MEC), and patients experienced a 250 percent increase in mobilization of AML blasts with no preferential mobilization over healthy cells, and 39% achieved complete remission (Uy 1). Another result of the plerixafor was in a clinical study (NCT01352650), where plerixafor was tested for its potential to induce sensitization of LSCs to decitabine (chemotherapy drug that reduces growth of cancer cells), and the plerixafor triggered LSCs mobilization, but these cells persisted in the bone marrow during treatment which gave it a response rate of 43% and an overall survival of 11 months (Roboz 1). The main result of plerixafor that this paper will be focusing on was in a phase 1 clinical study (NCT00990054),

where plerixafor (CXCR4) was combined with cytarabine and daunorubicin to test the safety and tolerability of plerixafor, and the study resulted with 67% of the patients achieving complete remission (Roboz 1). In the introduction, this literature review discussed the background of AML and the overview on the three studies conducted on AML, and the paper will further discuss the methods used to obtain sources, the detailed results of the clinical studies, and an analysis of the three sources and their impact on their field of research.

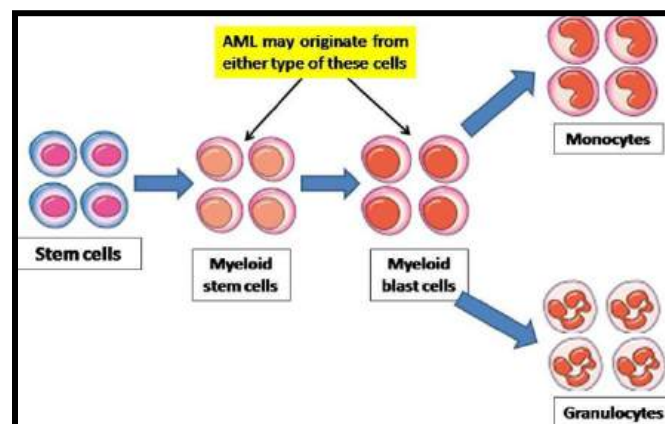


Fig 2: Origin of AML (Kabel 1)

Methods

The papers obtained for this literature review were found using search engines such as Google Scholar and Pubmed. The searches used to obtain these papers include “How do stem cells help against leukemia”, “Stem cells in regenerative medicines”, “Treating AML with stem cell niche”, “Treating cancer with regenerative medicines”, and “Regenerative medicine”. These papers were chosen due to their abundant amount of citations which proved their credibility, their time range which included papers from 2008 to 2021, and their relevance to certain topics which further enhanced the quality of this literature review. Certain papers were excluded due to low citations, which is anything lower than approximately one hundred citations, or for their lack of information about certain topics that would have improved the information in this literature review.

Results

The main result that this paper covers is a phase 1 clinical study (NCT00990054) where scientists combined plerixafor (CXCR4) with cytarabine and daunorubicin to test the safety and tolerability of plerixafor (Villatoro 3). A phase 1 clinical study means that scientists test the safety, side effects, dosage, and timing of a new treatment (Phase I Clinical Trial 1). The trial might also test the most effective way to distribute the new treatment whether it be by mouth, infusion to a vein, or injection (Phase I Clinical Trial 1). A phase 1 clinical study usually increases the dosage in small increments to find the highest dose that doesn't cause harmful or lethal side effects (Phase I Clinical Trial 1). Phase 1 clinical trials normally test on a small

amount of patients who have not been exposed to other treatments (Phase I Clinical Trial 1). Cytarabine is a chemotherapeutic medication that can be used alone or in combination with other chemotherapy drugs to treat certain types of leukemia including AML (Faruqi 1). Cytarabine can specifically cause a severe decrease in the amount of blood cells in your bone marrow (Faruqi 1). Daunorubicin comes in a powder form to be mixed with a liquid, and injected as a chemotherapeutic medication used with other chemotherapy drugs to treat AML by slowing or stopping the growth of cancer cells (Daunorubicin 1). Daunorubicin comes in a liquid solution form or a powder form to be mixed with a liquid for injecting intravenously (Daunorubicin 1). In the first trial, 23 patients received cytarabine on days 1–7, daunorubicin on days 1–3, and plerixafor on days 2–7 (Cancilla 6). This phase 1 clinical study (NCT00990054) yielded 67% of the patients achieving complete remission meaning all signs and symptoms of the cancer have disappeared (Cancilla 6). This trial differs from the others because it tests plerixafor in combination with cytarabine and daunorubicin, and it achieved a significantly higher remission rate than the other two studies (Villatoro 3).

Another result is a phase 1 clinical study (NCT01352650) where plerixafor was tested for its potential to induce sensitization of LSC to decitabine (Villatoro 3). The scientists conducting the trial hypothesized that combining the plerixafor with the decitabine would improve treatment outcomes through the mobilization of LSC, and through the alteration of the pharmacodynamics in decitabine (Roboz 1). Decitabine is a type of medication known as hypomethylation agents, and it works by helping the bone marrow produce normal blood cells and killing abnormal cells in the bone marrow (Decitabine Injection 1). Decitabine comes in a powder form which is added to a fluid, and slowly injected intravenously over three hours (Decitabine Injection 1). This trial's main inclusion criterias required that patients be at least 60 years old, ability to understand and willingness to sign a written informed consent document, and that patients haven't received prior chemotherapy, decitabine, or cytotoxic treatment (Roboz 1). The main exclusion criterias include prior treatment with decitabine, plerixafor, or the patient has an ongoing treatment for another malignancy (Roboz 1). The clinical trial studied each patient on an average of 4-6 months by testing their blood and bone marrow response to the treatment every month (Roboz 1). The results of the phase 1 clinical study (NCT01352650) were that plerixafor triggered LSC mobilization, but these cells persisted in the bone marrow during treatment which gave it a response rate of 43% and an overall survival of 11 months (Roboz 1).

The last result is a phase 1/11 clinical study (NCT00512252) where plerixafor was used to treat relapsed AML patients in combination with mitoxantrone, etoposide, and cytarabine (Uy 1). A phase 1/11 clinical study normally tests safety, side effects, dosage, and how well the cancer responds to the new treatment (Phase I/II Clinical Trial 1). In the phase 11 half of the trial, patients usually receive the highest dosage that didn't cause harmful effects from the phase 1 half of the trial, and combining both phases may allow research questions to be answered more efficiently (Phase I/II Clinical Trial 1). The scientists conducting this trial hypothesize that disrupting the interaction between AML blasts and the bone marrow environment with AMD3100, which is a bicyclam molecule that reversibly blocks CXCR4 from binding to SDF-1,

may sensitize AML blasts to the effect of chemotherapy (Uy 1). The results of the phase 1/11 clinical study (NCT00512252) were that patients experienced a 2.5-fold mobilization of AML blasts with no preferential mobilization over healthy cells, and 39% achieved complete remission (Villatoro 3).

Analysis

This section will discuss the benefits of each study and the differences and similarities between the studies in terms of patients, dosage, drugs, and delivery system. It will also cover the limitations of each clinical study including their inclusion and exclusion criteria, dosages, and where the trials were conducted.

The benefits of the main phase 1 clinical study (NCT00990054) include being able to make certain cells sensitive to killing by using plerixafor in combination with daunorubicin and cytarabine to help treat relapsed AML (Genzyme 1). The limitations of the inclusion criteria are having newly diagnosed AML which is defined as greater than 20% myeloblasts on the marrow aspirate or peripheral blood differential, having adequate cardiac function which is defined as greater than or equal to 40% left ventricular ejection fraction on echocardiography, MUGA scan, or similar radionuclide angiographic scan, toxicities from all prior treatments have reached baseline or Grade 1 prior to first dose of study drugs, and have adequate renal and hepatic function. The study was started in December 2009, and completed in March 2012 (Genzyme 1). The participants had to be at least 18 years-old, and at the most 70 years old. This study was conducted in multiple locations within the United States of America including Duarte, CA, Boston, MA, Ann Arbor, MI, St. Louis, MO, Rochester, NY, Cleveland, OH, Seattle, WA, and Houston, TX (Genzyme 1). The dosage used in the clinical study was 240 microgram/kilogram/dose which proceeded to increase dose levels for determination of the single-dose maximum tolerated dose (MTD), and during the trial there were no unacceptable dose limiting toxicities (Genzyme 1).

The benefits of the phase 1 clinical study (NCT01352650) include using a combination of plerixafor and decitabine as postremission therapy for older patients with AML to improve treatment outcomes through the mobilization of LSCs and alteration of the pharmacodynamics of decitabine (Roboz 1). Some inclusion criteria of the trial are having newly diagnosed AML patients which is defined as greater than 20% blasts on bone marrow, patients had to be at least 60 years old, AML patients with an antecedent hematologic disorder or myelodysplastic syndrome (MDS) are eligible for treatment provided that they have not received prior treatment with decitabine or prior cytotoxic treatment for AML, and AML patients with therapy-related myeloid neoplasms are eligible if they have not received chemotherapy for their primary malignancy or disorder for longer than 6 months (Roboz 1). This study was conducted by the principal investigator Gail J. Roboz, M.D. from the Weill Medical College of Cornell University, and her team on June 17, 2011, and it was completed on December 27, 2016 (Roboz 1). Another difference is that this trial focused testing on elderly patients meaning patients with at least 60 years of age while the other trials tested on patients with a wide span of ages.

The benefits of the phase 1/11 clinical study (NCT00512252) include enhancing the cytotoxic effect of chemotherapy for relapsed patients by disrupting the interaction between AML blast and the marrow microenvironment with plerixafor (Uy 1). Some limitations of the inclusion criteria are being diagnosed with AML whose criteria was primary refractory disease following at least 1 round of induction chemotherapy or first relapse and higher, being within the age range of 18-70 years-old, and adequate organ function. This study was conducted by the principal investigator Geoffrey L. Uy, M.D. from the Washington University School of Medicine, and his team in July of 2007, and it was completed by June 2010 (Uy 1).

The similarity between the three studies is that they all used plerixafor in combination with other chemotherapeutic drugs in their studies (Genzyme, Roboz, Uy 1). Another similarity between the trials is that they all had around the same number of participants with the main trial amassing the least with 36 participants, and the other two trials amassing 52 and 71 participants (Genzyme, Roboz, Uy 1). One difference is that the main study achieved 67% complete remission rate while the other studies achieved 39% and 43% complete remission rate respectively (Genzyme, Roboz, Uy 1). Another difference is that the main study administered an initial plerixafor dosage of 240 micrograms/kilogram/dose while the phase 1/11 trial administered doses of 80 mcg/kg/d and 160 mcg/kg/d in phase 1 with 240 mcg/kg/d in phase 11, and the other phase 1 trial administered 320 mcg/kg/d in cohort 1, 540 mcg/kg/d in cohort 2, and 810 mcg/kg/d in cohort 3 (Genzyme, Roboz, Uy 1). All sexes and genders were eligible with no acceptance of healthy volunteers (Genzyme, Roboz, Uy 1). The study that was conducted by the principal investigator Gail J. Roboz, M.D. is the most recent study being conducted from 2011-2016 in comparison to the other studies which were conducted from 2007-2010 and 2009-2012 (Genzyme, Roboz, Uy 1). The phase 1/11 clinical study (NCT00512252) is very distinguishable from the other two trials because it requires at least one round of induction therapy from the patients, while the other trials exclude patients if they've had any chemotherapy for a prior malignancy within a certain range of time (Genzyme, Roboz, Uy 1). Since the phase 1/11 study's goal was to enhance the cytotoxic effect of chemotherapy, the results being the remission rate may have been higher than what they should have been because the chemotherapy would have been in the patient's body more recently and for a longer time, therefore the chemotherapy's cytotoxic effect could have been greater than if the patients didn't experience chemotherapy for a certain period of time (Genzyme, Roboz, Uy 1). The final difference is that the main trial tested plerixafor in combination with cytarabine and daunorubicin while the other trials tested plerixafor in combination with other drugs such as decitabine or mitoxantrone, etoposide, and cytarabine (Genzyme, Roboz, Uy 1). This paper focuses on this particular study being phase 1 clinical study (NCT00990054) because it shows the most promise out of the three reviewed in this literature review meaning it had the most relevance to this paper's topic, and it also had the highest complete remission rate.

Conclusion

AML is a fast-growing blood cancer that develops and spreads quickly in the bone marrow's myeloid cells (*Buie 1*). Myeloblast production escalates due to bone marrow cell mutations (*Acute Myeloid Leukemia 1*). This causes the leukemia stem cells (LSCs) to grow uncontrollably (*Buie 1*). LSCs can be protected by microenvironments in the bone marrow called the hematopoietic stem cell niche (*Ferraro 1*). Some direct treatments for AML include chemotherapy and targeted therapy drugs, which in some situations are used with the stem cell niche to mobilize the LSCs out of the hematopoietic stem cell niche to sensitize them to chemotherapy, therefore eradicating AML relapse (*Villatoro 1*). Scientists have discovered various ways of dealing with LSCs (*Lanza 1*). This includes LSC eradication through anchoring interference or mobilization of LSCs out of the hematopoietic stem cell niche (*Lanza 1*). Scientists are able to use plerixafor which is a CXCR4 antagonist in combination with different chemotherapeutic medications to mobilize LSCs out of the hematopoietic stem cell niche (*Villatoro 1*). This paper reviews three different clinical trials that study the effects of plerixafor in combination with various drugs to treat AML relapse. The future directions of AML treatment should lean toward personalized therapies using combinations of mutation-targeted agents aiming at LSC eradication with reduced side effects, non-mutation-targeted novel agents directed to LSC eradication via interference with their anchoring to the stem cell niche (*Thakral 3*). The continued study of AML will increase patients' survival and remission rates for over twenty-thousand new patients every year (*Chen 1*).

Summary

This paper discusses three clinical trials and the use of plerixafor for mobilization of the Leukemia stem cells out of the hematopoietic stem cell niche to increase LSC sensitization to chemotherapy.

Works Cited

- Mahla, Ranjeet Singh. "Stem Cells Applications in Regenerative Medicine and Disease Therapeutics." *Stem Cells Applications in Regenerative Medicine and Disease Therapeutics*, International Journal of Cell Biology, 19 July 2016, www.hindawi.com/journals/ijcb/2016/6940283/.
- Villatoro, Alicia, et al. "Leukemia Stem Cell Release from the Stem Cell Niche to Treat Acute Myeloid Leukemia." *Frontiers*, Frontiers, 19 June 2020, www.frontiersin.org/articles/10.3389/fcell.2020.00607/full.
- Mao, Angela S., and David J. Mooney. "Regenerative Medicine: Current Therapies and Future ..." *Regenerative Medicine: Current Therapies and Future Directions*, PNAS, 23 Nov. 2015, www.pnas.org/doi/abs/10.1073/pnas.1508520112.
- Mansouri, Vahid, et al. "Recent Advances in Regenerative Medicine Strategies for Cancer Treatment." *Biomedicine & Pharmacotherapy*, Elsevier Masson, 3 July 2021, www.sciencedirect.com/science/article/pii/S0753332221006570.
- Chan, Wai-In, and Brian J.P. Huntly. "Leukemia Stem Cells in Acute Myeloid Leukemia." *Seminars in Oncology*, W.B. Saunders, 8 Aug. 2008, www.sciencedirect.com/science/article/abs/pii/S0093775408001139.
- Roboz, Gail J. "Decitabine and Plerixafor in Elderly Acute Myeloid Leukemia (AML)." *Clinicaltrials.Gov*, National Library of Medicine, 12 May 2011, classic.clinicaltrials.gov/ct2/show/NCT01352650.
- Uy, Geoffrey L. "Safety and Tolerability of Plerixafor in Combination with Cytarabine and Daunorubicin in Patients with Newly Diagnosed Acute Myeloid Leukemia- Preliminary Results from a Phase I Study | Blood | American Society of Hematology." *Safety and Tolerability of Plerixafor in Combination with Cytarabine and Daunorubicin in Patients with Newly Diagnosed Acute Myeloid Leukemia*, Ash Publications, 18 Nov. 2011, ashpublications.org/blood/article/118/21/82/83106/Safety-and-Tolerability-of-Plerixafor-in.
- M,; Lanza F;Gardellini A;Laszlo D;Martino. "Plerixafor: What We Still Have to Learn." *Plerixafor: What We Still Have to Learn*, U.S. National Library of Medicine, 15 Oct. 2014, pubmed.ncbi.nlm.nih.gov/25315706/.
- A,; Thakral D;Gupta R;Khan. "Leukemic Stem Cell Signatures in Acute Myeloid Leukemia- Targeting the Guardians with Novel Approaches." *Stem Cell Reviews and Reports*, U.S. National Library of Medicine, 28 Mar. 2022, pubmed.ncbi.nlm.nih.gov/35412219/.
- "Key Statistics for Acute Myeloid Leukemia (AML)." *Key Statistics for Acute Myeloid Leukemia (AML)*, American Cancer Society, www.cancer.org/cancer/types/acute-myeloid-leukemia/about/key-statistics.html. Accessed 17 Jan. 2024.
- Canadian Cancer Society / Société canadienne du cancer. *Treatments for Relapsed or Refractory Acute Myeloid Leukemia*, Canadian Cancer Society, cancer.ca/en/cancer-information/cancer-types/acute-myeloid-leukemia-aml/treatment/relapsed-or-refractory. Accessed 10 Jan. 2024.

“Blood Cancers.” *Blood Cancers*, American Society of Hematology, www.hematology.org/education/patients/blood-cancers Accessed 1 Jan. 2024.

Buie, Larry, “What Are the Types of Leukemia?” *Types of Leukemia*, Memorial Sloan Kettering Cancer Center, www.mskcc.org/cancer-care/types/leukemias/types. Accessed 8 Jan. 2024.

“Acute Myeloid Leukemia.” *Acute Myeloid Leukemia (AML) | LLS.Org*, Leukemia and Lymphoma Society, www.lls.org/leukemia/acute-myeloid-leukemia. Accessed 18 Jan. 2024.

“Risk Factors for Acute Myeloid Leukemia (AML).” *Risk Factors for Acute Myeloid Leukemia (AML)*, American Cancer Society, www.cancer.org/cancer/types/acute-myeloid-leukemia/causes-risks-prevention/risk-factors.html. Accessed 7 Feb. 2024.

Saleem, Tabinda. “Daunorubicin.” *StatPearls [Internet].*, U.S. National Library of Medicine, 28 Aug. 2023, www.ncbi.nlm.nih.gov/books/NBK559073/#:~:text=Daunorubicin%20is%20an%20anthracycline%20antibiotic,both%20acute%20and%20chronic%20leukemias.

Faruqi, Arjumand. “Cytarabine.” *StatPearls [Internet].*, U.S. National Library of Medicine, 8 Aug. 2023, www.ncbi.nlm.nih.gov/books/NBK557680/.

“Decitabine Injection: Medlineplus Drug Information.” *MedlinePlus*, U.S. National Library of Medicine, medlineplus.gov/druginfo/meds/a608009.html. Accessed 23 Jan. 2024.

Genzyme. *Clinicaltrials.Gov*, National Library of Medicine, clinicaltrials.gov/study/NCT00990054?cond=AML&term=Acute+Myeloid+Leukemia&intr=Plerixafor&rank=4. Accessed 12 Jan. 2024.

Ferraro, Francesca, et al. “Adult Stem Cells and Their Niches.” *Advances in Experimental Medicine and Biology*, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/21222205/. Accessed 9 Feb. 2024.

Chen, Pengyin, et al. “Global, National, and Regional Burden of Acute Myeloid Leukemia among 60-89 Years-Old Individuals: Insights from a Study Covering the Period 1990 to 2019.” *Frontiers in Public Health*, U.S. National Library of Medicine, 11 Jan. 2024, [www.ncbi.nlm.nih.gov/pmc/articles/PMC10808630/#:~:text=Approximately%2080%25%20of%20new%20cases,death%20\(6%E2%80%9338\)](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC10808630/#:~:text=Approximately%2080%25%20of%20new%20cases,death%20(6%E2%80%9338)).

“Phase I Clinical Trial.” *Comprehensive Cancer Information - NCI*, NCI Dictionary of Cancer terms, www.cancer.gov/publications/dictionaries/cancer-terms/def/phase-i-clinical-trial. Accessed 13 Mar. 2024.

Faruqi, Arjumand. “Cytarabine.” *StatPearls [Internet].*, U.S. National Library of Medicine, 8 Aug. 2023, www.ncbi.nlm.nih.gov/books/NBK557680/.

“Phase I/II Clinical Trial.” *Comprehensive Cancer Information - NCI*, NCI Dictionary of Cancer terms, www.cancer.gov/publications/dictionaries/cancer-terms/def/phase-i-ii-clinical-trial. Accessed 10 Mar. 2024.

Boulais, Philip E. Making Sense of Hematopoietic Stem Cell Niches, ScienceDirect, 14 Dec. 2020, www.sciencedirect.com/science/article/pii/S0006497120412297.

Cancilla, Daniel, et al. "Targeting CXCR4 in AML and All." *Frontiers*, *Frontiers*, 29 July 2020, www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2020.01672/full.

Fig 1: Shallis, Rory M, et al. *Epidemiology of Acute Myeloid Leukemia*. 29 Apr. 2019.

Epidemiology of Acute Myeloid Leukemia: Recent Progress and Enduring Challenges, ScienceDirect,

<https://www.sciencedirect.com/science/article/abs/pii/S0268960X18301395#preview-section-references>. Accessed 23 July 2024.

Fig 2: Kabel, Ahmed M., et al. 'Acute Myeloid Leukemia: A focus on Risk Factors, Clinical Presentation, Diagnosis and Possible Lines of Management.' *Journal of Cancer Research and Treatment* 5.2 (2017): 62-67.

The Differentiation of Viral and Bacterial Pneumonia Using Deep Learning

By Arnav Dhar

Abstract

This research investigates the use of Convolutional Neural Networks (CNNs) to classify X-ray scans as either bacterial or viral pneumonia. The goal is to support pulmonary healthcare in underprivileged regions by automating diagnostic tasks. Using Keras and TensorFlow, a CNN was trained on a dataset of 5,856 pediatric chest X-rays from children aged one to five, sourced from the Guangzhou Women and Children's Medical Center. After evaluating multiple architectures, the final model, based on Inception ResNet v2, achieved a 76.55% test accuracy while optimizing training efficiency through a minimal number of epochs. Key strategies, such as dropout, early stopping, L2 regularization, and data augmentation, were implemented to mitigate overfitting. This study demonstrates the potential of CNNs for pneumonia classification and highlights the role of deep learning in medical imaging, particularly in resource-constrained settings.

Introduction

Pneumonia is a serious respiratory infection affecting the lungs, causing inflammation, fluid accumulation, and breathing difficulties. While pneumonia impacts individuals of all ages, modern medical advancements have significantly reduced its threat, primarily affecting vulnerable populations like children and the elderly. In underdeveloped regions, pneumonia remains the leading cause of death among children (Why Pneumonia Is So Dangerous). Despite improvements in living conditions, vaccinations, and nutrition in developed countries, pneumonia continues to be a critical issue in developing regions due to limited healthcare resources (Dadonaite & Roser; Scott et al.).

Diagnosing pneumonia typically involves chest X-rays, blood tests like Complete Blood Count (CBC) and Pulse Oximetry, and clinical evaluations (Diagnosis of Pneumonia). However, distinguishing between bacterial and viral pneumonia is essential, as they require different treatments. For example, bacterial pneumonia often necessitates antibiotics, while viral pneumonia does not (Viral vs. Bacterial Pneumonia). Misdiagnosis can lead to inappropriate treatments, contributing to antimicrobial resistance or insufficient patient care, particularly in resource-constrained settings (Viral vs. Bacterial Pneumonia).

The advent of artificial intelligence (AI) and machine learning (ML) has paved the way for advancements in medical imaging diagnostics. Convolutional Neural Networks (CNNs) have proven successful in a range of medical applications, such as identifying cancer metastasis (Wang et al.) and fracture classification (Beyaz). CNNs, known for their ability to extract meaningful patterns from images, offer a promising solution for classifying pneumonia types based on chest X-rays.

This research focuses on developing a CNN model using the Inception ResNet v2 architecture to classify pediatric chest X-rays as either bacterial or viral pneumonia. The model

is trained on a dataset of 5,856 chest X-rays from children aged one to five, sourced from the Guangzhou Women and Children’s Medical Center (Mooney). The aim is to assess whether a deep learning-based solution can enhance pneumonia classification accuracy, particularly in under-resourced healthcare settings where timely diagnosis is critical to patient outcomes.

Methods

This model uses the Inception ResNet v2 architecture to classify X-ray images as either viral or bacterial pneumonia, as shown in Figure 1.

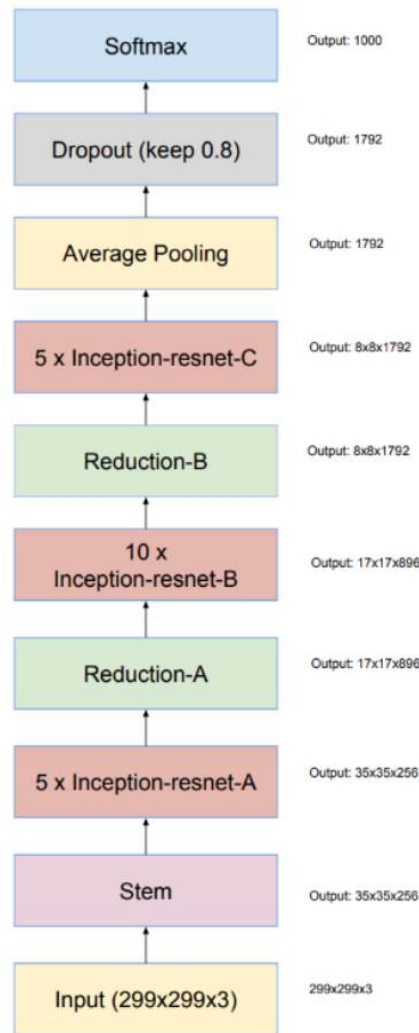


Fig 1: Using Inception ResNet v2 to classify X-ray images

While the base architecture is effective, hyperparameter tuning is essential to optimize performance for this specific task. A grid search was conducted to identify the most impactful hyperparameters and their effects. Table 1 shows a condensed version of the significant findings, followed by an explanation of the best hyperparameter combination.

Table 1
Summary of hyperparameter tuning for proposed model

Data Table 2: Summary of hyperparameter tuning for proposed model

Optimizer	Config	Dropout	Learning rate	Decay steps	Decay factor	Decay	Accuracy	Observations
Adam	1	0.4	1.00E-05	N/A	N/A	none	70.23%	
Adam	2	0.4	1.00E-04	N/A	N/A	none	72.44%	<i>Shows lower learning rate improves accuracy</i>
Adam	3	0.4	1.00E-03	N/A	N/A	none	68.44%	<i>A learning rate that is too high decreases accuracy</i>
Adam	4	0.4	5.00E-04	N/A	N/A	none	73.11%	
Adam	5	0.4	8.00E-04	N/A	N/A	none	74.32%	<i>Learning rate in between 0.001 and 0.0001 works best</i>
Adam	6	0.4	8.00E-04	N/A	0.95	(BIE) built-in exponential	74.22%	
Adam	7	0.4	8.00E-04	N/A	N/A	IBE+linear	70.67%	
Adam	8	0.4	8.00E-04	10	0.95	IBE+step	71.12%	
SGD	9	0.4	8.00E-04	10	0.5	step	75.35%	<i>SGD's step decay improves accuracy, but overfits</i>
SGD	10	0.8	8.00E-04	10	0.5	step	76.55%	<i>Inception ResNet v2's recommended dropout works best</i>
**for all tests, batch size = 30 and epoch count = 50								

Two optimization algorithms were tested: Stochastic Gradient Descent (SGD) and Adam. While Adam adapts the learning rate during training by combining the strengths of AdaGrad and RMSProp, it wasn't as effective for this model. SGD, which lacks built-in decay but allows for custom adjustments, proved more suitable and was ultimately chosen as the optimizer. The best performance was achieved using SGD with custom step decay, which improved test accuracy.

The dropout rate played a key role in regularization. While Adam worked better with a 0.4 dropout rate, testing with SGD revealed that a higher dropout rate of 0.8 was optimal. Additionally, L2 regularization was tested, and a higher value of 0.1 (compared to Adam's optimal 1e-5) yielded better results with SGD. This helped the model avoid overfitting by controlling the weights and biases more effectively.

An unusually low learning rate worked best for this model. While typical CNNs use a learning rate between 0.1 and 0.001, this model required a rate as low as 0.0001, with 0.0008 being the final selection after fine-tuning.

Batch size and epoch count had minimal impact on accuracy. A batch size of 30 was found to work best, with initial tests using 30 epochs to assess the model's capability without overburdening computational resources. After further research, it was determined that increasing the epoch count above 50 led to overfitting or performance plateauing, making 50 epochs optimal for this task. Tracking the model's accuracy in relation to training time helped to ensure an optimal train-test split and prevented overfitting to the dataset.

Table 2
The effects of different optimizers and learning rate schedules on accuracy

Data Table 1: The effects of different optimizers and learning rate schedules on accuracy

Optimizer	Decay Type	Avg. Test Accuracy	Avg. Test Loss
Adam	Built-In Exponential (BIE)	84.22%	1.77
Adam	BIE & Step Decay	72.11%	1.92
Adam	BIE & Linear Decay	78.92%	1.33
SGD	Exponential	89.47%	0.34
SGD	Step Decay	89.86%	0.35
SGD	Linear Decay	84.23%	0.87

****For all tests:**

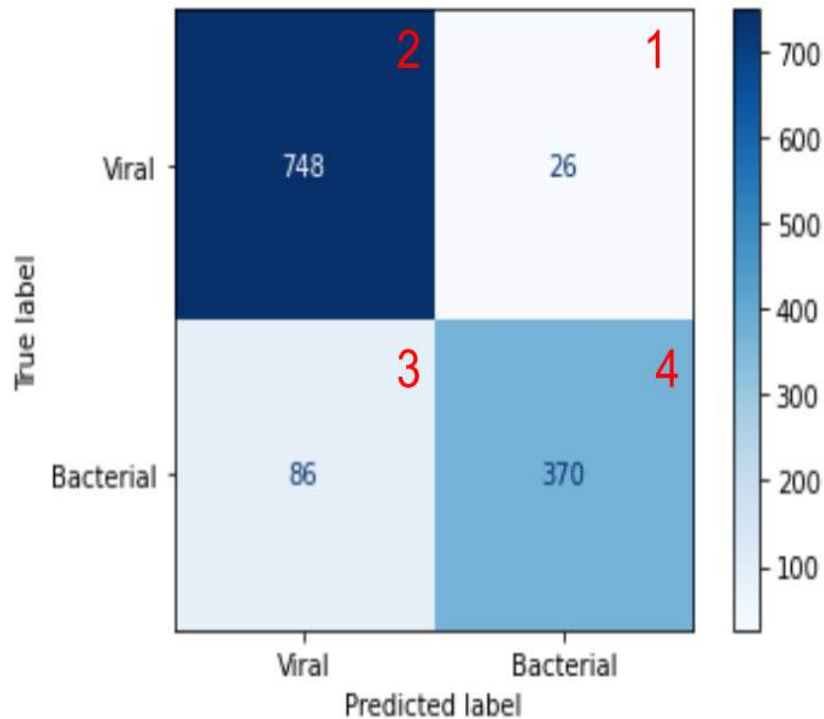
- Initial Learning Rate = 0.008
- Epoch Count = 50
- Batch Size = 30

Results and Discussion

The final CNN model achieved a test accuracy of 76.55%, with an area under the curve (AUC) of 0.94, indicating strong performance in distinguishing between bacterial and viral pneumonia. Although the results are promising, they highlight areas for improvement, particularly in bacterial pneumonia classification. The visual similarities between bacterial and viral pneumonia on X-rays likely contributed to the model's difficulty in accurately distinguishing the two.

Validation loss decreased consistently after an initial spike during the third epoch, mirroring the trend of training loss. This fluctuation is common in deep learning models as they adjust internal parameters during early training stages (Rendyk). Validation accuracy, however, plateaued at 0.75 after initially reaching 0.86, indicating that the model's generalization capacity had reached its limit. The model's reliance on dropout and L2 regularization helped reduce overfitting, though further refinements to these methods may yield better results in future iterations (Li).

Fig 2: Predicted label



The confusion matrix sheds light on the model’s strengths and weaknesses. The model correctly classified viral pneumonia with 97% accuracy, with only 26 false bacterial predictions out of 1,230 viral cases. This suggests that the model has a strong capacity for detecting viral pneumonia. However, its performance in bacterial pneumonia classification was slightly less reliable, underscoring the challenge posed by visual similarities between the two conditions (Cohen et al.). Misclassifications of bacterial pneumonia are particularly concerning in clinical settings, where prompt and accurate diagnosis is essential for appropriate treatment.

The consistent gap between training and testing accuracy, with training accuracy lagging 7-8% behind, suggests that regularization methods such as dropout and L2 regularization were effective in curbing overfitting during training but are not applied during testing (Rendyk). Additionally, the use of a predefined test/train split may have resulted in an easier test set, artificially inflating the test accuracy. A more randomized dataset split could mitigate this effect in future studies.

The choice of optimization algorithm also influenced the model’s performance. Stochastic Gradient Descent (SGD) with custom step decay outperformed the Adam optimizer, possibly because the manual control of learning rate decay allowed for more precise adjustments as training progressed. This result highlights the importance of fine-tuning hyperparameters to prevent overfitting while maximizing model performance (Rosebrock).

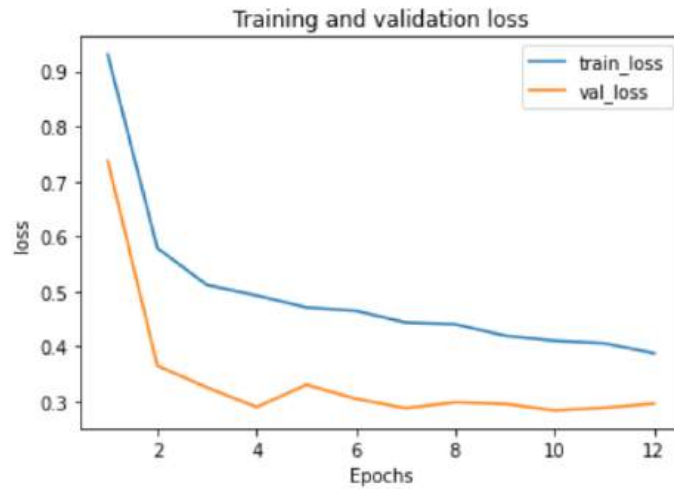


Fig 3: Training and validation loss

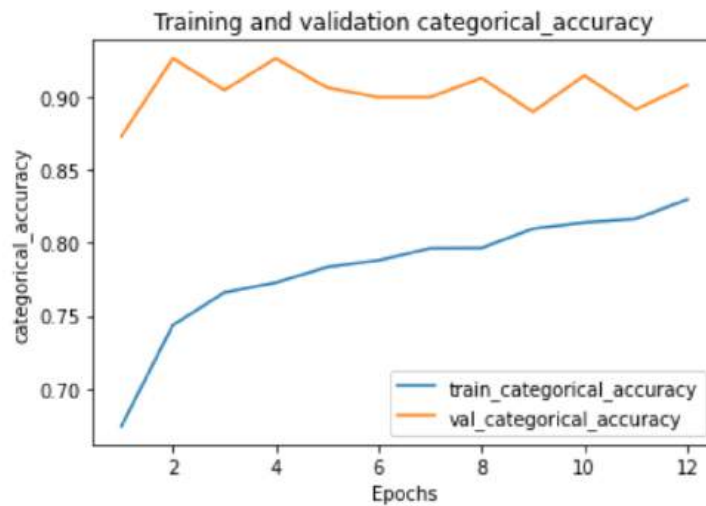


Fig 4: Training and validation categorical accuracy

Improvements to the model are planned for future development. One potential enhancement involves combining the train and test images into single ‘Viral’ and ‘Bacterial’ folders, allowing Keras to randomly split the dataset, ensuring no bias in test difficulty. Another improvement would incorporate the unused ‘Normal’ images to create a two-tiered model: first distinguishing pneumonia from healthy cases, then using the current model to classify the type of pneumonia. This approach is still being refined but could significantly improve both accuracy and functionality.

Conclusion

In conclusion, the Inception ResNet v2 model developed for classifying bacterial and viral pneumonia from X-ray images demonstrated solid performance, achieving an accuracy of 76.55% and an AUC of 0.94. While the model excelled in distinguishing viral pneumonia, further refinement is needed to improve bacterial pneumonia classification. The use of regularization techniques, such as dropout and L2 regularization, helped mitigate overfitting, but discrepancies between training and testing accuracy suggest opportunities for improvement, particularly in data handling and model tuning. Future enhancements, including a more randomized train/test split and the incorporation of healthy cases into a two-tier classification system, could significantly increase the model's effectiveness. These findings highlight the potential of deep learning models in medical imaging and their ability to assist healthcare professionals, particularly in under-resourced areas.

Works Cited

- Beyaz, Salih. "A Brief History of Artificial Intelligence and Robotic Surgery in Orthopedics & Traumatology and Future Expectations." Joint Diseases and Related Surgery, U.S. National Library of Medicine, 2020, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7607941/>
- Cohen, Joseph Paul, et al. "Predicting Covid-19 Pneumonia Severity on Chest X-Ray with Deep Learning." Cureus, Cureus, 28 July 2020, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7451075/>
- Dadonaite, Bernadeta, and Max Roser. Pneumonia, Our World In Data, 4 Nov. 2018, <https://ourworldindata.org/pneumonia#:~:text=Mortality%20from%20pneumonia%20i%20highest,the%20number%20deaths%20from%20pneumonia>
- "Diagnosis of Pneumonia." National Heart Lung and Blood Institute, U.S. Department of Health and Human Services, 2020, <https://www.nhlbi.nih.gov/health/pneumonia/diagnosis#:~:text=A%20chest%20X%20Ray%20is%20enough%20oxygen%20into%20your%20blood>
- Li, Katherine (Yi). "How to Choose a Learning Rate Scheduler for Neural Networks." How to Choose a Learning Rate Scheduler for Neural Networks, Neptune AI, 14 Nov. 2022, <https://neptune.ai/blog/how-to-choose-a-learning-rate-scheduler>
- Mooney, Paul. "Chest X-Ray Images (Pneumonia)." Kaggle, Kaggle, 24 Mar. 2018, <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>
- Rendyk. "Tuning the Hyperparameters and Layers of Neural Network Deep Learning." Tuning the Hyperparameters and Layers of Neural Network Deep Learning, Analytics Vidhya, 26 May 2021, <https://www.analyticsvidhya.com/blog/2021/05/tuning-the-hyperparameters-and-layers-of-neural-network-deep-learning/>
- Rosebrock, Adrian. "Keras Learning Rate Schedules and Decay." PyImageSearch, PyImageSearch, 28 Jan. 2022, <https://pyimagesearch.com/2019/07/22/keras-learning-rate-schedules-and-decay/>
- Scott, J Anthony G, et al. "Pneumonia Research to Reduce Childhood Mortality in the Developing World." The Journal of Clinical Investigation, vol. 118, no. 4. American Society for Clinical Investigation, Apr. 2008, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2276784/>
- "Viral vs. Bacterial Pneumonia: Understanding the Difference." Pfizer, 2022, https://www.pfizer.com/news/articles/viral_vs_bacterial_pneumonia_understanding_the_difference
- Wang, Dayong, et al. "Deep Learning for Identifying Metastatic Breast Cancer." ArXiv.org, Cornell University, 18 June 2016, <https://arxiv.org/abs/1606.05718>
- "Why Pneumonia Is so Dangerous for the Elderly." Northwest Pulmonary & Sleep Medicine, 2021, <https://pulmonarysleep.com/blog/why-pneumonia-is-so-dangerous-for-the-elderly/>

The Applications of CRISPR-Cas to Treating Genetic Diseases Caused By Repeat Expansions Such as Huntington's Disease, Friedrich Ataxia, and Fragile X Syndrome

By Gabriel Rui

Abstract

About 1 in 3,000 people are affected by a genetic disease caused by repeat expansions in their DNA. Historically, there have been no treatments for these diseases other than symptom management. However, with the advancement of gene editing, most recently CRISPR-cas, new tools, and methods are now available to address these diseases. The three prominent repeat expansion diseases are Huntington's Disease, Friedrich Ataxia, and Fragile X Syndrome. Here, I will discuss how non-homologous end joining, homology-directed repair, exon skipping, and dCas9 can be used for treating repeat expansion diseases, along with the primary methods of delivering the compounds. Studies on all of these diseases yielded compelling results. While gene editing to treat Fragile X Syndrome is currently among the most well-researched, Cas9 shows promise as an effective treatment method for all three diseases.

Introduction

Genetic disorders occur in around 2-5% of all live births and can cause a variety of severe symptoms.¹ There are numerous genetic diseases that can be categorized by the type of gene mutation that causes them. Three highly prevalent genetic diseases, Huntington's Disease, Fragile X Syndrome, and Friedrich Ataxia, are all caused by excessive repeat expansions in the genome. Each of these genetic diseases has its unique causes and symptoms. However, they all cause devastating symptoms in those afflicted. While these genetic diseases can appear de novo, because of their genetic basis, they can also be passed down family lines with varying severity and no foreseeable end.

Currently, there are no cures for any of these diseases. Instead, all medical professionals had been able to do was provide therapies to manage the symptoms. The techniques used for gene sequencing have long been available to scientists, and using this, they have been able to identify the causes of genetic diseases but could not do anything to fix them. What has changed the status quo is the recent developments in the field of gene editing. Scientists can now use the newest gene-editing techniques to target specific parts of the genome to be deleted entirely or to be replaced with a nucleotide sequence of their choosing. In the past years, scientists have been experimenting with a multitude of different techniques for gene editing, such as Restriction Enzymes, Transcription Activator-Like Effector Nucleases (TALENs), Zinc Finger Nucleases (ZFNs), RNA Interference (RNAi), and Meganucleases. Of the different techniques tested, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and the CRISPR-associated protein (cas) were generally the most efficient and effective. Currently, cas9 is the most widely used due to its ease of use.

Repeat expansions in DNA

Repeats are highly common in DNA; however, due to the nature of repeat expansions, they also cause nearly 30 different heritable genetic diseases.² Repeats in DNA can have varying numbers of nucleotides in each repeat, ranging from 3 to 12 nucleotides, but most genetic diseases caused by repeat expansions have trinucleotide repeats.² These repeat expansions can be acquired in one of two ways: either through fixed Mendelian inheritance or dynamic repeat expansions.

Due to the genetic nature of repeats in DNA, the dynamic repeat expansions causing various genetic diseases all follow similar patterns. One such pattern is that repeat expansion mutations all originate from normally polymorphic repeats. Furthermore, dynamic repeat expansions, or repeat sequences changing in length when passed from generation to generation, on the higher end of the normal range have a higher probability of further expanding into the range of genetic diseases.³

CRISPR-Cas

CRISPR was first discovered in the DNA of *Escherichia coli* (*E. coli*) bacteria.⁴ Following that discovery, scientists found CRISPR originated in prokaryotes as a natural defense mechanism against bacteriophages. Prokaryotes use CRISPR to retain a memory of foreign nucleotide sequences so that the bacteria can recognize and destroy any foreign substances with matching sequences.

Scientists have discovered how to use CRISPR-Cas from prokaryotes as a means for gene editing as it has both the necessary binding and cutting capabilities. Guide RNA (gRNA) can be used to direct the cas9 protein to target sites where double-strand breaks are created. When these breaks are created, the cell initiates its natural repair systems for DNA, two of the most prevalent being non-homologous end-joining (NHEJ) and homology-directed repair (HDR). In these processes, the double-strand breaks are repaired, and nucleotides are rebuilt. Scientists can then use an inserted template designed with the sequence of their choice to determine what DNA is put into the genome during the repair process.

Even as one of the most effective means of gene editing, CRISPR-Cas9 still has some downsides. One of its flaws lies in the functionality of gRNA. Because gRNA is usually about 20 bp long, it may recognize the wrong sequences. Another factor that can prove difficult when using CRISPR-Cas over other tools is that long repeats are more difficult to delete, though possible to deal with.

With all of the foundational scientific knowledge in the field of genetics that has been accumulated over time, the newest discoveries in gene editing tools offer a golden opportunity to address one of the most challenging healthcare obstacles. As expected, the compelling nature of these discoveries and the problems they solve have quickly garnered attention, opening up a new field of scientific research. In this paper, I will delve into the field of genetic diseases and their possible solutions. I will begin by discussing the basis of genetic diseases, focusing on Huntington's Disease, Fragile X Syndrome, and Friedrich's Ataxia. Then, I will review the theory behind possible genetic treatments revolving around using the CRISPR-Cas system.

Following the descriptions of the CRISPR-Cas techniques, I will review the latest treatments and clinical trials regarding the three repeat expansion genetic diseases, with an additional overview of clinical trials for Duchenne Muscular Dystrophy (DMD). DMD is not a repeat expansion disease; however, it is one of the genetic diseases with the most clinical trials and data on, so it serves as a compelling precedent for gene therapy targeting all genetic diseases. Finally, I will conclude this paper with a brief discussion regarding bioethics and my predictions for future developments in this field.

Repeats in genetic diseases

More than 40 diseases are caused by repeat expansions in the human genome.⁵ Of these 40 diseases, many affect the nervous system. This paper focuses on three of those diseases: Huntington's Disease, Fragile X Syndrome, and Friedrich Ataxia.

Huntington's Disease

Huntington's Disease is a neurodegenerative genetic disease that typically appears in people ages 30 to 50 years old and is known to cause nerve cells in the brain to decay over time, resulting in declining mobility, thinking, mental health, and usually death within 10 to 30 years of diagnosis. Juvenile Huntington's disease can also develop in people younger than 20, causing different symptoms and often death within 10 to 15 years of diagnosis.

Huntington's Disease is an autosomal dominant disease caused by CAG repeats in the Huntingtin gene (HTT) on chromosome 4.⁶ The Huntingtin gene codes for the huntingtin protein, which is necessary in the post-embryonic period and supports synaptic function.⁷ Excessive CAG repeats will result in a dysfunctional Huntingtin protein, making the number of CAG repeats highly critical. A standard genome is considered to have 26 or fewer CAG repeats. Those with 27 to 35 CAG repeats do not have any risk of developing Huntington's. However, due to the nature of repeats losing stability as they increase in length, those people have a high chance of having children with Huntington's. From 36 to 39 repeats, one may or may not develop Huntington's Disease, and beyond 40 repeats, one will develop symptoms of Huntington's Disease. Furthermore, having 60 or more repeats leads to the development of Juvenile Huntington's Disease.⁶ Huntington's Disease also appears to follow other hereditary patterns in addition to the size of the repeat expansion increasing instability. Research has revealed that in more than 80% of hereditary transmissions when Huntington's Disease is passed on maternally, the repeat expansion is reduced by a few repeats, but when it is passed on paternally, it tends to further expand by a few repeats.⁶

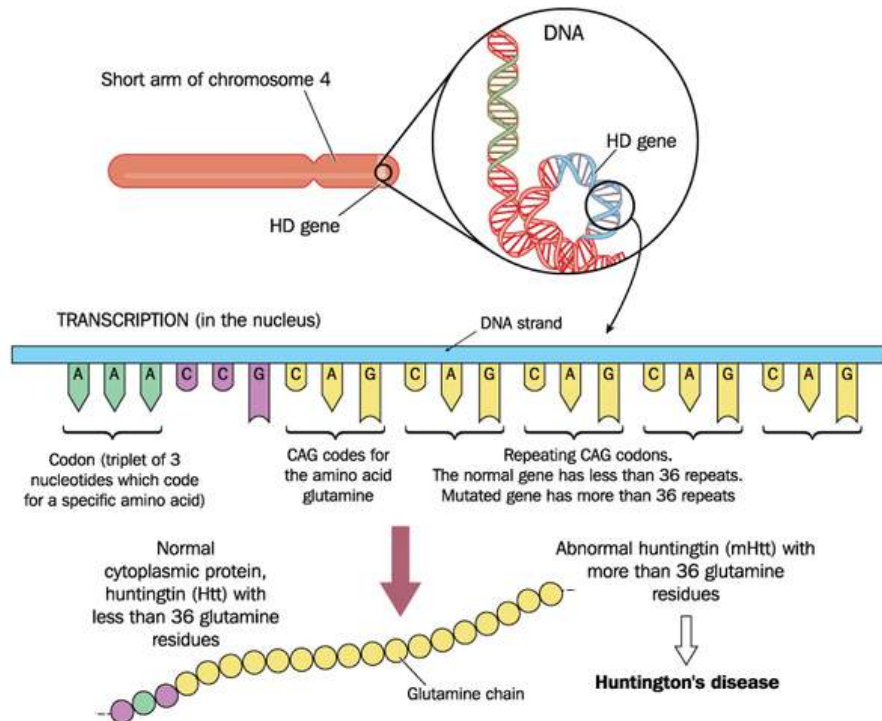


Figure 1 - Chromosome 4 with HD gene and repeating CAG codons causing abnormal huntingtin
Figure from Blamb (2015)⁸

Fragile X Syndrome

Fragile X Syndrome is a genetic disorder known to cause intellectual and developmental disabilities. Because it is an X-linked disease, it is generally more prevalent in males than females. Having this disorder means cognitive developmental problems from a young age and characteristic physical features such as a narrow face, large ears, and flexible fingers.

Fragile X Syndrome is caused by the CGG trinucleotide repeat in the *FMRI* gene on the X chromosome. The *FMRI* gene codes for the Fragile X Mental Retardation Protein (FMRP), and when this protein is absent, the symptoms of Fragile X Syndrome are caused. Unafflicted people generally have 30 CGG repeats, with the normal range extending to 40 repeats. The range for intermediate alleles is considered to be from 41 to 54 repeats. These alleles are longer than the wild type but do not risk symptoms or instability. The trinucleotide repeat sequence only becomes unstable when there are 55 to 200 repeats. It has been discovered that while individuals with this premutation can still produce FMRP and, therefore, do not suffer from Fragile X Syndrome, they do produce lower levels of FMRP. This class of alleles is known as premutations as it is meiotically unstable, often giving rise to the full mutation in females. The permutation cannot expand into a full mutation in males because it cannot be sustained in spermatogenesis. Once the CGG expansion reaches over 200 repeats, the *FMRI* gene is turned off due to the hypermethylation of the expanded track and CpG island, and *FMRI* is epigenetically silenced, resulting in FXS.⁹

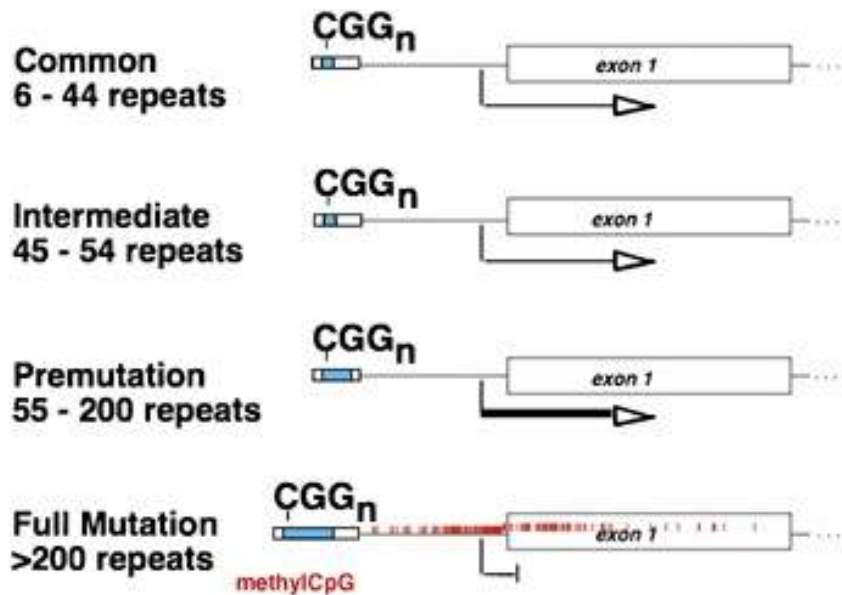


Figure 2 - different magnitudes of CGG repeats causing varying degrees of Fragile X mutation
Figure from Garber et al. (2008)¹⁰

Friedrich Ataxia

Friedrich Ataxia is another degenerative disease that causes damage to the spinal cord, peripheral nerves, and cerebellum. Like Fragile X Syndrome, this disease also develops in children and teens but gets progressively worse, causing symptoms such as loss of senses, loss of reflexes, heart palpitations, and shortness of breath.

Friedrich Ataxia is caused by GAA repeats in intron 1 of the *FXN* gene found on the proximal long arm of chromosome 9. The *FXN* gene codes for frataxin, a protein necessary for mitochondrial function, so when the *FXN* is not fully functional, low frataxin levels cause Friedrich Ataxia symptoms.¹¹ In unaffected individuals, there are normally about 38 or fewer GAA triplets. However, in individuals afflicted with Friedrich Ataxia, the repeat expansion can range anywhere from about 70 to greater than 1000 triplets, typically having 600-900 repeats. Studies have revealed that these GAA repeats cause a triplex structure to form with three strands of DNA. Furthermore, when the GAA expansion is long enough, two DNA segments join at the triplex region, forming the sticky DNA structure. Other studies have also revealed an association between GAA repeats and heterochromatin formation. All of these factors, stemming from the GAA repeat expansion, partially inhibit transcription of the *FXN* gene, blocking most of frataxin production. Because *FXN* is only partially silenced, most individuals afflicted by Friedrich Ataxia are homozygous for the disorder, and generally, those who are heterozygous do not clinically have Ataxia. There is a 2% to 5% minority of heterozygous individuals with Ataxia, but those cases are usually compounded with other mutations. While every cell carries the same genetic code, experiments have revealed that cells of different tissues vary in *FXN* gene expression. In the case of humans, frataxin mRNA is most abundant in the heart and spinal cord, making it one of the most crucial target regions for treatment.¹¹

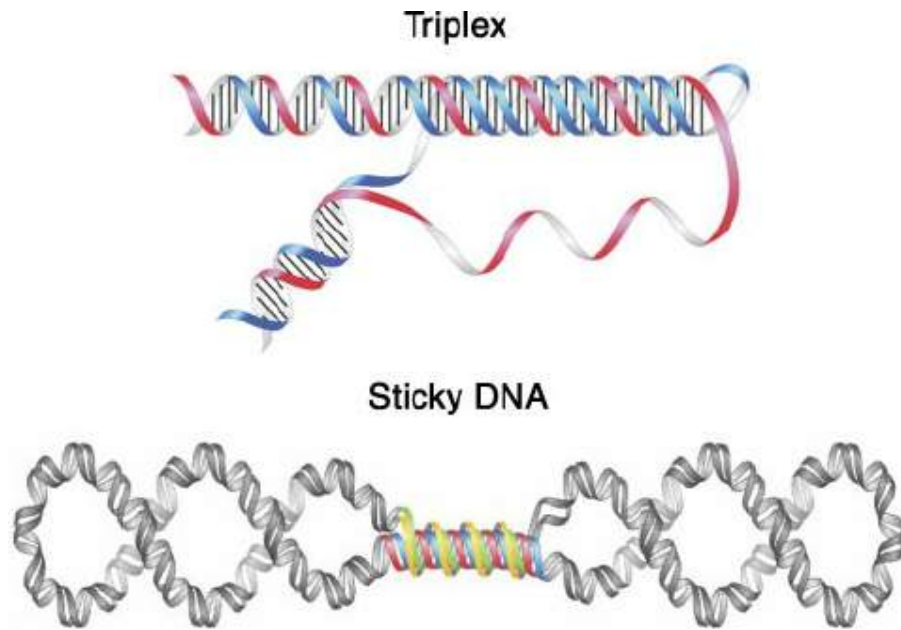


Figure 3 - Triplex and sticky DNA structures that form in DNA as a result of excessive GAA repeats
Figure from Wells (2008)¹²

How CRISPR-Cas9 targets and fixes repeats

CRISPR-Cas9 has taken the gene editing field by storm due to its high efficacy and ease of use. However, one of the greatest strengths of CRISPR-Cas9 is its versatility. Because Cas9 is so flexible in the compounds it can be paired with and targets it can be used for, even when fixing genetic diseases, there are various CRISPR-Cas9-based approaches. In the past years, scientists have conducted clinical trials on genetic diseases using various methods, with a few of the more prevalent being Non-Homologous End Joining, Homology Directed Repair, Exon Skipping, and dCas9.

Non-Homologous End Joining (NHEJ)

NHEJ is one of the primary ways cells repair double-stranded breaks, making it one of the go-to repair methods for pairing with CRISPR-Cas9 when deletions and edits are made in the genome. The system revolves around the Cas9 protein, which can cut DNA. This Cas9 protein is guided by a dual RNA complex that can identify specific sites for the Cas9 protein to cut. Using modified RNA complexes, scientists can accurately control which DNA is cleaved by the Cas9 protein, effectively isolating specific problem DNA segments of their choice to remove. Once the DNA is cut, the cell repairs the break through the NHEJ pathway without needing a homologous template.¹³

Homology Directed Repair (HDR)

HDR is one of the other primary ways cells repair double-stranded breaks in the genome. HDR is used in natural situations far less frequently than NHEJ because it occurs in the S/G2 phase of the cell cycle when there is an undamaged sister chromatid or donor DNA available. In

the first step of this process, a 5' to 3' resection of the DNA is initiated, creating a 3' single-stranded DNA overhang. This overhang allows the damaged DNA to be highly accurately restored using the functional sister chromatid or donor DNA. One downside to using HDR is that longer HDR is known to have lower efficacy, which becomes a problem when targeting extremely long repeat sequences.¹⁴

Exon Skipping

CRISPR-SKIP is a gene editing technique that controls gene expression using exon skipping without inducing any double-stranded breaks in DNA. Unlike the more traditional uses for CRISPR-Cas, CRISPR-SKIP offers the ability to skip exons in long non-coding RNAs selectively. Exon skipping is especially useful in the field of medicine when it comes to genetic diseases because it provides a more permanent fix than the alternatives with a much lower risk of creating chromosomal aberrations or off-target effects.¹⁵

dCas9

Recently, the mutant protein called dead Cas9 (dCas9) has been discovered, opening up a whole new world of possibilities for using CRISPR-Cas9 in gene therapy. Unlike the more standard ways of using CRISPR-Cas9, dCas9 can be designed to modify DNA epigenetics. To affect epigenetics, dCas9 can be engineered to add or remove methyl groups, forcing methylation or demethylation. It can also be used to modify histones, which are essential in the expression of DNA. By changing the chromatin state, the dCas9 does not need to change the DNA but can regulate gene transcription. dCas9 can also recruit transcriptional activators or suppressors to specific genes, offering promising methods for controlling gene expression to treat genetic disorders.¹⁶

Delivery

The CRISPR-Cas system must be able to be delivered for both *in vivo* and *in vitro* processes to be an effective and useable tool in disease treatment. Before the system can be delivered to the cells, it needs to be put into a form that can effectively enter the cells and carry out its function. There are currently three primary forms that the Cas9 protein can take. One method is to deliver the Cas9 protein in the form of mRNA. Due to the properties of mRNA and its ability to be translated directly in the cytoplasm, this method is the fastest and has the slightest chance of inducing other undesired effects. The downside of using mRNA is that because mRNA is rapidly degraded in the cytoplasm, it is unstable, and the Cas9 protein will only have a short duration of gene editing. The second option is using plasmids to deliver pDNA with the Cas9 protein and gRNA encoded in it. As it uses DNA that enters the nucleus, this method is more stable than using mRNA; however, there are many drawbacks to this option.¹⁷ The immediate restraints are that the cell must allow for transfection or viral transduction, and suitable promoters must be chosen for Cas9 and gRNA expression. Furthermore, it is slower and less efficient because the pDNA must enter the nucleus, be transcribed, and then be translated. A

more dangerous downside to using pDNA is that it may integrate into the DNA of the target cell, risking permanent off-target changes in the genome.¹⁸ The last option is ribonucleoprotein (RNP) delivery, in which the Cas9/gRNA RNP complex is generated by incubating Cas9 and gRNA together. This method is highly versatile in delivery, and its greatest strength is the low off-target effect as it is accurate and is degraded over time. On the other hand, this strength can also be a downside as it is unsuitable for long-term use. Additionally, the packaging itself is more complicated than the alternatives.¹⁸

Once the packaging of the Cas9 protein is prepared, scientists must choose one of multiple delivery methods. Currently, the three categorizations of delivery are biological, chemical, and physical. In terms of biological delivery, viral vectors have distinguished themselves as the most efficient delivery method based on the current technologies and information available. Through evolution, viruses have become highly efficient at infecting human cells, effectively protecting contents from degradation and replicating within cells for more permanent effects. Multiple vectors can be used, such as Adenoviral Vector (AdV vector) and Lentiviral vector (LV vector). Still, Adeno-associated Virus (AAV) is one of the most commonly used due to its versatility in terms of compatible target cells. The main weaknesses of using AAV are immunity to the AAV capsid, liver toxicity, and long-term impact due to integration into the genome. Chemical delivery options are also being explored to avoid immunogenicity.¹⁹ One of the most promising non-biological delivery methods revolves around polymer and lipid-based nanoparticles. Scientists have struggled with lipid nanoparticle (LNP) technology because cationic lipids are toxic and can be inefficient in transfection and passing through the cell membrane.¹⁹ Despite these drawbacks, one of the most recent successes in LNP technology can be seen in the COVID-19 mRNA vaccine. The last option for delivery is physical methods, including electroporation, microinjection, and sonoporation. These methods are in the early stages of research but prove promising as electroporation offers a solution for cells that are hard to transfect, and microinjection removes the size restraints on the substance being delivered.¹⁹

Latest treatments and clinical trials applying CRISPR-Cas

Duchenne Muscular Dystrophy

While repeat expansions do not cause Duchenne Muscular Dystrophy, clinical trials regarding its cure are relevant to all other genetic diseases as they spearhead research into gene therapy targeting genetic diseases in general. Duchenne Muscular Dystrophy (DMD) is the most common form of muscular dystrophy, causing muscle degeneration and necrosis. Generally, symptoms become present at 2 to 3 years of age, beginning with difficulties walking and climbing. These symptoms then increase in severity throughout the rest of the afflicted individual's life. Typically, one requires a wheelchair by 10 to 12 years of age, then a ventilator by around 20. The most common outcome is the death of the afflicted patient in their 20s to 40s. Research has revealed that DMD is caused by mutations in the gene encoding for dystrophin

production found on the X chromosome. DMD is specifically caused by either a frameshift mutation, which is the product of deletions or duplications, or a nonsense mutation.²⁰ Due to the wide range of mutations that can cause DMD, a variety of clinical trials have experimented with different approaches to effectively fix the mutation. Many of the most recent and successful clinical approaches use some form of the CRISPR-Cas system due to its versatility and efficiency.

Table 1

Compound	Target	Cells/Tissue	Delivery	Outcome	Source
NHEJ CRISPR-Cas9	Correct DMD by skipping mutant dystrophin exons	Postnatal muscle tissue in mdx mouse (in vivo)	Adeno-associated virus-9 (AAV9)	Restored expression of Dystrophin protein in cardiac and skeletal muscle. Varying levels of expression increased 3-12 weeks after injection. Enhanced skeletal muscle function	²¹
NHEJ CRISPR-Cas9	Restore reading frame by excision mutated DMD exon 23	Myofibers, cardiomyocytes, and muscle stem cells in mdx mouse (in vivo)	Adeno-associated virus (AAV)	Successful excision restored reading frame and resulted in truncated but functional protein. Partially restored muscle function and created endogenously corrected myogenic precursors	²²
NHEJ and HDR CRISPR-Cas9	Restore NHEJ reading frame by HDR exon 23 correction	Skeletal muscles and heart in mdx mouse	1-cell embryo injection	Restoration of dystrophin and 47-60% of heart and skeletal muscle fibers found dystrophin positive.	²³
NHEJ CRISPR-Cas9	Restore NHEJ reading frame by deletion in exons 45-55 and introns 44, 50, 51, 55	Immortalized patient DMD muscle cells, immunodeficient NSG mice	Electroporation	Successful deletion of large exon 44-45 region, mice with treated fibroblasts had dystrophin-positive fibers	²³
dCas9	Synthetic transcriptional activation of utrophin by targeting promoter	Immortalized patient DMD muscle cells	Electroporation	Achieved 1.7–6.9-fold upregulation of utrophin and restored β -dystroglycan expression	²³

Among most CRISPR-Cas-related trials targeting DMD, one notable commonality is using NHEJ over other methods. Due to the straightforward nature of using CRISPR-Cas to delete parts of the mutant genome and the natural process of NHEJ, which can also be directed, it makes sense that these types of trials are the most common. Furthermore, the trials show some levels of success as dystrophin was restored, leading to enhanced muscle function. As a majority of the NHEJ trials are performed in vivo and there have been multiple ways of successfully delivering the mechanism, this method of gene therapy shows promise for curing DMD and possibly other genetic diseases.

One notable case is the 2016 Wojtal et al. trial in which dCas9 is used to activate utrophin by targeting the promoter. Unlike the other trials, there is no attempt to restore the reading frame or target mutations to restore utrophin production. Instead, the dead Cas9 complex is used to directly restore the production of utrophin by transcriptionally activating the promoter that was inactivated by the mutant gene. This method is another possible solution to DMD because it does not risk any off-target, lasting effects on the genome since it cannot induce any genetic changes.

Huntington's Disease

Table 2

Compound	Target	Cells/Tissue	Delivery	Outcome	Source
RNAi	HTT lowering	Mouse neurons	intrastriatal injection	Mutant HTT silenced, abnormal behavioral phenotype delayed	²⁴
ZFP	HTT lowering	Mouse neurons	intrastriatal injection, AAV delivery	Reduced chromosomal expression of mutant HTT at protein and mRNA levels, low off-target effect	²⁴
CRISPR-Cas9 and gRNA	HTT inactivation by targeting mutant PAM sequences	In vitro untransformed primary fibroblast cells from humans	Vector	All large deletions only occurred in mutant alleles as intended and had successful knockout effects. CAG repeats removed from DNA and RNA, resulting in reduced mutant HTT.	²⁵

As Huntington's Disease is not as well known and experimented on as DMD, there have been fewer CRISPR-Cas9-related trials attempting to cure its symptoms; however, there have been promising trials for the future. In the 2016 clinical trial performed by Shin et al., a CRISPR-Cas9 and guide RNA system was able to successfully and accurately target mutant PAM sequences. This process successfully removed CAG repeats, and the reduced mutant HTT

signifies an effective method for gene editing with future promise. Even though it was so successful, gene editing to fix Huntington's does not come without risk. Because HTT is a vital protein to the body, targeting the HTT gene risks harming the wild type and further impairing function, increasing the need for safety precautions, even if it slows progression.

Other than using CRISPR-Cas9, clinical trials showed successful results using compounds such as RNAi and ZFP to lower HTT. While CRISPR-Cas9 appears to be the best all-round method of repairing these genetic disorders, the trials carried out in the table above prove that in specific cases, other methods of gene editing can also be effective solutions. Knowing this, scientists should not abandon any methods other than CRISPR-Cas9 in the future in hopes of discovering an even more effective method to apply specifically to one case or just a better method in general.

Fragile X Syndrome

Table 3

Compound	Target	Cells/Tissue	Delivery	Outcome	Source
SpCas9	Deletion of 47 bp upstream of CGG repeat	iPSC	Electroporation	CGG repeats deleted in 2-3% of expanded cells, sustained reactivation of FMR1	²⁶
SpCas9	Deletion of 40 bp upstream and 35 bp downstream of CGG repeat	iPSC	Nucleofection	CGG repeats deleted in 20% of cells, <i>FMR1</i> reactivation persisting 50 days in culture	²⁷
Cas9 and Cfp1	Target Grm5 to reduce activation	Striatum in mouse	Gold nanoparticles	Reduced Grm5 expression and return of phenotypic behaviors	²⁸
dCas9-VP192	Selective reactivation of silenced FMR1 locus targeting promoter and CGG repeats	Human ESC and neuronal progenitor cells	Lipid mediated Transfection	<i>FMR1</i> reactivation but low FMRP expression from CGG targeting and lower expression from promoter targeting	²⁹
dCas9-Tet1	Demethylation of CGG repeats	iPSC	Lenti virus	<i>FMR1</i> reactivated and CGG repeat locus demethylated. Low reactivation rates in neurons. Half of edited iPSC expressed FMRP.	³⁰

The various clinical trials conducted on Fragile X Syndrome using CRISPR-Cas9, show two primary methods for curing this genetic disease. The first method is seen in the first three trials of this table and uses CRISPR-Cas9 for the knockout of genes, showing varying degrees of success. The first study in the table used CRISPR to target 47 base pairs above trinucleotide repeat and complete deleted CGG repeats in induced pluripotent stem cells. Xie et al. used a similar strategy, but not all clones were reactivated with full efficiency. Alternatively, the last two trials on the table used dCas9 in an attempt to reactivate FMR1. In Liu et al. 2018, CRISPR was used epigenetically with dCas9 fused to Tet1 to incur epigenetic changes at the FMR1 locus. Another dCas9 trial by Haenfler et al. used dCas9-VP192 and guide RNAs to target the FMR1 promoter region without altering genes. While neither of these trials could use dCas9 to restore FMRP expression fully, they show that it is possible to see some progress using those methods.

Friedrich Ataxia

Table 4

Compound	Target	Cells/Tissue	Delivery	Outcome	Source
CRISPR-Cas9	Deletion of GAA repeats in intron 1 of FXN gene	Tibialis anterior muscle of the YG8R mice	DNA electroporation, (AAV-DJ vector in testing)	Successful deletion of GAA expansion, up to twice as much frataxin mRNA and protein levels.	³¹
CRISPR-Cas9 and NHEJ	Removal of GAA repeats in the first intron of FXN gene	YG8R mouse lymphoblasts (in vivo & in vitro)	Lipofectamine CRISPRMAX Cas9 transfection reagent kit	No difference in FXN expression in healthy lymphoblasts. Increase in FXN expression in treated cells, comparable to asymptomatic cells.	³²

Of the four genetic diseases discussed, Friedrich's Ataxia has proven the most difficult to fix for scientists. One of the biggest obstacles scientists face when approaching Friedrich's Ataxia is ensuring regular frataxin expression is maintained after gene editing since they are deleting genetic material. In many clinical trials, scientists could remove repeat expansions but actually reduced frataxin levels due to one whole copy being deleted in addition to repeats, leaving only one functional FXN copy. In some cases, a decrease in cell viability was observed after GAA removal.

Despite these challenges, the table above shows some cases where researchers have circumnavigated these obstacles, using CRISPR knockout to remove GAA repeats. The trial

carried out by Rocca et al. reveals that CRISPR-Cas9 gene editing was optimized using 4RNP specifically, offering future direction for developing more effective methods.

Bioethics

The recent developments in gene editing hold great transformative potential in the field of genetic engineering. The precision and control scientists are gaining over genetic information can allow for great advancements, such as the removal of genetic diseases, as discussed in this paper. However, it also brings up moral and legal conflicts. One of the biggest questions now being asked is if genetic engineering threatens the sanctity and integrity of human life. One example of this is the concept of “designer babies.” By altering natural human variability and possibly widening social inequalities, moral debates are being brought up by opponents of genetic engineering. There is also the conflict of consent. Because changes are being made to the genome, these changes will affect all individuals born from the individual with the edited genes. This is problematic because changes will be made that affect the unborn children who are unable to consent. Going beyond the individual cases, possibly one of the most critical issues is the threat to broader societal equality. Due to the nature of gene editing, it will most likely only be available to the wealthy. With disease prevention and genetic enhancements mirroring one’s socioeconomic status, class divisions will only widen. In addition to these moral and social issues, a slew of legal conflicts are brought up by new developments in gene editing. The problem arises because many legal codes were set up before the recent developments in genetic engineering, leaving this field lacking the proper legal guidelines. To prevent a chaotic environment where risky or unethical experiments take place, new regulatory procedures will need to be set up, along with a set of consistent international policies to prevent reproductive tourism.

The Future of the Field

I see a highly promising future for gene editing if proper precautions are taken. As seen in this paper, tools such as CRISPR-Cas9 have proven effective in fixing genetic diseases, one of the most difficult medical issues to fix. I believe that at the rate the field of genetic engineering is currently developing, we will soon see a world where genetic diseases will begin disappearing from the gene pool. I also foresee that gene editing will be able to fix more than just genetic diseases, possibly even being applied to problems not directly caused by incorrect genetic code. While the possible benefits are too great to refuse, my reservation comes from the gravity of the risks that come with gene editing. Because gene editing often induces permanent changes to the genome, any intentional changes, or worse, mistakes made during editing, will affect the whole gene line. That said, I do not believe this will be a problem, considering the safe ways gene editing is being tested, using both in vitro tests and in vivo trials in mice. From my perspective, scientists have done well to evaluate various tools and methods for gene editing. However, with CRISPR-Cas9 coming out as one of the best options, scientists should focus on broadening its application to different issues.

Works Cited

1. emhj. Genetic disorders and congenital abnormalities: strategies for reducing the burden in the Region. *World Health Organization - Regional Office for the Eastern Mediterranean*
<http://www.emro.who.int/emhj-volume-3-1997/volume-3-issue-1/article18.html>.
2. Kim, J. C. & Mirkin, S. M. The balancing act of DNA repeat expansions. *Curr. Opin. Genet. Dev.* **23**, 280–288 (2013).
3. Paulson, H. Chapter 9 - Repeat expansion diseases. in *Handbook of Clinical Neurology* (eds. Geschwind, D. H., Paulson, H. L. & Klein, C.) vol. 147 105–123 (Elsevier, 2018).
4. Gostimskaya, I. CRISPR–Cas9: A History of Its Discovery and Ethical Considerations of Its Use in Genome Editing. *Biochem. Biokhimiia* **87**, 777–788 (2022).
5. Armstrong, B. Repeat after me: what are repeat expansion disorders? *Genomics Education Programme*
<https://www.genomicseducation.hee.nhs.uk/blog/repeat-after-me-what-are-repeat-expansion-disorders/> (2022).
6. The Huntington Gene - UC Davis Huntington’s Disease Center of Excellence.
<https://health.ucdavis.edu/huntingtons/genetic-change.html>.
7. Roos, R. A. Huntington’s disease: a clinical review. *Orphanet J. Rare Dis.* **5**, 40 (2010).
8. Genetics Huntingtons Disease Showing Chromosome 4 Stock Vector (Royalty Free) 285928157. *Shutterstock*
<https://www.shutterstock.com/image-vector/genetics-huntingtons-disease-showing-chromosome-4-285928157>.
9. Fragile X syndrome | European Journal of Human Genetics.
<https://www.nature.com/articles/ejhg200861>.
10. Garber, K. B., Visootsak, J. & Warren, S. T. Fragile X syndrome. *Eur. J. Hum. Genet.* **16**, 666–672 (2008).
11. Friedreich Ataxia | Genetics and Genomics | JAMA Neurology | JAMA Network.
<https://jamanetwork.com/journals/jamaneurology/fullarticle/796230>.
12. Wells, R. DNA triplexes and Friedreich ataxia. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **22**, 1625–34 (2008).
13. CRISPR–Cas9 Structures and Mechanisms | Annual Reviews.
<https://www.annualreviews.org/content/journals/10.1146/annurev-biophys-062215-010822>.
14. Yang, H. *et al.* Methods Favoring Homology-Directed Repair Choice in Response to CRISPR/Cas9 Induced-Double Strand Breaks. *Int. J. Mol. Sci.* **21**, 6461 (2020).
15. CRISPR-SKIP: programmable gene splicing with single base editors | Genome Biology.
<https://link.springer.com/article/10.1186/s13059-018-1482-5#Sec11>.
16. IJMS | Free Full-Text | Dead Cas Systems: Types, Principles, and Applications.
<https://www.mdpi.com/1422-0067/20/23/6041>.
17. Huang, J., Zhou, Y., Li, J., Lu, A. & Liang, C. CRISPR/Cas systems: Delivery and application in gene therapy. *Front. Bioeng. Biotechnol.* **10**, (2022).
18. Hempstead, A. CRISPR 101: Ribonucleoprotein (RNP) Delivery.
<https://blog.addgene.org/crispr-101-ribonucleoprotein-rnp-delivery>.

19. Delivery of CRISPR-Cas tools for in vivo genome editing therapy: Trends and challenges - ScienceDirect.
<https://www.sciencedirect.com/science/article/pii/S016836592200027X#s0085>.
20. Duan, D., Goemans, N., Takeda, S., Mercuri, E. & Aartsma-Rus, A. Duchenne muscular dystrophy. *Nat. Rev. Dis. Primer* **7**, 1–19 (2021).
21. Long, C. *et al.* Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science* **351**, 400–403 (2016).
22. Tabebordbar, M. *et al.* In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science* **351**, 407–411 (2016).
23. Lim, K. R. Q., Yoon, C. & Yokota, T. Applications of CRISPR/Cas9 for the Treatment of Duchenne Muscular Dystrophy. *J. Pers. Med.* **8**, 38 (2018).
24. Wild, E. J. & Tabrizi, S. J. Targets for future clinical trials in Huntington’s disease: What’s in the pipeline? *Mov. Disord.* **29**, 1434–1445 (2014).
25. Shin, J. W. *et al.* Permanent inactivation of Huntington’s disease mutation by personalized allele-specific CRISPR/Cas9. *Hum. Mol. Genet.* **25**, 4566–4576 (2016).
26. Park, C.-Y. *et al.* Reversion of FMR1 Methylation and Silencing by Editing the Triplet Repeats in Fragile X iPSC-Derived Neurons. *Cell Rep.* **13**, 234–241 (2015).
27. Reactivation of FMR1 by CRISPR/Cas9-Mediated Deletion of the Expanded CGG-Repeat of the Fragile X Chromosome | PLOS ONE.
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0165499>.
28. Lee, B. *et al.* Nanoparticle delivery of CRISPR into the brain rescues a mouse model of fragile X syndrome from exaggerated repetitive behaviours. *Nat. Biomed. Eng.* **2**, 497–507 (2018).
29. Haenfler, J. M. *et al.* Targeted Reactivation of FMR1 Transcription in Fragile X Syndrome Embryonic Stem Cells. *Front. Mol. Neurosci.* **11**, (2018).
30. Liu, X. S. *et al.* Rescue of Fragile X Syndrome Neurons by DNA Methylation Editing of the FMR1 Gene. *Cell* **172**, 979-992.e6 (2018).
31. Ouellet, D. L. *et al.* 321. Deletion of Mutated GAA Repeats from the Intron 1 of the Frataxin Gene Using the CRISPR System Restores the Protein Expression in a Friedreich Ataxia Model. *Mol. Ther.* **24**, S129 (2016).
32. CRISPR-Cas9 Gene Editing of Hematopoietic Stem Cells from Patients with Friedreich’s Ataxia: Molecular Therapy Methods & Clinical Development.
[https://www.cell.com/molecular-therapy-family/methods/fulltext/S2329-0501\(20\)30078-4](https://www.cell.com/molecular-therapy-family/methods/fulltext/S2329-0501(20)30078-4).

Antigone and Thecla: Females In Martyrdom Literature By Jian Wang

Introduction

Women had always been a significant part of the martyrdom narration. They were at the forefront of the clashes. They were arguably the most controversial of all the martyrdom stories for their very nature of being more mystical, vague, and, to some extent, more malleable than male martyrs. Especially in Thecla's story, such narrations made it especially hard to categorize her actions, for she was a rebel against the social expectations of getting married. Still, at the same time, she was attracted by Paul and was following him as, to some extent, a zealous pursuer. Her actions were, at the same time, rebellious natures but also submissive traits. This interesting paradoxical discourse was rarely seen on male martyrs, who, in most martyrdom accounts, were perceived as more straightforward and were of a simpler motivation structure. Meanwhile, for female martyrs, their actions seem more obscure, and, under the expectations of women in the context of the patriarchic narration of the stories, they were harder to attribute to a unified and singular cause. Therefore, in this review paper, I will focus on female martyrs to decipher their seemingly more esoteric and complex motivations.

This paper will feature two of the primary sources that have been provided that describe female martyrs. The first one would be Antigone, and the second one would be Acts of Paul and Thecla. The reason why these two martyr stories were chosen is because they hold a common sense of defeminization of the two female martyrs, and therefore might provide novel insights into the matter of women martyrs.

By studying these two extremely representative female martyrs, it is possible to obtain a more thorough and complete view of the role and treatment that the authors of these martyrdom works of literature appended to them.

Review of Literature

The story of Antigone begins with tragedy: the two brothers Ismene and Antigone are killed. Moreover, they died in a fratricide. However, their fate after their death was different: Eteocles, the one who stood with the ruler of Thebes, Creon, was honored and buried in a tomb, whereas Polynices, who stood against the ruler, was left unattended in the fields. (Sophocles 159) Antigone, who was devastated by the news of losing two brothers in one day, decided to claim the body of his beloved brother Polynices and asked Ismene, her sister, to join. However, it was not that easy a promise to keep. Creon sent out a prohibition, sentencing anyone who dared to bury the body of Polynices to death. Here, Sophocles proposed a contrast between the two sisters, who reacted completely differently. Ismene, knowing that it would be fatal for Antigone to bury the body, refused the suicidal asked and pleaded with Antigone not to seek a determined death. As on page 162, Ismene proclaimed: "I shall do no dishonor. But to act against citizens. I cannot." She presented another sense of determination that she determined not to break the decree of the ruler and, more importantly, not to lavish her precious life on a determined death. But still, Antigone went. From this first scene, we can see two types of women being described.

Ismene and Antigone clearly showed opposite reactions towards the death of their brother and the prohibition of Creon.

The news of the burial of Polynices soon went into the ears of Creon. His decree was broken by not a man, but a woman! It was Antigone, a woman, who dared to violate his interdiction! In his conversation with his guard, Creon could not believe such a thing could happen, especially since a woman shook his authority. He kept on regarding the person who had done the deeds as a male figure and was utterly shocked when he knew that Antigone was, in fact, a woman. (Sophocles 171) Later, when Antigone was brought up against Creon, he said: "No woman rules me while I live!" (Sophocles 177) However, Antigone did, implying her masculinization. Creon's authority was shaken not only because his order was not obeyed but also because the act was done by a woman whom he thought was somewhat inferior to him. This clearly matches up with the hierarchy of a patriarchic order: through the defiance of the probability of the existence of women on par with Creon in terms of power, Antigone is depicted as a man. For announcing that no woman rules over Creon, it was implied that Antigone didn't fall within the category of a woman. Therefore, through Sophocles' narration, the power of Antigone was not constructed upon her identity as a woman but through defeminizing her. This can be reflected in the contrast between her and her sister. Even though Ismene ends up turning herself in after Antigone is caught, her figure is fundamentally different from Antigone's. Antigone's sense of resistance and Ismene's submission seem like a masculinization of Antigone. Since Antigone is the only family member of Ismene who's still alive, Ismene's without-condition following Antigone became a hierarchy of power in which Antigone is the more masculine and daring figure, and Ismene is a timid adherent—she is dependent on Antigone. Sophocles, in some sense, presented that the action of martyrdom is purely masculine, and an obedient figure, like Ismene, should not be the person to conduct such actions. This sense of defeminization in terms of empowerment was even more hyperbolized in the story of Thecla.

Thecla, in the story, was a virgin who was obsessed with Paul and became his disciple; she was already engaged with another person, Thamyris, by that time. Due to her love for Paul, she broke her vow to Thamyris, and therefore, she was sentenced to death. Later, when Paul was sent to jail by the local governor, Thecla bribed the guards and went in to meet Paul. Together, they were sent to the seat of judgment. The governor asked Thecla why she did not answer the question. She just stared at Paul. Her mother, Theocleia, then cried: "Burn the lawless one! Burn her that is no bride in the midst of the theatre, that all the women who have been taught by this man may be afraid!" (Schneemelcher 242) This represented a social norm that an engaged couple should eventually marry, and such a rule is constraining to the extent that the punishment for disobeying it is the sentence of death. The authority of the rule is so undeniable that even the family affection between mother and child cannot prevail. However, Thecla did; not only did she turn down the engagement, but she was in a relationship that was described as somehow erotic with Paul. She would even listen to Paul's sermons for three days without eating or drinking.

After she magically survived the punishments of the governor, Thecla's decisions were independent. Even though Paul thought otherwise, Thecla still proceeded to cut short her hair,

baptize herself, make herself clothes of men, and go back to Iconium. The revolting figure of Thecla surely did not correspond to the traditional submissive women's figure at her time, and she surely was completely different from Ismene. Her figure, in that fashion, was more masculine than the other women's figures depicted, such as her mother, Theocleia.

Constantinou, however, discovered a discrepancy within the narration of the work. In Thecla's story, empowerment seemed to intertwine with the erotic male gaze and control over Thecla, as she mentioned in her essay, "Instead of offering role models for actual Christian women, female martyr legends generate and perpetuate discourses and social structures that sanction men's control over women's bodies and sexualities." (Constantinou 75) However, the empowerment of Thecla to become a teacher after her erotic tortures were sheerly masculine, transforming Thecla from a woman into a man. In the narration of the author of Thecla's story, the female body of Thecla was not actually the martyr; it was inferior to hold God's power. The body of Thecla that was able to hold the might of God and the honor of becoming a martyr was the male body of Thecla, the body that had been masculinized. Only the body of a man was considered an adequate body of a martyr, and Thecla, therefore, became a man. Even during her torturing, the reveal of her feminine body was considered inappropriate by God. Therefore, God summoned a storm to protect her.

Both Antigone and Thecla's actions were apparently not on par with what was commonly expected of women from a patriarchic viewpoint. In a greater sense, through the exaggeration of the masculinization of Antigone and Thecla, women were deprived of the rebellious context of martyrdom.

Another essential trait of the female martyr narration was the dehumanization of the martyrs. For instance, in Antigone's story, through Teiresias's signs, the fate of Creon is revealed: if death comes to Antigone, the same thing happens to Creon's son and wife. (Sophocles 193-196) Later, when Creon tries to save Antigone from the fate of death that he himself assigned to her, a messenger comes with the news of his suicide of Antigone, alongside the news of the death of his son and wife.

In this part of the story, Antigone not only loses her femininity but also her human character. She was no longer a person but a container, a container of wrath, hatred, and vengeance. Her life no longer matters as the life of a living person but serves as a tool of punishment. At this point, Antigone's dehumanization came to completion. A similar description in Thecla's story would be her submission to Paul. His power over her was so great that she would give up her life to be with him. In this case, Thecla could only become the submissive side of the power structure between her and Paul. She was a possession of Paul. Her autonomy, on the other hand, seemed to be controversial to her submission to Paul. She disobeyed Paul's orders and baptized herself, and she went back to Iconium when Paul asked her to go and teach the world of God. However, her power to reject Paul was not provided by her identity as a woman but by her identity as a man. By cutting short her hair, dressing up like a man, and becoming a teacher of God's words after her departure to Seleucia, Thecla became an imitator of Paul; she followed Paul to the extent that she became "Paul," and it was her becoming of Paul that made her independent.

Conclusion

In conclusion, the narration of the martyrdom stories of Antigone and Thecla hold a common characteristic of defeminization of the female martyrs. From the stand-off between Antigone and Creon to Thecla's becoming an independent disciple, such empowerment nevertheless came from the transformation of femineity to masculinity. In other words, within the narration of these stories, the virtues of a martyr were bound to the body of a man, so for women to become martyrs, they must become men in advance. Another characteristic was their sense of dependence on Paul, which showed her inferiority and submission compared to Paul. From that, she became too powerful for a woman because of her martyrdom, yet too feminine to really become a man. Therefore, it can promptly be concluded that in the sense of female martyr narrations, the defeminization and dehumanization of the martyr were key in creating and shaping a triumphant female martyr. In order to become a martyr, one must become a man.

Works Cited

Sophocles, and Elizabeth Wyckoff. "Antigone."

Schneemelcher, Wilhelm. "Acts of Paul and Thecla." *New Testament Apocrypha*, Westminster John Knox Press, pp. 239–246.

Constantinou, Stavroula. "Thekla the Virgin: Women's Sacrifice and the Generic Martyr." *The "Other" Martyrs: Women and the Poetics of Sexuality, Sacrifice, and Death in World Literatures*, Harrassowitz Verlag, pp. 73–86.

Autonomous Sensory Meridian Response (ASMR)'s Effectiveness on Lowering Stress and Tension By Kexin Lyu

ABSTRACT

Noted for its tingling sensation that is triggered by specific auditory or visual stimuli, Autonomous Sensory Meridian Response (ASMR) has been a popular research target in areas such as inducing relaxation and promoting mental well-being. The present article outlines the studies on the effectiveness of ASMR videos and apps in reducing stress and tension, aiding sleep, and promoting overall well-being in young people. The following content delves deeply into the research on the vulnerability to stress, sensory processing sensitivity, and the ASMR phenomenon. Studies included here show unique neural activation associated with ASMR, suggesting potential benefits in treating mental disorders. The combination of ASMR with Virtual Reality (VR) technology also showed promising potential in enhancing relaxation. Moreover, emerging research on using ASMR in treating Attention Deficit Hyperactivity Disorder (ADHD) presents ASMR as a promising alternative treatment option for individuals with the disorder. Future research should put more efforts on deciphering the full potential of ASMR in mental health treatment and the corresponding mechanisms.

Keywords Behavioral and Social Sciences, Neuroscience, Stress Reduction, ASMR, Sleep Aid, ADHD

1. INTRODUCTION

Among the younger population (*e.g.* children and adolescents) in many countries, there has been an increasing concern on the issues of mental health problems. Statistical analysis showed that the recent rates of anxiety were found to be 24.9%, the rates of depression was found to be 19.7%, and the recent rates of stress was found to be 15.2% (Tang et al., 2021). According to a meta-analysis research done by Racine et al., about 20% of the young were identified with anxiety symptoms. (Racine et al., 2021). Another review done by Sharma et al. (2021) reported that among children and adolescents, 54% have sleep disorders, 27% experience worsening sleep, and 16% experience irregular sleep duration. These phenomena lead to the search of self-help tools by many individuals to support their well-being. Apps have been developed to help these individuals. Autonomous Sensory Meridian Response (ASMR) videos is one such tool that has gained popularity.

ASMR is a special media or video tool that, through the tingling sensation, calms down the individuals and triggers pleasurable feelings (Taylor, 2014). ASMR videos and apps may utilize various forms such as whispering, tapping, and role-playing, and are specifically designed to carry out the relaxing and therapeutic functionalities. As evidenced by many studies, ASMR videos and apps have shown to be helpful in reducing stress, improving sleep quality, and reaching for better health outcomes.

Therefore, the aim of this research essay is to explore the effectiveness of using ASMR videos and apps to reduce stress and tension, aid sleep, and promote overall well-being.

2. DISCUSSION

It was found that some people are more sensitive to sensations than others, indicating that those people feel the physical effects of stress and tension more frequently and intensely. It is said that 15%-20% of the population experience Sensory Processing Sensitivity (SPS).

Individuals with SPS process sensory information more deeply, increasing emotional reactions and empathy. While SPS is not a disorder, it can impact well-being, especially when managing sensitivity to negative stimuli, potentially resulting in stress, tension, and sleeping difficulty.

Research done by Belsky and Pluess (2009) showed that some individuals are more susceptible to the effects of their environment. These people tend to seek relief from negative physical and emotional reactions to their environment to reduce their stress, tension, anxiety, and insomnia.

People more prone to experiencing negative sensations can also experience positive sensations from the surrounding environment. ASMR videos combine auditory triggers (e.g., whispering and tapping) and nature sounds with visual stimuli such as eye contact and hand movements. As a result, they induce a pleasant tingling sensation and bring feelings of “comfort, bliss, and euphoria” (AdenClements, 2012; Tufnell, 2012). Therefore, ASMR media may be utilized as a potential strategy to reduce insomnia.

Research done by Lochte et al. (2018) further showed the sensation perception of ASMR media. They measured the activated human brain regions while listening to ASMR video and revealed that whispering and personal attention induced the strongest tingling feelings. These findings suggest that the ASMR experience can create a unique neural activation that generates relaxing sensations similar to social bonding and musical frisson. Another study done by Inagaki and Ohta (2022) recorded the electroencephalograms (EEGs) of their studied participants. The participants in the survey were listening to different ASMR sounds and performing maze tasks while their EEGs were being measured. Inagaki and Ohta (2022) found that ASMR sounds helped the participants to show recovered EEG activation patterns during maze tasks. This indicates that ASMR could reduce stress caused by mental workload. Another study done by Wood and Cobbs (2023) found that participants with ASMR sessions decreased heart and respiration rates. These are the typical physiological changes that occur when people experience relaxation and euphoria. These studies provided a biological explanation that ASMR media produces positive sensations in the body that reduce stress, tension, anxiety, and negative emotions. It is a self-help method with positive results for reducing stress and tension, aiding sleep, and promoting overall well-being.

In addition, the combination of ASMR and virtual reality (VR) technology has shown promising potential for bringing relaxation and promoting mental well-being. More and more research is also emerging on ASMR in treating Attention Deficit Hyperactivity Disorder (ADHD) presents ASMR as a promising alternative treatment option for individuals with the disorder. Because few scientific studies have been conducted on ASMR media use for mental

health purposes, more studies are needed to explore the long-term health benefits, as current studies show short-term benefits. More research is needed for understanding the underlying mechanisms of ASMR in relation with mental health improvement, and discovering the full potential of ASMR.

2.1 Vulnerability to Stress

Sensory Processing Sensitivity (SPS) describes those who are extremely sensitive to the senses in the environment, such as scents, sounds and tastes. People with SPS have also been shown to feel greater emotions and empathy, and be easily overstimulated (Aron et al., 2012). Personality traits that are related to SPS are "introversion (or low extroversion), neuroticism (or irritability/ negative emotionality), and openness to experience" (Asscher et al., 2016). SPS is a temperament trait, and not a disorder. It can, however, have a negative effect on well-being, especially when people try to manage their sensitivity to negative or unpleasant environmental stimuli. Therefore, young people who have SPS might experience stress, tension, inability to sleep, and even ADHD more intensely than others (Greven et al, 2019).

Belsky and Pluess (2009) conducted a study exploring individual differences in environmental impacts. They found that some people are more sensitive to adversity, while others are more susceptible to supportive experiences. This can be attributed to temperamental characteristics, phenotypes, and specific genes often explained by the “diathesis-stress model” that focuses on vulnerability in understanding people’s response to stress. However, Belsky and Pluess (2009) suggest that sometimes, these genetic factors function less like “vulnerability factors” and more like “plasticity factors.” They can be malleable or changed as a result of the environment. Some children are empirically observed to be more sensitive to experience due to their nervous system, which registers experiences more strongly.

There is evidence supporting the “differential-susceptibility hypothesis” in understanding that some individuals are more affected by environmental experiences than others. It appears that individuals vary in their plasticity, with some being more susceptible to adverse effects of unsupportive environments and beneficial effects of supportive ones. This includes differential-susceptibility effects of parenting, child-care quality, life events, rural versus urban residence, and birth season on various outcomes such as attachment, externalizing problems, antisocial behavior, depression, respiratory health, and endophenotypes. These findings perhaps explain why young people often experience Sensory Processing Sensitivity (SPS), a result of not only genes but also the environment (Aron, 1997). Therefore, it might be assumed that stressful environmental factors have increased worldwide affecting more and more young people, regardless of their genetics.

2.2 Autonomous Sensory Meridian Response (ASMR)

The ASMR community on YouTube creates videos that aim to produce a relaxing and shivering sensation for the viewer. The phenomenon known as “the shivers” exemplifies how emotions and physical sensations are interconnected. Without the emotional story to accompany

it, a touch or pat from a friend would not have the same effect. Similarly, the story would not create the same response without the physical sensation. There is a mutual relationship between touch and emotions (Andersen, 2015).

The sensations that these videos seek to recreate are described by users as “head tingles” (Cheadle, 2012) or “brain orgasms” (AdenClements, 2012). It was named the Autonomous Sensory Meridian Response (ASMR) by the founders of the website (<http://www.asmr-research.org>) in 2009 (AdenClements, 2012; Cheadle, 2012). Users claim that the shivers produce “intense feelings of relaxation” (AdenClements, 2012) that begin at the head and then extend to the limbs. ASMR is a sensory experience that involves emotions and associations of intimacy, care, and attention. Though ASMR videos draw on real-world paradigms of care, the attachments that ASMRers have to them beyond their sensory power help us envision a future where these experiences of computer-enabled intimacy are as emotionally compelling as those of maternal care, personal touch, or sex. ASMR videos become popular on platforms like YouTube and TikTok, attracting millions of viewers and subscribers. These videos, particularly roleplay videos, often feature ASMR content scenarios like makeup artists brushing viewers' faces, hairdressers giving haircut or massage, or doctors performing examinations. These videos may also include whispered that give positive words, such as "good job" , "perfect" and "great". And various triggers to induce relaxation and a sense of intimacy.

2.3 New Area of Research

Although much of the research up-to-date has focused on sleep aid and stress reduction, some of the research also sheds light on alleviating the signs and symptoms of ADHD as well as the autism spectrum disorder (ASD). Developed typically during childhood, ADHD is a neural disorder that is associated with forgetfulness, restlessness, fidgeting, insomnia, excessive talking, and difficulty in sustaining attention and mental effort. It is also associated with a higher incidence of other psychiatric problems. If left untreated, ADHD can develop a range of negative problems that interfere with one’s daily life, such as delinquency, motor vehicle safety issues, and difficulties with personal relationships (Wilens and Spencer, 2013). It is worth noting that many of ADHD associated issues - anxiety, depression, insomnia, and restlessness - may be alleviated by using ASMR media.

Abnormal sensory sensitivity is a common feature of both ASD and ADHD, which often presents comorbidly. A study with 116 adults with ADHD compared their sensory sensitivities and autism-spectrum symptoms to norm values, indicating that adults with ADHD have more hyper- and hyposensitivity to external stimuli compared to norm groups, or individuals without this sensitivity, particularly in activity level and auditory sensory modalities (Biljenga et al., 2017).

The relationship between ADHD and ASMR could have significant implications. Currently, the common practice for treating ADHD involves medication (*e.g.* Adderall) (Sharkey, 2023). Immediate-release and extended-release are two main types of stimulants currently on market. Immediate-release stimulants may cause a significant drop in energy after their effect

wears off. In contrast, extended-release stimulants may take longer to start working, but can provide longer relief throughout the day, so they typically are not for alleviating sudden attacks of restlessness or anxiety. However, these stimulants may have side effects such as loss of appetite, stomach aches, nervousness, and disordered sleeping patterns. More research is needed in future to explore alternative or more optimal treatments for individuals with ADHD. In that case, ASMR shows promising candidacy.

3. CONCLUSION

To summarize, mental health has become a big concern among children and adolescents globally. Public health data further confirms that by showing increased rates of emotional distress, anxiety, depression, and stress that negatively affect their overall well-being. Sleep disorders have also become very prevalent among young individuals, further affecting their mental states. ASMR videos and apps may be a promising tool for individuals to reduce stress, improve sleep quality, and enhance overall well-being. Further research is needed to explore more potentials of ASMR in treating mental health disorders, and the underlying mechanisms. With advances in science and research, ASMR may continue to play a vital role in promoting mental well-being in the younger population as well as the individuals facing various mental health challenges.

Works Cited

- Andersen, J. (2014). Now You've Got the Shiveries: Affect, Intimacy, and the ASMR Whisper Community. *Television & New Media*, 16(8), 683–700.
<https://doi.org/10.1177/1527476414556184>
- Aron, E. N., & Aron, A. (1997). Sensory-processing sensitivity and its relation to introversion and emotionality. *Journal of Personality and Social Psychology*, 73(2), 345–368.
<https://doi.org/10.1037/0022-3514.73.2.345>
- Barratt, E. L., & Davis, N. J. (2015). Autonomous Sensory Meridian Response (ASMR): a flow-like mental state. *PeerJ*, 3(3:e851), e851. <https://doi.org/10.7717/peerj.851>
- Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: Differential susceptibility to environmental influences. *Psychological Bulletin*, 135(6), 885–908.
<https://doi.org/10.1037/a0017376>
- Bijlenga, D., Tjon-Ka-Jie, J. Y. M., Schuijers, F., & Kooij, J. J. S. (2017). Atypical sensory profiles as core features of adult ADHD, irrespective of autistic symptoms. *European Psychiatry*, 43, 51–57. <https://doi.org/10.1016/j.eurpsy.2017.02.481>
- Cheadle, H. (2012, July 31). *What is ASMR? That Good Tingly Feeling No One Can Explain*. VICE; VICE.
<https://www.vice.com/en/article/asmr-the-good-feeling-no-one-can-explain/>
- Clements, A. (2012). - *YouTube*. www.youtube.com.
https://www.youtube.com/watch?v=1IXrWgO-Odg&feature=youtu_gdata_player
- Greven, C. U., Lionetti, F., Booth, C., Aron, E. N., Fox, E., Schendan, H. E., Pluess, M., Bruining, H., Acevedo, B., Bijttebier, P., & Homberg, J. (2019). Sensory Processing Sensitivity in the context of Environmental Sensitivity: A critical review and development of research agenda. *Neuroscience & Biobehavioral Reviews*, 98, 287–305.
<https://doi.org/10.1016/j.neubiorev.2019.01.009>
- Grewe, O., Katur, B., Kopiez, R., & Altenmüller, E. (2010). Chills in different sensory domains: Frisson elicited by acoustical, visual, tactile and gustatory stimuli. *Psychology of Music*, 39(2), 220–239. <https://doi.org/10.1177/0305735610362950>
- Kaleva, I., & Riches, S. (2023). Stepping inside the whispers and tingles: multisensory virtual reality for enhanced relaxation and wellbeing. *Frontiers in Digital Health*, 5.
<https://doi.org/10.3389/fdgth.2023.1212586>
- Poerio, G. (2018). *RESTLESS COMPENDIUM : interdisciplinary investigations of rest and its*.
- Racine, N., McArthur, B. A., Cooke, J. E., Eirich, R., Zhu, J., & Madigan, S. (2021). Global Prevalence of Depressive and Anxiety Symptoms in Children and Adolescents during COVID-19. *JAMA Pediatrics*, 175(11), 1142–1150.
<https://doi.org/10.1001/jamapediatrics.2021.2482>
- Smith, S. D., Fredborg, B. K., & Kornelsen, J. (2020). Functional connectivity associated with five different categories of Autonomous Sensory Meridian Response (ASMR) triggers. *Consciousness and Cognition*, 85, 103021. <https://doi.org/10.1016/j.concog.2020.103021>

- Tang, S., Xiang, M., Cheung, T., & Xiang, Y.-T. (2020). Mental Health and Its Correlates among Children and Adolescents during COVID-19 School Closure: The Importance of Parent-Child Discussion. *Journal of Affective Disorders*, 279(1), 353–360. <https://doi.org/10.1016/j.jad.2020.10.016>
- Wilens, T. E., & Spencer, T. J. (2010). Understanding Attention-Deficit/Hyperactivity Disorder from Childhood to Adulthood. *Postgraduate Medicine*, 122(5), 97–109. National Library of Medicine. <https://doi.org/10.3810/pgm.2010.09.2206>
- Woods, N., & Turner-Cobb, J. M. (2023). “It’s like Taking a Sleeping Pill”: Student Experience of Autonomous Sensory Meridian Response (ASMR) to Promote Health and Mental Wellbeing. *International Journal of Environmental Research and Public Health*, 20(3), 2337. <https://doi.org/10.3390/ijerph20032337>

Homelessness: Working Towards Helping Poverty By Sonya Gupta

Introduction

Whether you are walking in NYC, driving in Los Angeles, or using the "L" train in Chicago, it is estimated that around 1 out of every 83 individuals you encounter is homeless ("Homelessness Is a Shared Experience in New York City,"). Around 36% of homeless people in LA have been homeless for more than a year (Tinoco). Envision living for over a year without access to fundamental requirements such as nourishing food, clean drinking water, and a cozy sleeping arrangement. 402,399 people in NYC, LA, and Chicago do not have to imagine it. The majority of individuals experiencing homelessness are women living in poverty. Women who often face wage gaps, domestic violence, and limited employment opportunities, have a higher risk of falling into poverty. The persistent issue of urban homelessness and women in poverty is driven by a complex interplay of economic inequality, housing shortages, and systemic injustices creating a clear moral imperative to take action.

The Cost of Living Crisis

Cost of living is one of the main reasons for homelessness in the United States, especially New York City, where escalating housing costs paired with the rise in prices for homes have and will continue to lead to homelessness. In July 2024, the median home price in NYC was around \$800,000 (realtor.com). As of 2022, the average individual income was around \$40,000 and the average household income was around \$76,000 (realtor.com). NYC homes are worth 10-20 times more than the average household income, creating larger issues around homelessness as families will become unable to afford a house in the city. To rent a one bedroom apartment in NYC it costs, on average, \$3,863 per month (apartments.com). For a family, a two bedroom apartment costs \$5,249 per month (apartments.com). The average yearly rent for a one bedroom apartment in NYC is higher than the average individual income, and even two people making an average salary would need 75 percent of their combined income to cover rent for an average 2-bedroom. This is without including basic necessities such as food, water, or clothes in their expenses. With this high of a rent and such a low income, homelessness in New York is inevitable. As housing prices continue to rise, and living wage jobs continue to wane, the city's poverty numbers will increase exponentially.

LA's enormous homeless population is also driven by its extremely expensive housing market. LA has the most expensive housing market in the US at an average of \$809,750 (Mehta). Los Angeles also boasts the highest homeless population in the United States, accounting for 25% of the nation's homeless population (Mehta). The median individual income for residents in LA is \$28,072 (Mehta). This is 29% of the cost of an average house. To rent a 2-bedroom apartment in LA, the average rent is \$2,719 per month (Rent Cafe). Although this amount is significantly cheaper than NYC, the yearly cost for a 2 bedroom apartment is still 58% of an average family's total income in a dual income household. This contributes to the massive homeless population of 45,252. This issue is similar to NYC, because if housing

costs are high and median income is low, people will be homeless. The presence of the most expensive housing market in the United States accounts for the poverty rates as affordable housing is rare for many without financial stability, such as the homeless population.

Chicago has a marginally cheaper rent than NYC and LA, but housing costs still drive high levels of homelessness. 185,155 residents are low-income renter households, but Chicago only yields 58,990 affordable and available rental homes for people at a low income level ("New Data Reaffirms Long Standing Shortage of Affordable Homes for Extremely Low-income Renters in Illinois and Chicago"). Therefore, a total of 126,165 low-income individuals do not have an affordable place to live. As of 2022, the average individual income for Chicago was around \$40,000 while the average house cost is around \$545,000. The average rent for a one bedroom in Chicago is \$1,863 per month, and a two bedroom is \$2,363 per month on average. This means a typical Chicagoan will spend 55% percent of their earnings on an average apartment. That enormous cost will continue to affect Chicago's homeless population of 5,149 people.

Gender inequality exacerbates the impacts of the cost of living crisis, driving especially high levels of homelessness for women. Women living in poverty face more challenges because of the disparities in income and their limited access to jobs. While women compose nearly half of the workforce, they account for 60% of the nation's lowest-paid workers ("Women and Poverty in America"). Women who might work harder than their male co-workers, have seniority over them, or have more experience are still often paid less ("Women and Poverty in America"). If many urban families are unable to afford housing, it follows that even more single mothers, who are paid significantly less than their male counterparts, will not be able to afford housing. Until systemic inequalities affecting women are eliminated, women will always be more vulnerable to homelessness.

Other legal precedents, though, have furthered the harms of homelessness caused by cost of living, particularly those affecting women. In 1980, the case of *Harris v. McRae* was where The U.S. Supreme Court upheld the Hyde Amendment, which restricted the use of federal funds for abortions, except in cases of life endangerment, rape, or incest (*Harris v. Mcrae*). This ruling disproportionately affected low-income women who rely on these funds for healthcare, limiting their access to abortion services and raising their healthcare costs, highlighting the intersection of poverty and reproductive rights. With the ongoing increase of homelessness in the United States caused by economic disparity, it is important to question why assistance for impoverished women is not being provided, since this will not effectively address the issue of homelessness.

Housing Shortages

Homelessness cannot be solved without adequate housing supply along with sufficient shelters for the unhoused, and the impacts of failures in both areas can be seen in major cities. In NYC, April 2024, 147,518 people slept each night in shelters and more than 200,000 people temporarily doubled-up in the homes of others ("Basic Facts about Homelessness: New York

City"). In April 2023, in NYC, only 24,730 single shelter beds were available ("Comptroller Lander Unveils New Dashboard to Track Shelter Population, Eviction, Housing Vouchers, & More"). That means 122,788 NYC citizens were without the ability to seek refuge in a homeless shelter, forcing the shelter to exceed its capacity. These overcrowded shelters then quickly become dangerous living situations for their residents. In 2021, significant incidents of violence, overdoses, and sex offenses reported were connected to the number of deaths in the city's shelter system, which rose by 58% compared to 2019 (Gross). One in five adult New York residents or 1.2 million people, live below the poverty line, and between 1996 and 2017, NYC lost 1.1 million affordable housing units. NYC now has 450 facilities across the city with a total capacity of 58,000 beds ("Comptroller Lander Unveils New Dashboard to Track Shelter Population, Eviction, Housing Vouchers, & More"). The most recent Housing Vacancy Survey indicates a vacancy rate of less than one percent for affordable apartments in NYC. As housing prices in NYC continue to rise, and living wage jobs continue to wane, NYC's poverty numbers will increase exponentially. The shortage of shelter beds and affordable housing increases the threat of danger for the thousands of unhoused as in NYC.

Chicago has made better efforts to improve the quality of life for their unhoused population, contrary to New York. The estimated number of individuals experiencing unsheltered homelessness decreased from 2022 to 2023, going from 1,263 to 990 ("Estimate of People Experiencing Homelessness in Chicago"). Chicago has invested time, effort, and money in housing to reduce the unsheltered homeless population. Since October 2020, the Expedited Housing Initiative's rapid rehousing efforts housed over 330 unsheltered households, including 117 individuals in 2022 ("Estimate of People Experiencing Homelessness in Chicago"). This lowered their homeless population by 13%. NYC and LA must take Chicago as an example as they continue to help its unhoused population. But Chicago still needs to do more to provide affordable housing outside of shelters— in Chicago, 185,155 residents are low-income renter households, but Chicago only yields 58,990 affordable and available rental homes for people at this income level. Therefore, a total of 126,165 low-income individuals do not have a place to live.

Women in shelters are especially vulnerable to sexual violence and abuse, making it especially important that affordable housing takes them out of that situation, and that less crowded shelters provide safer environments as a result. In a study of homeless women in shelters, 48% had experienced physical violence without a weapon, 18% with a weapon, and 18% had experienced sexual violence within the past six months ("Violence against People Who Are Homeless: The Hidden Epidemic"). Residing in a shelter should ideally provide a sense of security, but unfortunately, the majority of women in shelters have encountered acts of abuse. Many metropolitan areas have focused on providing adequate housing for the homeless population and women to combat the issue. The Violence Against Women Act or VAWA has led to lowered rates of gender-based violence in the United States due to its local shelters, transitional housing programs, and culturally specific support services ("The Basic Facts about Women in Poverty").

In addition to the VAWA, several instances have arisen that provide alternative living spaces for the homeless if a homeless shelter is unavailable. In 2018, *Martin v. The City of Boise* ruled that cities cannot enforce anti-camping ordinances if they do not provide enough shelter beds for homeless individuals (*Martin v. City of Boise*). This was a major case that changed the lives of many unhoused in Boise. This case established that punishing homeless individuals for sleeping on public property when no alternative shelter is available constitutes cruel and unusual punishment under the Eighth Amendment. In 2014, *Desertrain v. City of Los Angeles* struck down a Los Angeles ordinance that prohibited people from living in their vehicles, deeming it unconstitutionally vague (*Desertrain v. City of Los Angeles (9th Cir.)*). The ruling highlighted the need for clear and fair laws concerning homelessness and brought attention to the growing number of people living in vehicles due to economic hardship.

Other legal cases have prevented cities from abusing homeless people by establishing their right to live in public spaces when they couldn't afford a home. In 1992, *Pottinger v. City of Miami* was a landmark case that resulted in a consent decree that required Miami to provide more services and protections for homeless individuals and limited the city's ability to arrest people for life-sustaining activities in public spaces. The plaintiffs' complaint alleged that the City of Miami had a custom of arresting, harassing, and interfering with homeless people for engaging in basic activities of daily life including sleeping and eating in the public places where they were forced to live (*Pottinger v. City of Miami (Southern District of Florida)*). Due to inadequate income in Miami, citizens were forced to live in uninhabitable areas and were subject to harassment, and this case made a big impact for them by establishing the responsibilities the city had to them that weren't being met.

Although several organizations and cases have been established to assist the homeless community, there have been instances where cases, albeit appearing helpful, have inadvertently facilitated the occurrence of violence. In 1979, *Callahan v. Carey* paved the way for further legal victories that ensured the right to shelter for homeless men, women, children, and families in New York City. While this lawsuit guarantees that every homeless individual in NYC is entitled to a homeless shelter, likely, the prevalence of overcrowded housing, violence, and individuals turning to living on the streets would likely rise as a result. If every homeless individual is granted access to a shelter with limited capacity, the shelter would inevitably experience severe overcrowding, leading to a substantial increase in incidents of violence. With the ongoing increase of homelessness in the United States and New York especially caused by housing shortages, people in need should not face overcrowded housing, violence, and the need to resort to living in the streets.

Systemic Injustice

Systemic social injustices such as racism and sexism deeply entrench urban homelessness and disproportionately impact women in poverty, creating a cycle of hardship and limited opportunities for these vulnerable populations. In NYC, Black, Hispanic, and Latinx New Yorkers are disproportionately affected by homelessness. Approximately 56% of members

in shelters are Black and 32% are Hispanic ("Basic Facts about Homelessness: New York City"). Systemic racism in a variety of contexts limits opportunity for people of color, contributing to higher rates of homelessness. In NYC, 88% of the homeless population in shelters are people of color emphasizing how our system has failed to address the deep injustice people of color face. 4,727 Hispanic people and 5,528 Black people were recorded in DHS shelters in March 2022 ("Comptroller Lander Unveils New Dashboard to Track Shelter Population, Eviction, Housing Vouchers, and More"). By May 2023, the population of homeless Hispanics had a 415% increase and the homeless Black population had increased by 250% ("Violence against People Who Are Homeless: The Hidden Epidemic").

Similar problems exist in LA, where officials acknowledge the role of race in the crisis. In LA, African Americans represent 8% of the population ("Impacting Black History"). In a homeless count, Black people make up 34% of the 66,000 total ("Impacting Black History"). The Los Angeles Homeless Services Authority released the summation of county statistics this week saying that "Without institutional racism, there would be 15,000 fewer people experiencing homelessness, almost all coming from Black and Native American populations." The majority of the homeless population in LA consists of people of color, highlighting the structural inequalities that have compelled them to seek refuge in shelters.

The issue of urban homelessness reveals a complex web of socioeconomic challenges, one of which is the disproportionate impact of poverty on women, often exacerbated by systemic sexism. Women face many injustices such as the pay gap, caring for children, and their access to education. In 2002, women made 80¢ to the dollar ("For Women's History Month, a Look at Gender Gains – and Gaps – in the U.S."). After 20 years, it is reasonable to anticipate an increase in this figure but this number has only risen by 2¢. This sets women back severely as they are unable to accumulate savings, invest in education, or build financial security. For women of color facing multiple levels of systemic disadvantage, the effects are magnified.

The gender pay gap is not the only main factor of societal injustices creating poverty as single mothers also face significant challenges in achieving financial stability and escaping the cycle of poverty. In the United States, single mothers are more likely to live in poverty than single fathers or married couples ("Facts and Figures: Economic Empowerment"). A single parent must still allocate funds to provide for their child, but a mother, who earns significantly less than a father, will face greater challenges in accomplishing this. Approximately 30% of single mothers live in poverty, compared to 17% of single fathers and 6% of married couples. A single mother who works extremely hard to take care of her child will be restrained from being able to work a full-time job that will pay enough to support her family. Sexism plays a role as societal norms and workplace policies often fail to support single mothers adequately. Discrimination in hiring practices, lack of affordable childcare options, and limited access to social safety nets further constrain economic opportunities for single mothers.

Moreover, the economic constraints caused by limited access to education perpetuate a cycle of limited opportunity and reinforce gender inequalities. Women and girls in poverty often have less access to education compared to their male counterparts, leading them into a slippery

slope that leads to joblessness. Only 39 percent of women in the United States had completed four years or more of college in 2022 (“Percentage of the U.S. Population Who Have Completed Four Years of College or More From 1940 to 2022, by Gender”). Sexism restricts a woman's access to quality education, perpetuating cycles of poverty. This limits their ability to secure better-paying jobs and escape poverty, reinforcing gender disparities in economic outcomes.

Many legal precedents protect people of color because of their systemic disadvantages. In many areas, there have been instances where cases against homeless people of color have been overruled. *Seattle Housing Authority v. Collier* in 2008 was a major case where SHA initiated an unlawful detainer action to evict Collier, a single mother who is African American (*RESIDENT ACTION COUNCIL v. SEATTLE HOUSING AUTHORITY*). The Washington Supreme Court ruled in favor of Collier, finding that SHA did not comply with the specific notice requirements mandated by state law. This court decision reinforced the importance of following proper legal procedures before evicting tenants. This is crucial for people of color, who are disproportionately represented in public housing and are more likely to face eviction due to systemic inequalities.

Conclusion: What Can Be Done

While homelessness may appear to be an intractable issue in numerous cases, there are several measures that the average individual can undertake to enhance their standard of living. Assisting the homeless community requires little exertion, and there are numerous methods to contribute, such as providing donations or working at shelters. These straightforward actions can significantly enhance the well-being of the homeless community.

Donating to your local homeless shelter, an unhoused individual resorting to the street, or an organization or charity is one way to help the community. Donating clothing, socks, garbage bags, hygiene items, soap, toothpaste, feminine hygiene, toiletries or blankets can help many homeless people. Donations of clothes, blankets, and hygiene products offer protection from the elements, improve health, and boost self-esteem. These donations to organizations enhance medical provisions for the homeless, addressing health concerns that may be exacerbating their condition and reducing their reliance on healthcare expenses.

Connecting to donating, volunteering at a shelter is another way to support the homeless community. In a homeless shelter, many struggle with social isolation. Volunteers can provide companionship, build trust, and offer a sense of belonging. Interacting with caring volunteers can offer emotional support and encouragement, which can be crucial for overcoming challenges. Volunteering plays a significant, supportive role in helping individuals experiencing homelessness and strengthens organizations working towards solutions. As we tackle homelessness as a neighborhood, we all must tend to the needs of the homeless community and help improve their well-being.

Works Cited

- "5 Facts about the Homeless Crisis in Los Angeles." In *Los Angeles Mission*.
<https://losangelesmission.org/facts-about-homelessness/>.
- 2023 *Snapshot of Homelessness in Chicago*. Illustration.
https://www.chicago.gov/content/dam/city/depts/fss/supp_info/Homeless/2023PITCOUNT/2023%20PIT%20Count%20one%20pager%20FINAL.pdf.
- Apartments.com. Accessed September 20, 2024.
<https://apartments.com/rent-market-trends/new-york-ny/>.
- "Basic Facts about Homelessness: New York City." In *Coalition for the Homeless*. Last modified June 2024.
<https://www.coalitionforthehomeless.org/basic-facts-about-homelessness-new-york-city/>.
- "The Basic Facts about Women in Poverty." In *American Progress*. Last modified August 3, 2020. <https://www.americanprogress.org/article/basic-facts-women-poverty/>.
- "Comptroller Lander Unveils New Dashboard to Track Shelter Population, Eviction, Housing Vouchers, and More." In *New York City Comptroller Brad Lander*. Last modified April 18, 2024.
<https://comptroller.nyc.gov/newsroom/comptroller-lander-unveils-new-dashboard-to-track-shelter-population-eviction-housing-vouchers-more/>.
- Desertrain v. City of Los Angeles (9th Cir.).
<https://law.justia.com/cases/federal/appellate-courts/ca9/11-56957/11-56957-2014-06-19.html>.
- "Estimate of People Experiencing Homelessness in Chicago." In *Chicago Homeless*.
<https://www.chicagohomeless.org/wp-content/uploads/2023/08/Homeless-Estimate-2023-v7.pdf>.
- "Facts and Figures: Economic Empowerment." In *UN Women*.
<https://www.unwomen.org/en/what-we-do/economic-empowerment/facts-and-figures>.
- "For Women's History Month, a Look at Gender Gains – and Gaps – in the U.S." In *Pew Research Center*. Last modified February 27, 2024.
<https://www.pewresearch.org/short-reads/2024/02/27/for-womens-history-month-a-look-at-gender-gains-and-gaps-in-the-us/>.
- "Gender Equality and Education." In *UNESCO*.
<https://www.unesco.org/en/gender-equality/education>.
- Gross, Courtney. "Exclusive: Videos and Data Show Increased Violence in City Shelters." *Spectrum News*.
<https://ny1.com/nyc/all-boroughs/public-safety/2022/09/20/exclusive--videos-and-data-show-increased-violence-in-city-shelters>.
- Harris v. McRae, No. 79-1268 (Congress). Oyez.
- "Homelessness Is a Shared Experience in New York City." *The Bowery Mission. Rebuilding Lives Since 1879*. Last modified September 2023.
<https://www.bowery.org/homelessness/#:~:text=In%20a%20city%20of%20more,one%20of%20the%20other%20boroughs>.

"Impacting Black History." In *NHSLA*.
<https://www.nhslacounty.org/impacting-black-history-black-people-make-up-8-of-l-a-population-and-34-of-its-homeless/>.

Martin v. City of Boise (9th Cir.).
<https://law.justia.com/cases/federal/appellate-courts/ca9/15-35845/15-35845-2018-09-04.html>.

Mehta, Yanit. "L.a. Housing: Racism, Skyrocketing Prices and Now a Homeless Crisis." In *UCLA Blueprint*. Last modified 2021.
<https://blueprint.ucla.edu/feature/l-a-housing/#:~:text=The%20dimensions%20of%20the%20issue,poveerty%20rates%20in%20the%20country.>

"New Data Reaffirms Long Standing Shortage of Affordable Homes for Extremely Low-income Renters in Illinois and Chicago." In *Housing Action Illinois*. Last modified March 14, 2024.
<https://housingactionil.org/blog/2024/03/14/new-data-reaffirms-long-standing-shortage-of-affordable-homes-for-extremely-low-income-renters-in-illinois-and-chicago/#:~:text=For%20a%20single%20person%2C%20the,is%20urgent%20and%20long%20overdue.>

NSVRC. <https://www.nsvrc.org/statistics>.

"Percentage of the U.S. population who have completed four years of college or more from 1940 to 2022, by gender." Statista.
<https://www.statista.com/statistics/184272/educational-attainment-of-college-diploma-or-higher-by-gender/#:~:text=Education%20%26%20Science-,Percentage%20of%20the%20U.S.%20population%20with,degree%2C%20by%20gender%201940%2D2022&text=In%20an%20impressive%20increase%20from,more%20of%20college%20in%202022.>

Pottinger v. City of Miami (Southern District of Florida).
<https://law.justia.com/cases/federal/district-courts/FSupp/810/1551/1500109/>.

"Poverty Increases the Likelihood of Becoming Unhoused." Covenant House.
<https://www.covenanthouse.org/homeless-issues/poverty-and-income-inequality#:~:text=Poverty%20and%20income%20inequality%20play,of%20a%20job%2C%20and%20discrimination.>

Realtor.com. Accessed September 20, 2024.
https://www.realtor.com/realestateandhomes-search/New-York_NY/overview.

Rent Cafe. <https://www.rentcafe.com/average-rent-market-trends/us/ca/los-angeles/>.

RESIDENT ACTION COUNCIL v. SEATTLE HOUSING AUTHORITY, No. 80006-5..
<https://caselaw.findlaw.com/court/wa-supreme-court/1114220.html>.

Tinoco, Matt. "Your Homeless Neighbors Have Probably Lived in LA Longer than You." *LAIST*, December 7, 2018.
<https://laist.com/news/los-angeles-homeless-demographic-race-gender>.

"Violence against People Who Are Homeless: The Hidden Epidemic." In *UCSF*. UCSF.

"Women and Poverty in America." In *Legal Momentum*.
<https://www.legalmomentum.org/women-and-poverty-america#:~:text=70%25%20of%20the%20Nation's%20Poor%20are%20Women%20%26%20Children&text=Currently%2C%2035%20percent%20of%20single,raise%20their%20families%20in%20poverty.>

Considering Biological Sex in The Development of Dementia By Hoa-Yen Trinh

Abstract

Dementia is a neurological, cognitive disease characterized by a patient's impaired memory and judgment, often seen during the geriatric stages of life. It is estimated that close to 10.7% of people worldwide over the age of 65 have some form of diagnosed dementia. This large percentage creates undue financial and caretaking burdens on family members and heavily impacts healthcare systems. While a cure for dementia is not known, researching this condition paves the way for treatments that could help millions of families. However, for these treatments to be effective, one should consider demographic information, such as biological sex, as a key variable in their studies. The correlation of demographic differences to the development of dementia has yet to be thoroughly studied. To investigate potential sex differences in dementia-related cognitive decline, we used data from The Aging, Dementia, and TBI Study from the Allen Institute for Brain Science that included a variety of patient cognitive assessments—such as the Braak and the NIA-Reagan criteria. Our results indicate a significant correlation between biological sex and the severity of neurodegeneration. Our findings suggest that sex as a biological variable should be given deep consideration when assessing the prognosis of neurodegenerative disorders in human patient populations.

Introduction

Dementia is a neurocognitive disease that impairs memory and judgment and is often observed in the late stages of natural aging. It is thought to be caused by an excess of non-functioning proteins that build up in the brain, impairing regions and leading to malfunctions, including cognitive, memory loss, and disorientation [1]. The most common form of dementia is Alzheimer's disease, which is further characterized by the rapid deterioration and malfunction of one's memory abilities [1]. Furthermore, there are also other types of dementia, including Huntington's disease—where the nerve cells in the brain break down [2]—and Vascular dementia—a type of dementia that occurs when blood vessels in the brain are damaged [3]—each with their own symptom profiles and hypothesized causes [4]. The variation in types of dementia creates a challenge for scientists to develop any one treatment for all. There is an urgent need to devise treatments for this family of diseases, as dementia affects not only the diagnosed but also their loved ones, who are most often their caretakers or the ones spending thousands of dollars on care services. There are also societal impacts that come with the disease, with research and treatment also affecting global economies and healthcare systems. Treatment for dementia is limited, often serving just to diminish individual symptoms as there is no cure for the disease. Some current treatments for dementia include acetylcholinesterase inhibitors—donepezil, rivastigmine, and Reminyl—which are medicines that prevent an enzyme from breaking down acetylcholine [5]. There is also memantine, a medicine that blocks the effects of excessive glutamate [5]. However, it is unknown how the efficacy of these treatments relates to demographic information, such as biological sex. Addressing these knowledge gaps

regarding dementia will advance research to eventually reach a cure for dementia and help millions of families affected by the disease. One of the most significant challenges to developing treatment for dementia is the possible implications biological sex could have. Biological sex poses a challenge because if there is a difference, then doctors will have to curate specific treatments for each biological sex. Previously, the risk of developing dementia differed between men and women, which is an important starting point [6]. Research has shown women have two times the risk of developing dementia [7]. On average, females typically live longer than males, greatly increasing their chances of developing the disease [7]. What are the biological sex differences in the development of dementia, and what are their implications? Researching the differences is crucial to finding a cure and treatment for the neurodegenerative disease. It will also help advance the development of personalized therapies focusing on helping patients. We will focus on how tests about the development of dementia—specifically Braak staging and the NIA-Reagan criteria—are different in males and females.

Methods

The Aging, Dementia, and TBI study from the Allen Institute for Brain Science includes three data sets used in the current study to examine the differences between sexes: Donor Information, Protein and Pathology Quantifications, and Description of Stains, which explains the histological methods used in the quantifications used in the former dataset [8]. The dataset was found online. There were 107 donors in the study, all over the age of 77 [8]. Using the Donor Information data set from the Aging, Dementia, and TBI study, we isolated females and males by first sorting the sex column of the data set. There are different types of indicating tests for dementia, and we specifically used the Braak Staging and the NIA-Reagan criteria as a metric for the risk of developing dementia. The National Institute on Aging (NIA)-Reagan's neuropathologic criteria, which relies on neurofibrillary tangles and neuritic plaques, ranging from lowest to highest based on the amount of each [9]. Braak Stages go from I to IV, evaluating the progression of neurofibrillary tangle growth within the medial temporal lobe memory circuit from the lowest to the highest stage [10]. Using the averages of Braak stages and NIA-Reagan values from males and females, we created bar graphs to examine the differences between biological sexes in the development of dementia. Using alpha-synuclein immunohistochemistry, we further focused on the temporal cortex's deterioration, a key memory formation region. Each comparison between males and females will undergo a Student's two-sample t-test [11], assuming equal variance between groups to test for statistical differences between the means of the two groups.

Results

After creating graphs of the Braak staging method and the NIA-Reagan criteria, females were found to have greater values in both indicating methods. Females also had a higher average number of alpha-synuclein proteins in the temporal cortex.

Figure 1

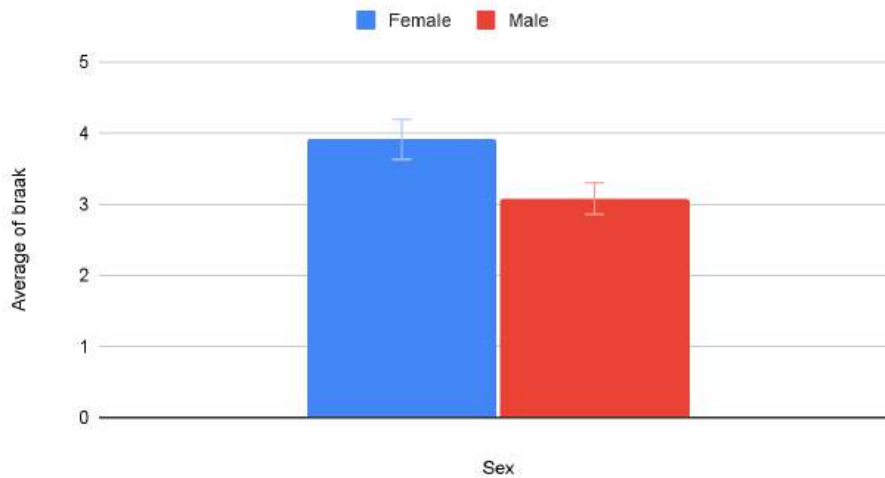
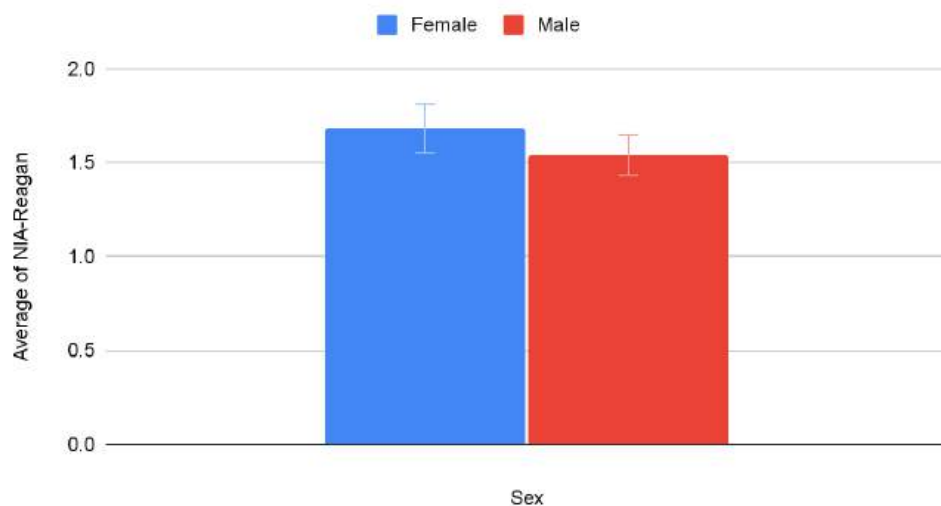


Figure 1 An unpaired-sample t-test was conducted to compare the means of Braak stage values between the two sex groups. Regarding the Braak staging method, females had higher averages than males, averaging stage 3.909, while males averaged 3.079. The standard error for females was 0.280, while males was 0.224. The p-value calculated was 0.013, below our predetermined alpha = 0.05 criterion for statistical significance. Females having a higher Braak stage suggest that, on average, they have a higher amount of neurofibrillary tangles in the brain than males do.

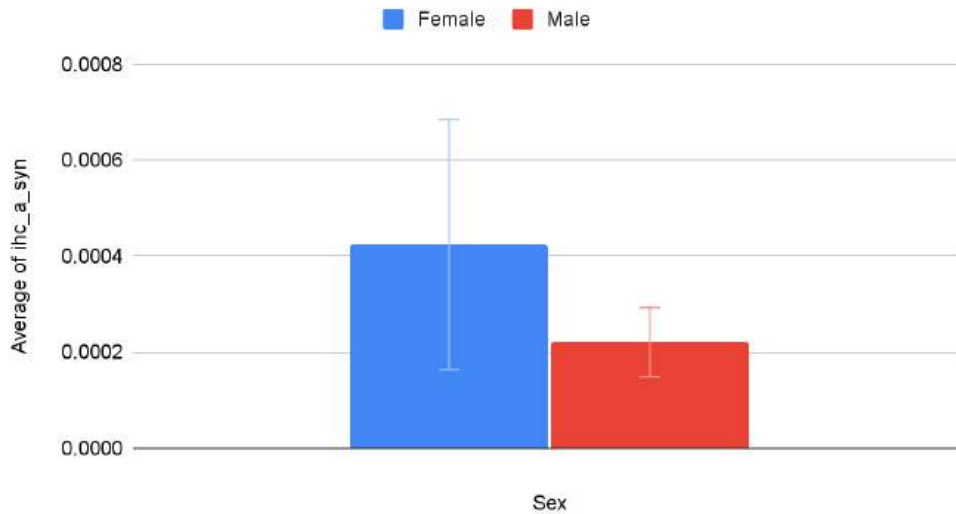
Figure 2



An unpaired-sample t-test was conducted to compare the means of NIA-Reagan values between the two sex groups. The average NIA-Reagan criteria values for females were higher than for males, averaging 1.68, while males averaged 1.54. The standard error for females was

0.129, while the standard error for males was 0.108. The p-value calculated was 0.40, which indicated that the difference between means was not statistically significant.

Figure 3



The average percent of alpha-synuclein proteins in females was higher than in males, at 0.000424%, while in males, it was 0.000221%. The standard error for females was 0.000260%, and for males it was 0.0000721%. The p-value calculated was 0.38, which indicated that the difference between the means was not statistically significant.

Discussion

Females have greater amounts of neurofibrillary tangles and neuritic plaques. Neurofibrillary tangles are accumulations of a protein called tau in neurons [12]. Tau collects together and creates tangles in the neurons, which hinder the neuron's functions and cause malfunctions in brain areas [12]. Females averaged at a higher Braak stage, which implies they have more neurofibrillary tangles in the brain, causing them to have dementia. On the other hand, males, on average, do not have as many neurofibrillary tangles, meaning they have a lower chance of getting dementia. The second method used is the NIA-Reagan criteria. Females, on average, also have a higher value. Neuritic plaques consist of deteriorating neuronal material [12]. A higher value in the NIA-Reagan criteria suggests females have greater neurofibrillary tangles and neuritic plaques in the brain, which hinder brain function. Conversely, males have lower amounts of neuritic plaques and neurofibrillary tangles, meaning they have lower amounts of non-functioning proteins, causing them to have a lower chance of getting dementia. Lewy bodies and Lewy neurites are abnormal protein deposits, including alpha-synuclein proteins, which positively correlate to neurological degeneration [13]. The average Braak stage graph's p-value was 0.013. This suggests a significant difference in the means of female and male Braak stages, which suggests a biological difference in the development of dementia between the sexes.

Previously, researchers had found that females were at a greater risk of developing dementia than males. It has been attributed to biological differences between females and males, such as menstruation, pregnancies, and menopause [7]. Still, it has also been attributed to traditional societal roles, specifically differences in work life and lifestyle [7]. The Braak stage difference between both sexes supports the idea that females have a greater risk of developing dementia. With females averaging a Braak stage of 3.909, while males with a 3.079, there is a clear disparity between the two. Females have a higher average Braak stage, meaning they are at a greater risk of developing dementia. The causes for this are still unknown and will have to be advanced by further research.

The neurological differences between sexes regarding dementia have a greater impact on treatments and how doctors diagnose patients. Doctors should prescribe medication and treatment accordingly and even take preventative measures, especially for biological females. Further, considering these biological and neurological differences between males and females will enhance knowledge and development of treatments, as well as future studies. Diagnosis will also change with this knowledge. It is important to note these different methods were used to diagnose dementia, and knowing that they differ between sexes can further inform scientists and doctors about advancing diagnosis. Knowing females have a higher chance of developing dementia, doctors should encourage screening females for dementia earlier in their lives to be proactive. Analyzing test results for females should also change; the scale of the tests should differ between sexes, with females having a lower value to be diagnosed with dementia, as they have a greater risk of developing it. Doctors should also curate specific tests between sexes, testing for sex-specific neurological dysfunction. Management of the disease will also be impacted, as there is a greater risk for females, which should be made known to the public to help prevent the disease.

Works Cited

- CDC. "What Is Dementia?" *Centers for Disease Control and Prevention*, 5 Apr. 2019, www.cdc.gov/aging/dementia/index.html.
- National Institute of Neurological Disorders and Stroke. "Huntington's Disease." *National Institute of Neurological Disorders and Stroke*, 25 July 2022, www.ninds.nih.gov/health-information/disorders/huntingtons-disease.
- National Institute on Aging. "Vascular Dementia: Causes, Symptoms, and Treatments." *National Institute on Aging*, 1 Nov. 2021, www.nia.nih.gov/health/vascular-dementia/vascular-dementia-causes-symptoms-and-treatments.
- National Institute on Aging. "Understanding Different Types of Dementia." *National Institute on Aging*, 2024, www.nia.nih.gov/health/alzheimers-and-dementia/understanding-different-types-dementia.
- National Health Service. "What Are the Treatments for Dementia?" *NHS*, 18 Aug. 2023, www.nhs.uk/conditions/dementia/about-dementia/treatment/.
- Li, Rena, and Meharvan Singh. "Sex Differences in Cognitive Impairment and Alzheimer's Disease." *Frontiers in Neuroendocrinology*, vol. 35, no. 3, Aug. 2014, pp. 385–403, <https://doi.org/10.1016/j.yfrne.2014.01.002>.
- Alzheimer's Society. "Why Is Dementia Different for Women? | Alzheimer's Society." *Www.alzheimers.org.uk*, 8 Mar. 2023, www.alzheimers.org.uk/blog/why-dementia-different-women#:~:text=Women%20have%20a%20greater%20risk.
- "Data Download :: Allen Brain Atlas: Aging, Dementia and TBI Study." *Aging.brain-Map.org*, 2004, aging.brain-map.org/download/index.
- Newell, Kathy L., et al. "Application of the National Institute on Aging (NIA)-Reagan Institute Criteria for the Neuropathological Diagnosis of Alzheimer Disease." *Journal of Neuropathology and Experimental Neurology*, vol. 58, no. 11, Nov. 1999, pp. 1147–1155, <https://doi.org/10.1097/00005072-199911000-00004>. Accessed 23 Mar. 2021.
- Malek-Ahmadi, Michael, et al. "Braak Stage, Cerebral Amyloid Angiopathy, and Cognitive Decline in Early Alzheimer's Disease." *Journal of Alzheimer's Disease*, vol. 74, no. 1, 10 Mar. 2020, pp. 189–197, <https://doi.org/10.3233/jad-191151>. Accessed 1 July 2020.
- Wadhwa, Raoul R., and Raghavendra Marappa-Ganeshan. "T Test." *PubMed*, StatPearls Publishing, 2023, www.ncbi.nlm.nih.gov/books/NBK553048/.
- Shwartz, Michael. "Alzheimer Disease | Definition, Causes, Symptoms, & Treatment | Britannica." *Www.britannica.com*, 29 Apr. 2024, www.britannica.com/science/Alzheimer-disease#ref107060.
- Rocha Cabrero, Franklyn, and Elizabeth H. Morrison. "Lewy Bodies." *PubMed*, StatPearls Publishing, 2020, www.ncbi.nlm.nih.gov/books/NBK536956/.

- Beach, Thomas G., et al. "Evaluation of α -Synuclein Immunohistochemical Methods Used by Invited Experts." *Acta Neuropathologica*, vol. 116, no. 3, 15 July 2008, pp. 277–288, www.ncbi.nlm.nih.gov/pmc/articles/PMC2708176/, <https://doi.org/10.1007/s00401-008-0409-8>.
- World Health Organization. "Dementia." *Www.who.int*, World Health Organization, 15 Mar. 2023, www.who.int/news-room/fact-sheets/detail/dementia#:~:text=Key%20facts.
- Bennett, D. A., et al. "Neuropathology of Older Persons without Cognitive Impairment from Two Community-Based Studies." *Neurology*, vol. 66, no. 12, 26 June 2006, pp. 1837–1844, <https://doi.org/10.1212/01.wnl.0000219668.47116.e6>.
- "What Happens to the Brain in Alzheimer's Disease?" *National Institute on Aging*, 19 Jan. 2024, www.nia.nih.gov/health/alzheimers-causes-and-risk-factors/what-happens-brain-alzheimers-disease#:~:text=Neurofibrillary%20tangles%20are%20abnormal%20accumulations.

The Intelligence Divide: Examining Gender Roles in Modern Marriage

By Katherien Lu

“I was not worried [that men would be intimidated by me]...A man who would be intimidated by me is exactly the kind of man I would have no interest in,” said Chimamanda Ngozi Adichie—feminist, activist, and acclaimed author (Adichie). This quote encapsulates the struggles and resilience of intelligent women in a world that often values them less than their male counterparts. While intelligent men are celebrated and overwhelmingly benefit from marriage, intelligent women frequently face challenges that stem from traditional views of gender roles. This essay explores why there is such a disparity in the likelihood of getting married between intelligent men and women.

This study is based on data from the UK and thus cannot be completely generalized. In a study that collected data from 42 countries, the UK ranked second highest in individualism (Diener et al.). Individualism is associated with support for gender equality in all aspects of society, along with higher levels of female employment and education, which challenges traditional gender norms (Davis and Williamson). Thus, we cannot assume that the same would be true for all countries, and particularly not for countries that score higher on collectivism than individualism. However, for the purposes of this paper, the cited data will be primarily from the UK & USA, and given that both countries score the highest on metrics of individualism, it can be inferred that they have cultural similarities in the context of gender equality and marriage (Diener et al.).

Marriage, as an institution, is inherently economic. What distinguishes marriage from a similar, long-term, romantic relationship is the added officialization that comes with merged assets, tax benefits, and legally bound commitments. This officialization transforms marriage from just a personal and emotional bond to something that is also a socio-economic partnership with distinct financial factors.

Historically, marriage was a way to ensure economic stability for many women who were not allowed to or had limited work opportunities. In the early 1900s, traditional gender norms reinforced the man as the breadwinner in the public sphere and confined the woman as the homemaker in the private sphere, effectively “trading” women’s housework for men’s wages. However, after World War II, there was a shift in women’s roles and perceptions towards work. As the majority of men were drafted to fight on the front lines, women were left to take over many essential jobs, such as manufacturing, agriculture, and healthcare. While most of these jobs did not last post-war, they were an indication not only to society but also to women that they were capable of working and earning money. This realization resulted in newfound confidence and self-actualization in women, creating waves of feminism and advocacy for gender equality following the war. As women gained more rights and working opportunities, they were less restricted by the economic constraints of marriage, and with increased self-sufficiency, they were able to increasingly devalue marriage as an economic investment and consider other factors before getting married.

Nowadays, women's attitudes towards marriage have completely shifted. Disregarding intelligence and other variables, most modern women are less likely than men to see marriage as necessary for fulfillment; in fact, "37 percent of men feel that a man can have a fully satisfying life without marriage," whereas "59 percent of women feel that a woman can have a fully satisfying life without marriage" (Kaufman and Goldscheider).

This discrepancy increases when taking intelligence into account. Intelligent individuals are generally more strategic when making decisions, often prioritizing long-term benefits over immediate gratification. They tend to take more time when approaching complex issues, and with marriage being a significant commitment, more intelligent men and women are typically more thoughtful and deliberate in their approach to it (Schirmer et al.). For men, this often means choosing to marry, which typically provides significant socio-emotional and economic benefits to them. For women, however, this may mean living a satisfying life outside of marriage, as they are more likely to recognize the potential drawbacks for women, such as invisible labor and lack of time affluence. Given that men typically benefit more from marriage, while women benefit less, the statistic that "for each 15-point increase in IQ, the likelihood of getting married increases by around 35% for a man but decreases by around 58% for a woman" aligns with the tendency of intelligent people to make decisions that best serve their long-term interests.

Overall, marriage offers individuals emotional stability and companionship, which are crucial for mental health. Studies have shown that married individuals report higher levels of happiness and lower levels of depression compared to their single counterparts. The emotional support that a spouse provides can be a buffer against stress and other life challenges, improving one's mental health (Wadsworth). This companionship ultimately reduces feelings of loneliness and isolation, which are significant factors in emotional suffering.

Furthermore, married individuals benefit from the structure and routine that marriage can often offer. With a spouse and potentially kids, there is more pressure to maintain a regular schedule, stick to healthier lifestyle choices, including regular exercise and balanced diets, listen to medical advice, and ensure overall well-being for the family. These healthy behaviors, which can be less common among single individuals, lead to better physical and mental health reported by married people (Robert G. Wood et al.). While both married men and women gain these health benefits, men are more likely than women to fully reap these benefits, as other factors, like invisible labor, may overshadow these perks for women.

Many studies have shown the overwhelming benefits that men typically receive from marriage, including social, emotional, and economic improvements. Research has shown that women provide "kin-keeping" benefits to men by strengthening their relationships with their children and other relatives (Kaufman and Goldscheider). This kin-keeping role often involves organizing family gatherings and events, maintaining communication with extended family, and nurturing social bonds, which men might not prioritize or manage as effectively on their own. As a result, married men often enjoy richer family relationships and a more robust support network, improving their overall emotional well-being.

From an economic standpoint, married men also overwhelmingly benefit. According to national USA studies, married men appear to earn 10 to 20 percent more than comparable never-married men (Ginther and Zavodny). This increase in income is called the marriage premium and could stem from a variety of factors, including the stability and motivation provided by a supportive partner, as well as societal perceptions that view married men more favorably. Both scenarios can positively influence their careers. Married men might be viewed as more responsible and stable, as they are now responsible for a spouse and often kids, revealing that they can be loyal and committed. These are qualities that are highly valued in professional settings, which can lead to better job opportunities and career advancement.

Not only do men benefit from being married, but more intelligent men may also be viewed especially favorably as marriage candidates. Higher intelligence is often associated with traits that are highly valued in long-term partners, such as problem-solving abilities, financial stability, social awareness, and competence. These traits make intelligent men more attractive to potential partners and also adhere to traditional gender norms of the male breadwinner, as they can offer better resources and stability for future family life. According to a study by Jaakko Aspara and colleagues, even when controlling for income, men with higher IQs are more likely to get married and stay married, suggesting that intelligence itself is an attractive trait beyond the economic benefits it may offer (Aspara et al.). Ultimately, highly intelligent men are more able to recognize the overwhelming benefits of marriage and are more desirable as marital candidates. Thus, they are more likely to get married, corroborating the 35% increase as IQ points increase.

On the other hand, women, especially more intelligent women, tend to overwhelmingly experience a disadvantage in marriage. This is mainly due to the phenomenon of invisible labor, which includes undervalued and overlooked forms of work, such as unpaid household tasks (Budd). Traditional conceptualizations of work are often focused on paid employment, neglecting the significant contributions of unpaid household work and caring for the family. Specifically, caring for kids, cleaning the house, cooking, etc., requires a significant physical, cognitive, and emotional effort that often is brushed off as “a woman’s job.” Many stay-at-home mothers are often viewed as jobless or that their work is “less important,” although the emotional and physical labor they do in the home is just as valid and crucial as an “official,” paid job.

Invisible labor in marriage often results in women managing most of the household responsibilities, even when they are employed full-time. This imbalance is evident in a study that surveyed 393 mothers in the USA and found that despite shifting gender roles, women still bear the brunt of household labor (Ciciolla and Luthar). Indeed, in the US, married women are more likely to have kids than unmarried women, and gender differences in household labor are most pronounced among couples with children (Osterman et al.; Ciciolla and Luthar). The time and effort dedicated to invisible labor not only inhibits women’s careers and economic advancement but also decreases and disrupts their “time affluence” — the feeling of having enough time to do what one wants and needs (Kasser and Sheldon). Without “time affluence,” many women may feel unfulfilled and lack personal time, leading to a loss of sense of self, decreased overall happiness, and poor mental health. Thus, strategic intelligent women, who are more likely to

have and value their higher-paying, full-time jobs, may avoid marriage if it hinders their career and well-being (Strenze). Moreover, intelligent, married women often face challenges in the workplace. Many employers may fear the costs of maternity leave, leading to covert discrimination when hiring and during work. These biases may hinder their careers and add another layer of difficulty for women who are already juggling significant responsibilities at home and work.

Intelligence doesn't only affect how a woman will consider marriage; it may also impact how others view intelligent women as potential marriage candidates. Contrary to how high intelligence increases a man's favorability as a marital candidate, high intelligence can potentially be harmful in the realm of marriage for women. Traditional gender norms often dictate that men should be the primary breadwinners and decision-makers within a household, creating male-centric power dynamics. As a result, women who demonstrate high levels of intelligence and professional success may seem more intimidating and less feminine to potential partners. This perception is rooted in long-standing cultural associations between femininity and qualities like docility and submissiveness, which may disadvantage intelligent women.

Furthermore, research indicates that women who out-earn their husbands may inadvertently challenge their partner's sense of masculinity. This phenomenon is called "status inconsistency," where, in this case, a woman's high status in one area (such as career or education) can potentially conflict with traditional expectations of her role in marriage and the quality of that relationship (Min Gong). Women in such positions may experience conflict or unhappiness within their marriages as traditional expectations clash with their professional ambitions and responsibilities. Studies have shown that a woman's employment is a potential risk factor for divorce, and that this risk is particularly pronounced when the wife's income exceeds that of her husband (South; Ono).

Despite significant progress, much remains to be done to achieve true gender equality. As Adichie's quote highlights, finding supportive partners can be difficult, and while intelligent men overwhelmingly benefit from marriage, intelligent women still face many challenges. Today, intelligent women frequently juggle professional success and household responsibilities, a burden not equally shared by men. Society must challenge traditional norms to ensure both men and women can equally enjoy the benefits of marriage. By valuing and supporting women's ambitions, we can move closer to true gender equality.

Works Cited

- Adichie, Chimamanda Ngozi. *We Should All Be Feminists*. Fourth Estate, 2014.
- Aspara, Jaakko, et al. "Types of Intelligence Predict Likelihood to Get Married and Stay Married: Large-Scale Empirical Evidence for Evolutionary Theory." *Personality and Individual Differences*, vol. 122, Feb. 2018, pp. 1–6. *DOI.org (Crossref)*, <https://doi.org/10.1016/j.paid.2017.09.028>.
- Budd, John W. "2. The Eye Sees What the Mind Knows: The Conceptual Foundations of Invisible Work." *Invisible Labor*, edited by Marion Crain et al., University of California Press, 2019, pp. 28–46. *DOI.org (Crossref)*, <https://doi.org/10.1525/9780520961630-004>.
- Ciciolla, Lucia, and Suniya S. Luthar. "Invisible Household Labor and Ramifications for Adjustment: Mothers as Captains of Households." *Sex Roles*, vol. 81, no. 7–8, Oct. 2019, pp. 467–86. *DOI.org (Crossref)*, <https://doi.org/10.1007/s11199-018-1001-x>.
- Davis, Lewis S., and Claudia R. Williamson. "Does Individualism Promote Gender Equality?" *World Development*, vol. 123, Nov. 2019, p. 104627. *DOI.org (Crossref)*, <https://doi.org/10.1016/j.worlddev.2019.104627>.
- Diener, Ed, et al. "Similarity of the Relations between Marital Status and Subjective Well-Being Across Cultures." *Journal of Cross-Cultural Psychology*, vol. 31, no. 4, July 2000, pp. 419–36. *DOI.org (Crossref)*, <https://doi.org/10.1177/0022022100031004001>.
- Ginther, Donna K., and Madeline Zavodny. "Is the Male Marriage Premium Due to Selection? The Effect of Shotgun Weddings on the Return to Marriage." *Journal of Population Economics*, vol. 14, no. 2, June 2001, pp. 313–28. *DOI.org (Crossref)*, <https://doi.org/10.1007/s001480000058>.
- Kasser, Tim, and Kennon M. Sheldon. "Time Affluence as a Path toward Personal Happiness and Ethical Business Practice: Empirical Evidence from Four Studies." *Journal of Business Ethics*, vol. 84, no. S2, Jan. 2009, pp. 243–55. *DOI.org (Crossref)*, <https://doi.org/10.1007/s10551-008-9696-1>.
- Kaufman, Gayle, and Frances Goldscheider. "Do Men 'Need' A Spouse More Than Women?: Perceptions of The Importance of Marriage for Men and Women." *The Sociological Quarterly*, vol. 48, no. 1, Feb. 2007, pp. 29–46. *DOI.org (Crossref)*, <https://doi.org/10.1111/j.1533-8525.2007.00069.x>.
- Min Gong. "Does Status Inconsistency Matter for Marital Quality?" *Journal of Family Issues*, vol. 28, no. 12, Dec. 2007, pp. 1582–610. *DOI.org (Crossref)*, <https://doi.org/10.1177/0192513X07300708>.
- Ono, Hiromi. "Husbands' and Wives' Resources and Marital Dissolution." *Journal of Marriage and the Family*, vol. 60, no. 3, Aug. 1998, p. 674. *DOI.org (Crossref)*, <https://doi.org/10.2307/353537>.
- Osterman, Michelle J. K., et al. "Births: Final Data for 2022." *National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, vol. 73, no. 2, Apr. 2024, pp. 1–56.

- Robert G. Wood, et al. *The Effects of Marriage on Health: A Synthesis of Recent Research Evidence*. U.S. Department of Health and Human Services, 30 June 2007, <https://aspe.hhs.gov/reports/effects-marriage-health-synthesis-recent-research-evidence-research-brief>.
- Schirner, Michael, et al. “Learning How Network Structure Shapes Decision-Making for Bio-Inspired Computing.” *Nature Communications*, vol. 14, no. 1, May 2023, p. 2963. *DOI.org (Crossref)*, <https://doi.org/10.1038/s41467-023-38626-y>.
- South, Scott J. “Time-Dependent Effects of Wives’ Employment on Marital Dissolution.” *American Sociological Review*, vol. 66, no. 2, Apr. 2001, p. 226. *DOI.org (Crossref)*, <https://doi.org/10.2307/2657416>.
- Strenze, Tarmo. “Intelligence and Socioeconomic Success: A Meta-Analytic Review of Longitudinal Research.” *Intelligence*, vol. 35, no. 5, Sept. 2007, pp. 401–26. *DOI.org (Crossref)*, <https://doi.org/10.1016/j.intell.2006.09.004>.
- Wadsworth, Tim. “Marriage and Subjective Well-Being: How and Why Context Matters.” *Social Indicators Research*, vol. 126, no. 3, Apr. 2016, pp. 1025–48. *DOI.org (Crossref)*, <https://doi.org/10.1007/s11205-015-0930-9>.

Understanding Portugal's Drug Decriminalization Policy: Public Awareness and Perceptions of Law 30/2000 By Agastya Rai Gupta

Introduction

Until recently, the general attitude of policymakers and law enforcers towards non-prescription or recreational drug use and addiction has been one of intolerance. Users are often perceived as lawbreaking criminals committing petty crimes, chasing their next fix due to their lack of self discipline and their pursuit of euphoria. This perspective fails to recognise the true nature of drug use, and how users often fall victim to their circumstances, or conditions outside their own control. Widespread and illicit drug use does, of course, have a variety of dangers, including but not limited to health issues (including death), addiction, as well as societal costs such as disturbance, accidents, costs of treatment in government facilities, and so on. Distribution of such substances is also a large contributor to violence, as a means of enforcement within distribution networks, or as a show of force for the benefit of rival distribution networks (European Union Drugs Agency, 2024). Largely, as of today drug policies exist solely to prevent the use of drugs and similar substances, clearly perceiving all drugs in a common light. However, upon the onset of the 21st century, nations have started incorporating a new perspective into their policymaking surrounding drug abuse and have seen favourable results. This perspective surrounds the idea that drug users are victims and patients rather than hardcore criminals striving to break the law as an act of greed or rebellion.

Current drug policies around the world still largely perceive users as criminals. Most notably in the West, the USA restricts drug use and possession through federal law (with the exception of 21 states which allow certain substances), using harsh laws and fines to punish. Drug policy in Mexico has allowed for more government organisations to tackle drug and drug related operations. Countries outside this sphere also have harsh drug penalties. Nations like Iran and Germany also have harsh drug laws, going so far as to conduct border checks for the purpose of drug enforcement. Iran, alongside nations such as Saudi Arabia and Vietnam, use executions/death penalty as means of punishing drug users.

Adopting this particular attitude ignores the underlying structural issues leading to uncontrollable drug use, which could include mental health issues, poverty, and so on. Nations such as Portugal and the Netherlands provide support frameworks for drug users, recognising addicts as patients in need of support rather than criminals wreaking havoc. These nations provide mental health and rehabilitation services rather than incriminating those vulnerable enough to resort to drug use.

In contrast, drug policy in Portugal has historically been less restrictive, allowing for a certain amount of drug use. This lets a regulated drug industry rise, rather than an illicit market remain underground and cause harm. Portugal's policy focuses on harm reduction over punishment, using social services, resulting in its population being better off. Against this backdrop, this paper attempts to explore the perception of ordinary Portuguese citizens on the drug problem, the existing policies and the need for political intervention in this matter.

In 2024, the Portuguese government ...

With the introduction of such policies, using a qualitative approach comprising semi-structured surveys and in-depth interviews, this paper attempts to test the awareness, understanding and perception of Portuguese citizens regarding the new law.

The knowledge gap identified and fulfilled by this paper is the relative beneficial effect of Portuguese drug policy and its potential for implementation in other nations in pursuit of the same effect. This study aims to uncover the effect of revamped Portuguese drug policy and how it has impacted locals, through the use of structured questionnaires and semi structured interviews. The study's objective is to determine the specific consequences of particular changes in Portuguese drug law.

Literature Review

Global Responses to Drug Use, Control and Prevention

International organisations, such as the United Nations, have adopted several conventions (such as the “Single Convention on Narcotic Drugs , 1961”) that act as barriers for individual nations to allow for less restrictive drug policy, since they strongly advocate for criminalising drug use, drug production and distribution. While attitudes of member nations have been shifting, policies have remained unchanged due to a lack of consensus, mostly from countries resistant to change in drug policy, mainly including nations from Eastern Europe, the Middle East, Asia. For the most part, the United Nations has not allowed the distribution or use of drugs, having banned it under numerous conventions (including the UN’s Convention in 1961) in the past. This has served as a substantial barrier to legalisation or decriminalisation in countries worldwide. However, of late, one observes that UN experts have issued statements to support the decriminalisation of drugs, stating that there is more harm done than good by maintaining the war on drugs.

Drug Policy in Portugal

Portugal has been one of the few pioneering nations to decriminalise drugs in the year 2000, much before other nations. Portugal has typically had a less restrictive but effective policy, reducing their overall drug use to a sustainable level, ending the drug crisis they faced in the 1990s. Before the implementation of the landmark 30/2000 policy, the Portuguese government’s first notable response to the drug crisis was the establishing of treatment and rehabilitation centres, starting with the TAIPAS centre in the 1980s. Additionally, general hostility by law enforcement and lack of provisions for drug users (such as needle exchange programs) served to discourage the use of drugs. Later in 1999, the Portuguese Institute for Drugs and Drug Addiction was established to gather information regarding drug use, available to the public, but also used for the educating of experts in the field of rehabilitation and treatment. In 1998, the Commission for the National Strategy to Combat Drugs was created, with the purpose of suggesting policy/policy changes to reduce drug addiction. Eventually, this Commission

proposed the National Strategy to Combat Drugs, which proposed the decriminalisation of drug use and the implementation of treatment and welfare services for drug users. This policy was further modified through discussions in Parliament as well as in public sessions to eventually reach Law 30/2000 (Domosławski, 2011).

It is strongly believed that the drug crisis was a result of Portuguese soldiers who occupied other countries (to uphold Portuguese colonisation) bringing back such drugs. Whether this is true or not, the return of Portuguese soldiers and the rise in drug use has a strong positive correlation (Domosławski, 2011). Portugal's 30/2000 law has allowed citizens to hold a certain amount of drugs for ten days (1 gram of heroin or 2 grams of cocaine as per Balch O. (2024)) without paying any penalties or facing any charges. The law has also resulted in the creation of Commissions of Dissuasion, comprising of medical and social professionals, which has been able to significantly reduce the incidences of excessive drug use by individuals. They determine and implement measures on a case to case basis to reduce harm to an individual who indulges in excessive drug use. 30/2000 is also accompanied by a variety of programs such as needle exchanges meant to prevent the reuse, or use of the same needle by multiple people, when consuming drugs in order to prevent the spread of infections such as HIV and AIDS. Considering around 1.3 million needles are distributed yearly, this step was highly effective in reducing the level of HIV and AIDS infections in the country (Slade H. (2021)). Portugal has also implemented increased treatment services including the introduction of mobile overdose prevention centres and the direct distribution of safe smoking kits and packs.

The impact of the 30/2000 Policy in Portugal

As a result, Portugal and its citizens have experienced a variety of benefits. Portuguese deaths caused by drugs fell only a few years after the implementation of the policy. When compared to the rest of Europe, Portugal had a much lower drug death rate after implementing the policy. There have overall been fewer arrests and generally less crime associated with drugs and drug use. Portugal has also seen much lower rates of HIV, regarded as the result of lesser use of injected drugs, as well as the use of clean needles amongst those who do use injected drugs. The clean needles are credited to the previously mentioned needle exchange programs.

Portuguese law enforcement has also been able to reallocate efforts towards detecting and dismantling larger criminal systems which enable the illegal sale of drugs in harmful quantities or of unsafe qualities. Drug users in Portugal are, in general, treated much better than they were before the implementation of this policy, which focuses on harm reduction rather than punishing drug users. Users with addiction or overuse problems are provided with support services and facilities to rehabilitate.

Methodology

This study uses a qualitative approach to study people's perception, understanding, and awareness of the drug policy in Portugal. Semi structured interviews were taken, followed by online surveys distributed to interviewees.

For the semi structured interviews, participants were randomly selected from an office outdoor leisure area, and then brought to a separate room for questions. Participants were of young age, of about 20-35 years. All were office workers with stable jobs. All participants were native to, and residents of, Portugal, more specifically Oporto (or Porto) city where this study was conducted. Interviews were conducted during break time for these employees. All interviews were conducted in the month of June 2024.

For the online questionnaire, participants of the semi structured interviews were provided with a link, and made to enter responses, after which they were asked to send the link to others they frequently interact with. This is a form of snowball sampling. For the survey, respondents were assured of anonymity and confidentiality and were asked for their due consent. The respondents of the survey largely had similar characteristics to the participants of the semi-structured interview, especially considering the questionnaire was made available to networks of those participants.

Research Findings and Analysis

1. Awareness

For the success of any policy, it is essential for citizens to be aware about its provisions and implications. The survey revealed that a majority of people, 68.8% specifically, were not aware of the policy at all, with only 18.8% firmly reported that they were aware. The remaining 12.5% were unsure, opting for “maybe”; when collated with the interview responses, it can be inferred that such respondents had heard of the policy, but were unaware of the provisions and its possible implications. In a similar fashion, 64.3%, a majority once again, had not heard of the Commissions for the Dissuasion of Drug Addiction, which are a vital component of the policy, playing an active role in making it effective and driving harmful drug use downwards. The remaining 35.7% of participants were aware of the Commissions.

Broadly, the responses to the policy and to the Dissuasion Commission could be categorised as:

- a. Participant does not have any knowledge.
- b. Helps by providing mental healthcare and addiction treatment.
- c. Monitors individuals with addiction or overuse problems.

Since few people were aware of the policy itself, it is not surprising that there is low awareness about the commissions as well, which are a critical component of the policy. Given the prevalence of a lack of knowledge on the end of the participants regarding this policy and its provisions, it is possible that this policy was not widely communicated or advertised by the government. For a policy that, in addition to state support, is strongly dependent on active, response and participation of citizens, such a gap in awareness could prove to have a strong impact in terms of implementation. It is worth noting that multiple participants, at numerous

times in the questionnaire or during the survey, stated they cannot make informed opinions or that they do not know enough to offer informed opinions.

2. Perceptions regarding the policy

Keeping in mind that only 18.8% of participants knew about the law and its provisions, some participants were made to speculate about these provisions when asked about it. 80% of participants believed medical drugs were the primary focus of the law, and that it was majorly these sorts of drugs that had been decriminalised. The remaining 20% believed it was mainly non medical drugs allowed by the policy. 86.7% of participants further speculated that this policy would benefit medical patients, while 13.3% of participants did not believe so and argued that decriminalisation of any sort of drug may prove to have harmful implications. A 20% minority of participants believed that this policy could not be misused, with a third of participants believing that this policy could, in fact, be misused. The remaining 46.7% gave no definite response, opting for a “maybe.”

Overall, participants were broadly observed to either argue that drug use and distribution, in general, irrespective of whether they are medical or non-medical, need to be controlled, or that there might be an increase in the use of drugs due to easier accessibility. During the in-depth interviews, participants mentioned several stories of hearsay incidences of drug misuse.

Supporters for this law argued that such a policy could be beneficial for medicinal purposes, especially for individuals who suffer from some form of health morbidity such as diabetes, cancer, chronic kidney or lung diseases, moto-neuro disorders, or diagnosed mental health issues where patients are expected to be on regular prescription drugs and are in constant need to buy medicines in bulk for regular usage. Participants further added that such a policy would also establish greater control over illegal consumption of drugs, as it would regulate and control the purchase of non-prescription or recreational drugs and in turn lead to minimisation of drug-related crimes. Finally, in case of the latter situation, supporters of this Act were also pleased as this law would treat individuals who overuse drugs as patients rather than criminals, which they argued could help such individuals move closer to a path of recovery and rehabilitation.

From the above responses, it seems that participants observe or speculate the downsides of the law revolve solely around the potential of increased use, likely keeping in mind public health rather than other factors such as economic impacts, associated crimes, since there is no mention of this in responses to questions asking about negative impacts of this policy. Positive impacts suggested by participants however, do consider another factor, namely drug abuse and crime, alongside public health and the treatment of drug users as patients.

Despite the divided opinion on the law, both groups of respondents also spoke about certain areas of caution. They added that it was important to keep in mind that easier access to drugs as a result of this policy had the risk of causing greater addiction. They also believe this could cause people to be less aware regarding the danger of drug use, assuming the decriminalisation means drugs are “safe” or have no impact on health.

Participants believe that drug decriminalisation largely involves the increased sale and consumption of drugs without penalty as well as treating drug users as patients rather than criminals.

Approximately 67% of participants argue that decriminalisation of drugs as per the policy has not led to safer use of drugs. The remaining third of participants has observed that decriminalisation has led to safer use.

Participants generally seem to understand that this policy involves greater drug use, and also believe that this policy has had the possibility of increasing drug addiction by reducing restrictions on drugs. Participants also understand the potential of other negative impacts also existing, while a majority also state that they observe no benefit of this policy in terms of drug use and safety.

3. Perceptions regarding drug use at large

60% of participants believe there is a drug problem in their city, while 26.7% of participants did not believe so. The remaining 13.3% of participants opted for “maybe.” Participants largely believe that increased drug use is correlated with factors such as age and income level. Specifically, most agreed that it was majorly younger people (ages 15-25) or poorer people who used more drugs. Some, but few, participants also voiced the opinion that people of ages 30-50 also made up a substantial portion of drug users.

Clearly, individuals observe, or believe they observe, drug problems in their city, despite the deployment of various policies that other sources claim to be effective.

Discussion

During interviews, a participant mentioned a drug and crime related issue, which was later confirmed by other participants. Participants spoke of the demolishing of a slum area, which they also claimed was inhabited majorly by drug users, and hence was a concentrated area of drug related activity. The demolishing and repurposing of this area for the construction of a luxury apartment complex led to the displacement of these drug users, which participants claimed led to drug related activity, particularly crime, being more spread out in the city. “Other neighbourhoods were flooded with criminals and drug addicts.”

The findings serve to question the effectiveness of Portuguese drug law, showing that citizens are not only largely unaware about it, but also do not believe it has been effective, contrasting previous findings shown in the Literature Review section. Most participants used in the study did not realise this policy exists and, after being introduced to it, stated that it was not effective since they still noticed drug related (disruptive) activities such as crime in their city. A majority of participants claimed they believed there is a drug problem in their city, with less than 30% of respondents being able to firmly argue otherwise.

It is feasible to assume that Portuguese policy has in fact been effective in reducing the large amounts of drug related crime and health issues observed in the 1990s, as posited by past research (seen in the Works Cited section, and also very briefly mentioned in the Literature

Review section), but that drug related crime still persists to a noticeable degree, or has at least been noticeable to the participants used. Participants may have responded in this way as they genuinely have noticed disruptive drug related activity, however are unable to compare current levels of nuisance caused by drug activity with those before the implementation of the law, perhaps due to young age, lack of exposure, insufficient information, lack of awareness, among other reasons.

Additionally, since most participants have not known about the policy, it is logical to infer its success or impact would also be relatively less known as well. Lack of awareness could be a result of poor communication on the part of the Portuguese government, meaning this program may not have been as effective as intended. If properly communicated to citizens, this policy could have helped a greater number of citizens.

Anecdotes brought up by interview participants (corroborated by other participants) claimed that the demolishing of a slum caused “other neighbourhoods [to be] flooded with criminals and drug addicts.” This brings to mind related failures of the government such as :

1. A lack of aid (or accessibility of aid) for drug addicts in slums similar to the one in the mentioned anecdote.
2. Preventing harm to third parties not involved in drug use, perhaps due to weak law enforcement strategies
3. The ability of the government to address underlying structural causes of drug use such as previous housing crises (possible among other causes).

Since broader research conducted on a larger scale concludes that Portuguese policies have been effective in reducing the harm of drug use, while the respondents of this study continue to argue that drug related disruption still exists, it is reasonable to come up with the following conclusion. In the last 20 years of this policy, Portugal has successfully reduced the level of individual harm to drug users, reducing the number of death and disease caused by drugs and related activity. However drug use still exists today to a level which is noticeable to citizens who have nothing to do with it.

The Portuguese people, government, and the policy may be better off with more community involvement, such as through increased community meetings or leveraging the pre existing Commissions for Drug Dissuasion by extending their roles to also include collecting information which can be used for further policy making.

This paper implies that Portuguese drug policy has been effective, however aspects such as government-to-citizen communication must be relooked at by policy makers. Furthermore, this paper also implies that this policy could be further revised to tackle lower level drug use observed by citizens, potentially by involving citizens in the process of identifying and reporting harmful or disruptive drug use and related activities (perhaps through local level community meetings for example).

Conclusion

This study aimed to determine the effectiveness of Portugal drug policy, specifically the 30/2000 law implemented in November, 2000. The literature review shows that this policy has in fact been effective in reducing the harms of drug use, and reduced drug use in general, however the results of the questionnaire used as the chief source of primary data in this study show reservations still maintained by citizens. The knowledge gap addressed here is the real world impact of these policies, as observed by citizens. This gap was addressed through primary research, used to collect data from Portuguese citizens.

The findings of this paper suggest that while overarching coverage posits Portuguese drug policy as entirely successful, citizens do not perceive it in the same light. This could serve as a basis to question the effectiveness of this policy. In order to ensure accurate interpretations of this research by readers, it is essential to consider potential bias or confounding variables that may exist that could serve to distort the findings or conclusion. Firstly, the sample size was quite small, consisting of 16 participants who were administered a questionnaire, and an additional 5 participants who took part in face to face semi structured interviews. The small sample size limits the generalisability of the opinions and perceptions presented by participants in the study. Furthermore, being a non-resident, the researcher lacks a local context, meaning the paper may not be entirely accurate in reporting social or cultural settings. Furthermore, the difference in languages may act as a potential language barrier. While face to face interviews were conducted in English, and participants were able to respond in coherent English, their interpretation skills were untested, meaning it is not known if they interpreted (and hence responded) to questions in a manner they did not mean to. The same issue persisted for participants who were administered a questionnaire. Overall, the paper challenges the idea that Portuguese has been as effective as displayed in public discourse, by showing that the Portuguese citizenry finds it ineffective.

Works Cited

- Domosławski A. (2011), “Drug Policy In Portugal: The benefits of decriminalising drug use”, Open Society Foundations.
- Cima J. & Almeida A. (2015), “The Needle Exchange program in Portugal- the social health gains”, European Journal of Public Health.
- Slade H. (2021), “Drug Decriminalisation in Portugal: Setting the Record Straight”, TRANSFORM Drug Policy Foundation.
- Bajekal N. (2018), “Want to win the war on drugs? Portugal might have the answer”, TIME.
- “Time to end Prohibition” (2021), Global Commission on Drug Policy.
- Doward J. (2016), “The UN’s war on drugs is a failure. Is it time for a different approach?”, The Guardian.
- “UN experts call for end to global “war on drugs”” (2023), Office of the United Nations High Commissioner for Human Rights.
- “Special points of interest” (2023), United Nations Office on Drugs and Crime.
- Kundrod K. (2015), “Decriminalization of drugs in Portugal: a controversial experiment for public health”, Perspectives on Business and Economics: Lehigh Preserve.
- O Neill Institute, Georgetown Law (2023), “Quick Take: Portugal’s pioneering approach to drug policy: Lessons Learned.”
- European Union Drugs Agency (2024), “EMCDDA and Europol spotlight drug-related violence in latest market overview.”
- Johnson B. (2009) “Patterns of Drug Distribution: Implications and Issues.”
- Drug Policy Alliance (2019) “Drug Decriminalisation in Portugal; Learning from a Health and Human-Centered Approach.”
- Moury C. and Escada M. (2022) “Understanding successful policy innovation: The case of Portuguese drug policy”
- Balch O. (2024) “‘It beats getting stoned on the street’: how Portugal decriminalised drugs- as seen from the ‘shoot up centre’”, The Guardian
- Félix S., Portugal P., Tavares A (2017) “Going after the Addiction, Not the Addicted: The Impact of Drug decriminalization in Portugal”, Institute of Labor Economics
- Blackstone S. (2012) “Portugal Decriminalized All Drugs Eleven Years Ago And The Results Are Staggering”, Business Insider.
- BBC (2023) “Portugal’s drug consumption rooms are important, users say.”
- McCarthy N. (2020) “Then and Now Portugal’s Drug Decriminalization”, Statista.com
- Mann B (2024) “How Portugal eased its opioid epidemic, while U.S. drug deaths skyrocketed.”, NPR.
- Brown C. “How Europe’s heroin capital solved its overdose crisis” CBC.

Exploring Contextual Learning Using Convolutional Neural Networks to Enhance Autonomous Driving By Soo Ban Yoon

Acknowledgments

Made possible with the guidance of Jason Jabbour from Harvard University throughout the development of this research paper.

Abstract

Autonomous vehicles depend on perception systems to navigate and make critical driving decisions. These systems play a pivotal role in tasks ranging from path determination to dynamic response strategies like stopping at traffic signals or adjusting speed around unpredictable hazards. However, current autonomous systems lack the nuanced contextual understanding that human drivers naturally employ, such as distinguishing between actual road signs and similar images on billboards. This paper seeks to bridge this gap by enhancing the contextual awareness of autonomous vehicles. We introduce a novel dataset featuring typical driving scenes with intentionally placed anomalies to challenge perception accuracy. Utilizing this dataset, we train two Convolutional Neural Network (CNN) models: one to detect the presence of out-of-context objects within the scene, and another to pinpoint their locations. This research represents a step towards providing autonomous vehicles with a more human-like understanding of their environment, thereby improving their operational safety and efficiency.

Index Terms Autonomous Driving, Contextual Understanding, Convolutional Neural Networks, Machine Learning

I. Introduction

The concept of autonomous vehicles has recently captured widespread global attention, leading to the integration of automated driving features in modern vehicles. Tests with driverless cars are performed all around the world, suggesting the beginning of the global shift to fully automated vehicles (Hörl et. al., 2016). Some, including Bansal and Kockelman (Bansal et. al., 2016) predict 25% of vehicles being level 4 autonomous – meaning that they are essentially fully autonomous and will not require “drivers” to actively be prepared to take over (What Are the Levels of Autonomy in Self-Driving Cars? – Imagination, 2023) – by the year 2045. Regardless of these expectations, currently, no car is yet fully autonomous due to various flaws in the autonomous vehicle systems. Consequently, drivers are required to keep their hands on the wheel at all times, even while the vehicle is driving itself.

The development of fully autonomous vehicles and the rise in the number of these vehicles are expected to solve many of the problems regarding transportation. For example, autonomous vehicles have tremendous potential in reducing car accidents happening every day throughout the world. In the United States in 2011 alone, there were more than 5.3 million car automobile crashes, resulting in more than 2.2 million injuries and more than 32,000 deaths. Alcohol was

involved in more than 39% of the crashes recorded above. This alone demonstrates that nation-wide transition to level 4 autonomous vehicles could prevent approximately a third of the crashes in the United States (~1.7 million crashes) (Anderson et. al., 2014). With the imminent perfection of image classification algorithms implemented in the decision-making process, it's only a matter of time before people are finally able to trust these vehicles with their lives and take advantage of the benefits that follow.

However, the field of autonomous vehicles is still in its developmental stages and has a lot of room for improvement. The predominant failure with self-driving features stems from poor image classification describing how a vehicle interprets the world around it through a suite of sensors such as radars, LiDARs, and cameras. Their functionality relies on using sensory data to conduct image classification and obstacle detection. Normally, there is a rotating LiDAR sensor on the top of the vehicle with a 60-meter range that creates a 3-D map of the environment around the car. Ultrasonic sensors and radars on the front, back, and sides of the car determine the short and long-range distances from objects surrounding the car, respectively. Finally, autonomous vehicles have dozens of cameras placed around the ego to gain a 360-degree view of the world. For example, Waymo's autonomous vehicle has 20+ sensors around the vehicle consisting of cameras, radars, and lidars (Learn How Waymo Drives - Waymo Help). Motional, a Hyundai subsidiary, also includes a suite of 30+ sensors on their IONIQ 5 Robottaxi (How Motional Built Our Next-Generation Robotaxi in Less than a Year, 2024; Super Sensors: How Motional's AVs Can "See" Better than a Human Driver, 2022).

Machine learning utilizes sensory data to learn from and interpret a vehicle's surroundings, determine the presence or lack of anomalies, calculate planning trajectories, and control the necessary actions to reach the destination by plotting a motion plan (Alqahtani & Kumar, 2024). For this process to be effective, the interpretations of the vehicle's surroundings provided to the final decision-making algorithm must be perfectly accurate to ensure operation without any crashes. Unfortunately, there have been numerous errors in these interpretations, such as the inability to use contextual clues effectively. For example, an autonomous vehicle may not distinguish between an actual fire hydrant and a picture of a fire hydrant printed on the back of someone's sweater.

A person can quickly determine if an object is an obstacle, whether it's dangerous, and if action needs to be taken to remove it. This is because humans can utilize their surrounding environment to assess the state of the environment and gauge appropriate responses. Unlike humans, currently, self-driving vehicles have not been very successful in assessing the environment around these objects and responding accordingly. For instance, according to ABC News (Waymo's Robotaxi Expansion beyond SF Temporarily Suspended as Part of Review, 2024), a self-driving taxi in San Francisco, California, called Waymo, was driving a passenger to their destination when it approached a squad of firefighters extinguishing a fire. The Waymo failed to recognize the scene and drove over a fire hose, which, to the vehicle, appeared as just a small bump in the road. For autonomous driving to be widely accepted by society, it must be perfected. Any small error can result in fatal consequences, and there is no room for such errors.

This paper explores a method to enhance the image classification currently used in autonomous vehicles by incorporating contextual clues from an object’s surroundings to improve a machine-learning model for image classification that also accounts for environmental assessment.

To tackle this challenging task, we propose a novel dataset comprising street images collected from cars driving around a city that are modified with out-of-context objects. This dataset includes images captured from multiple perspectives: front, front-left, back-left, front-right, back-right, and back. We enhance a selection of images from other datasets by inserting out-of-context objects to test whether a convolutional neural network (CNN) can detect and locate these anomalies. We aim to train a CNN not only to classify objects but also to consider the surrounding environment. A CNN operates by detecting features across a series of layers and identifying patterns within the images. For autonomous vehicles, this network incorporates images from various angles, layering different types of data so that the final output reflects a comprehensive consideration of all factors.

II. METHODOLOG

II.1 Novel Dataset Creation

To train a neural network capable of making decisions based on contextual clues, we created a modified dataset featuring a background image with an added anomaly. This dataset would be used to develop a neural network that could perform two tasks: detect the presence of anomalies in the image and identify their locations.

The background dataset we used is called NuScenes (Nuscenes.org, 2021), which consists of street-view images captured by six different cameras, each providing a unique perspective. These images serve as the backgrounds for the modified dataset with added anomalies. The perspectives include front, front left, front right, back, back left, and back right. This multi-camera setup gives the program and the neural network a comprehensive view of the car’s surroundings, aiding in the decision-making process. The dataset contains a total of 2,424 images, with 404 images per camera perspective, depicting the environment around the car, including buildings, roads, other vehicles, trees, people, and more.

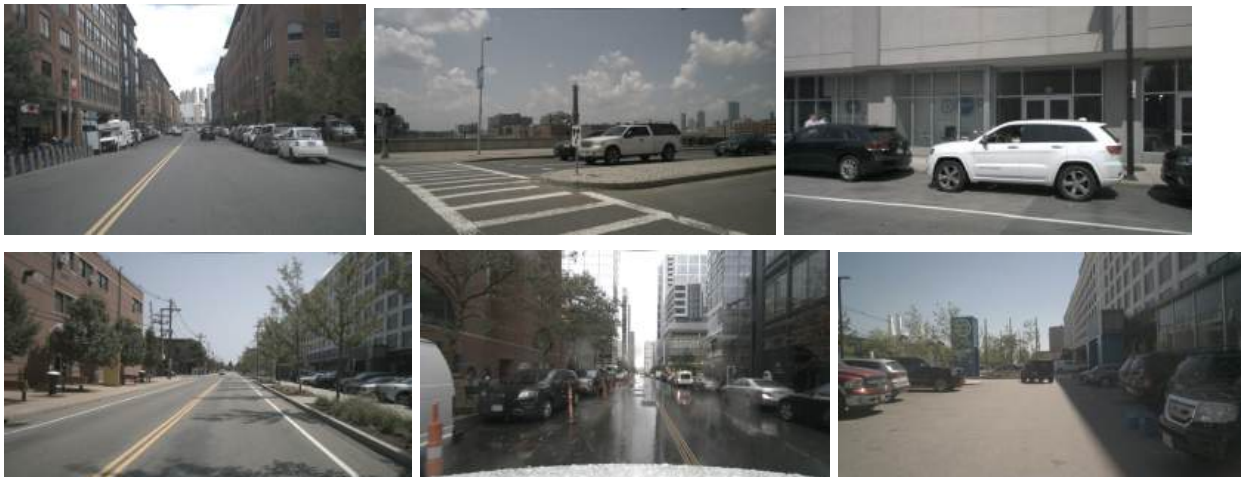


Figure 1: A set of example images from the NuScenes Data showing a diverse range of driving scenarios.

To create anomalies for the background images, we used a selection of images from three different datasets featuring various types of abnormal obstructions. The first dataset, Animal-10 (Animals-10), contains images of ten different animals, including dogs, cats, and butterflies. The second dataset, Oxford 102 Flower Dataset (Visual Geometry Group - University of Oxford), features 102 different types of flowers. The third dataset, Office-31 (Domain Adaptation - UC Berkeley, 2024), includes images of 31 different office items, such as cups, calculators, and computers.



To add these anomalies to the background images, we removed the backgrounds from the anomaly images, placed them at random locations on exactly half of the background images, and recorded the coordinates where they were added in a separate csv file. Each of the anomaly images that were overlaid on top of the NuScenes data was also randomly resized to create more variation in the generated dataset. This data alteration was done on only half of the NuScenes data (the left half of the NuScenes background images completely unaltered, giving us a modified dataset and an unmodified dataset).

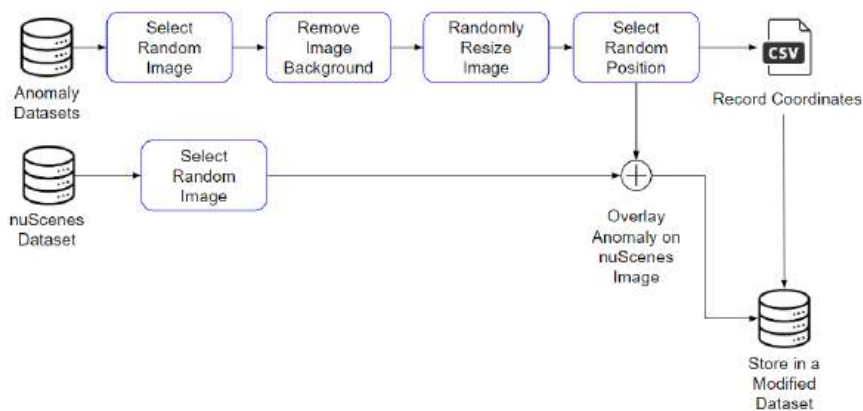


Figure 2: This diagram illustrates the process of creating a modified dataset, starting with two distinct data sets: the background NuScenes dataset and our anomaly datasets. The anomaly datasets undergo a series of transformations including background removal, resizing, and random position selection before being overlaid onto the NuScenes

dataset. The coordinates of the overlaid images are recorded in a CSV file. Only half of the NuScenes dataset is altered with anomaly data. All the modified data is then stored in a final, modified dataset.

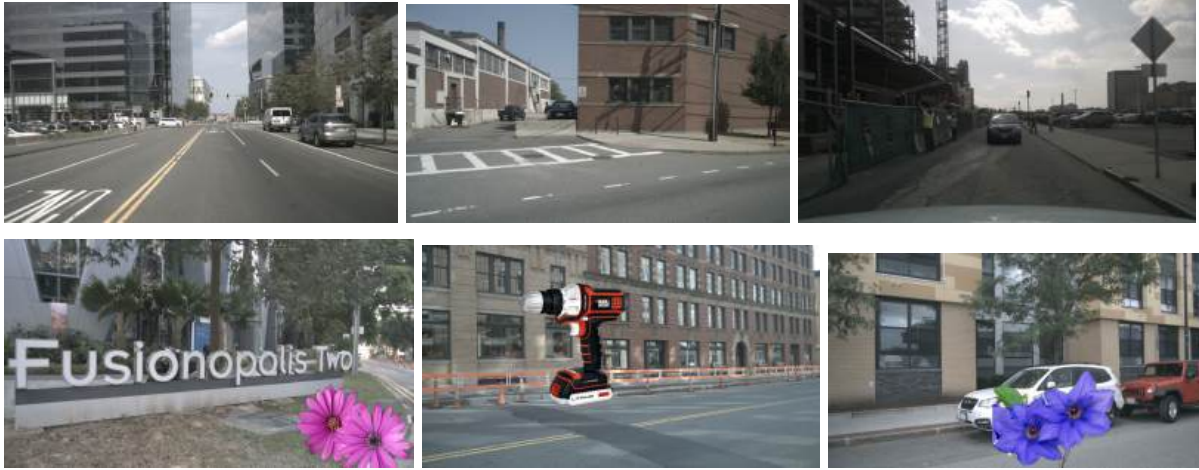


Figure 3: Sample images from the modified and unmodified datasets. On the top are samples of the unmodified images. On the bottom are samples of the modified ones with the pasted anomalies.

II.2 Anomaly Detection Model

An object detection model processes an image as input and performs a binary classification (Kumari & Srivastava, 2017) to determine whether an anomaly is present within that image. This task contrasts with regression tasks that would typically predict continuous values, such as the coordinates of anomalies in the form of bounding boxes. For this project, the focus is on classification to identify the presence or absence of anomalies, not their exact locations. The dataset preparation was integral to training this model. Two types of data were utilized to build the final modified dataset: the original unmodified NuScenes data and a version that was altered by overlaying anomalies. Each image in the dataset was labeled with a target value of 0 (unmodified) or 1 (modified), which the neural network used to train its anomaly detection capabilities.

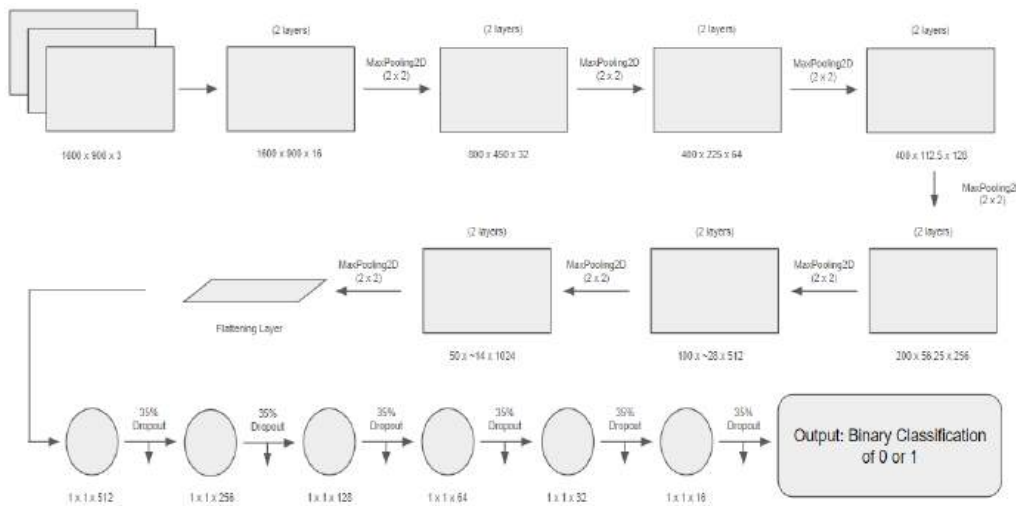


Figure 4: Diagram demonstrating the process behind the anomaly detection convolutional neural network.

This binary classification system is foundational to our method for using environmental context to detect anomalies, paving the way for more advanced applications such as precise localization in subsequent developments. The architecture of the model comprises 14 convolutional layers. These layers utilize a mathematical operation known as convolution (O'Shea, 2015), where a small matrix called a kernel slides over the image to extract essential features like edges and textures. These features are vital for the model to recognize different objects within an image.

Inserted after every two convolutional layers are pooling layers, which serve to reduce the dimensionality of the data. By doing so, pooling layers help condense the image information, maintaining only the most relevant features while making the data more manageable and reducing computational complexity. This process not only speeds up the processing but also enhances the model's ability to handle large inputs more efficiently.

As the network deepens, the number of kernels in these convolutional layers doubles every two layers, starting from 16. This increase allows the model to detect more complex features in the images, which is crucial for identifying anomalies with higher accuracy. After the convolutional and pooling layers, a flattening layer transforms the multi-dimensional array of features into a one-dimensional array that can be fed into the subsequent stages of the network.

The next section of the network consists of fully connected layers that interpret the features extracted by the convolutional layers to make a final classification decision. Between each fully connected layer, dropout layers are introduced. These dropout layers randomly ignore a subset of neurons during the training phase, which helps prevent the model from becoming overly dependent on any specific set of features, thus reducing overfitting. This technique is particularly effective in improving the model's generalization, meaning it can perform well on

unseen data. The culmination of this network is a binary output, where '0' indicates no anomaly detected and '1' indicates the presence of an anomaly.

The network uses binary cross-entropy loss (Ruby & Yendapalli, 2020) as its loss function. This choice is suitable for binary classification tasks where the model predictions are probabilities between 0 and 1 describing how likely the image is an anomaly. Binary cross-entropy loss measures the divergence between the actual class labels and the predicted probabilities, driving the model to minimize the difference for more accurate predictions. Alongside this, the metric of accuracy is used to gauge the percentage of total predictions that are correct, providing a straightforward measure of the model's performance on the training and validation datasets.

II.3 Anomaly Location Model

After successfully detecting anomalies in images, the next step is to build on this by developing a neural network that can identify and pinpoint the coordinates of anomalies within those images. Creating this neural network is challenging due to multiple computational issues. The model needed to be more complex than the binary classification model since it requires outputting more detailed information. The increase in convolutional layers and their sizes requires a very large amount of computational resources, which overwhelmed the computer system's capacity. Reducing the number of layers and making the model smaller did not capture enough complexity to accurately output a bounding box.

Additionally, we generated a much larger dataset consisting of 17,000 images. To improve the model's learning capabilities, we increased the data volume eightfold, potentially alleviating the data limitation problem. Despite these improvements, we faced additional challenges when training the model. The dataset was too large to process all at once, causing memory issues that led to system crashes. To overcome this, we developed code to process the images in batches, allowing the model to load only a portion of the dataset at a time. After implementing this solution, we retrained the model, but the results were still unsatisfactory. The loss remained high, and the accuracy of the bounding boxes was inconsistent; some images showed accurate predictions, but others had significant deviations.

Due to computational resource limitations, we shifted our focus to fine-tuning a pre-existing model. Fine-tuning is a process where a pre-trained model, already trained on a large dataset to solve a problem similar to the one at hand, is adapted to fit a specific need (Bergmann, 2024). This method is less computationally demanding because it involves adjusting the final layers of an already established model rather than training a new model from scratch. By fine-tuning, we essentially leverage the complex feature-extracting capabilities of the model and retrain it slightly with our specific dataset, which requires significantly less computational power and time.

The pre-trained convolutional neural networks used in this project include state-of-the-art architectures like ResNet50, VGG16, NASNetLarge, Xception, InceptionV3, MobileNetV2, and DenseNet201. These models are considered advanced due to their proven effectiveness and

efficiency in a variety of image recognition tasks. They have been trained on millions of images and have achieved high performance in distinguishing and classifying a vast range of image types. By fine-tuning these models, we aimed to utilize their robust, pre-trained feature detectors for our specific task of anomaly detection.

The bounding boxes are defined by four coordinates: x_1 , y_1 , x_2 , and y_2 . Here, the first two (x_1 , y_1) represent the top-left corner of the box, while the last two (x_2 , y_2) represent the bottom-right corner. This task is a regression task rather than a classification task. To evaluate the model's performance, we use mean squared error (MSE) and Huber loss, which both measure the error in the bounding box coordinates.

III. RESULTS

III.1 Anomaly Detection Model

The anomaly detection model achieved an accuracy of approximately 84-86%. This means that the model was generally effective at determining the presence of anomalies in images. However, there remains potential for enhancing its performance further.

Trial #	Test Accuracy
1	84.43%
2	84.43%
3	85.66%
4	85.83%

Table 1: Table showing the test accuracies of the anomaly detection model trained four different times.

III.2 Anomaly Location Model

The model designed to predict the locations of anomalies demonstrated notable progress. For instance, after fine-tuning the NASNetLarge model the mean squared error (MSE) was reduced from 129,796 to 81,723, indicating that learning was indeed occurring. This reduction in MSE underscores the model's enhanced ability to predict the locations of anomalies more accurately across many images. However, despite this improvement, the loss did not continue to decrease, suggesting that there are complexities within the data that the model cannot fully capture. In some instances, the predictions for certain outliers were significantly far from the actual anomalies, highlighting the model's limitations.

We employed both Huber loss and mean squared error as metrics to evaluate model performance. Huber loss (Chen, 2024) is particularly useful because it is less sensitive to outliers

in data compared to MSE. By combining MSE, which magnifies the effects of larger errors, and Huber loss, which helps maintain robustness to outliers, we gain a more comprehensive understanding of the model's accuracy and sensitivity. Examples of the resulting bounding boxes are shown below, illustrating both the successes and limitations of the model's performance.

Pre-Trained Convolutional Neural Network	Initial Huber Loss	Final Huber Loss	Initial Mean Squared Error	Final Mean Squared Error
NASNetLarge	280	210	129,796	81,723
Xception	269	200	122,154	76,808
InceptionV3	261	207	106,125	76,232
ResNet50	264	237	112,282	98,482
VGG16	283	247	134,048	95,933
DenseNet201	263	237	110,970	93,484

Table 2: Table showing the initial and final huber losses and mean squared errors of the anomaly location model where initial values are gathered from the result of the first epoch and final values are gathered from the result of the last epoch.

The difference between figures 5 and 6 illustrates the effectiveness of transitioning to a fine-tuned convolutional neural network. This model successfully identified the anomaly in the modified images, while the custom model showed that the bounding boxes (highlighted in red) were far from the actual anomaly (highlighted in green). However, there were instances where the fine-tuned model still mispositioned the bounding box, indicating room for improvement. Nonetheless, in figure 7, the bounding boxes are relatively similar in size and closer to the anomaly than those in figure 5, suggesting that the severity of the errors were significantly reduced.



Figure 5: Examples of output images of the custom-made model (before changing into fine-tuned models).

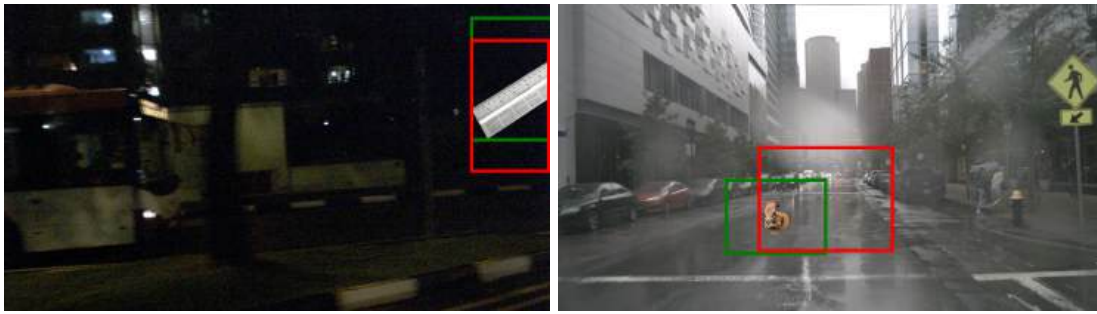


Figure 6: Examples of output images where the fine-tuned model was successfully able to determine the location of the anomalies and place bounding boxes around it.



Figure 7: Examples of output images where the fine-tuned model was unable to locate the anomalies.

IV. DISCUSSION

IV.1 Analysis of the Results

After training and evaluating two neural networks, one for anomaly detection and the other for pinpointing anomaly locations, the results demonstrated some success but also highlighted areas for improvement.

The anomaly detection model achieved an accuracy of 86%. This indicates that the model was generally effective at detecting the presence of anomalies within images. However, for practical applications in autonomous vehicles, this accuracy would need to be closer to 100% to ensure safety and reliability. While it is understood that no system is entirely infallible and occasional errors are inevitable, the goal for developers is to minimize these errors and achieve as high an accuracy as possible.

For the location model, while there were notable improvements, there is still room for enhancement. The model successfully generated bounding boxes around anomalies, and iterative adjustments led to progressively better results, as evidenced by comparisons between the fine-tuned and custom-made models. Despite these advancements, many bounding boxes were either not precisely aligned with the anomalies or were positioned at a distance from them. The immediate goal is to refine the model to reduce these discrepancies, ensuring that bounding boxes are more accurately centered around the anomalies.

We concluded that, based on our results, to truly get a model to have contextual understanding around its environments it needs to be provided more than just an image. Additionally, it needs to have a more broad understanding about the environment it lives in. To identify whether an object is an anomaly, the system needs to understand what the object is and why it might or might not belong in a particular context. Currently, the neural network lacks knowledge of the categories and types of objects present in the image, relying solely on pixel colors to detect anomalies. This is similar to a newborn baby, who wouldn't recognize anomalies in an image just by viewing 17,000 different pictures. The baby must first learn to categorize what is typically found in a street environment and understand what types of anomalies there would be, such as animals or office items.

Recently, with the rise of language models that are trained on a huge amount of the internet, we have seen these models imitate a more holistic understanding of our world. To achieve this contextual understanding, integrating a large foundation model would likely help improve results. Recent advances in vision and language models are currently being published (Elhafsi et. al., 2023; Cai et.al., 2023). A lot more data is needed to get this contextual understanding and data from different modalities such as both text and images.

Overall, this project's aim was to develop a neural network capable of detecting and locating anomalies using contextual information through a convolutional neural network. The results demonstrate that this objective is achievable with further refinement and optimization. A significant limitation was the lack of access to high-performance computing resources similar to those available to major tech companies like Google. With more advanced hardware, a larger dataset, an expanded neural network, and the implementation of a language model, it is definitely possible that the model's performance could be greatly improved, leading to more precise anomaly detection and location.

V. CONCLUSION

In summary, the anomaly detection model achieved an accuracy of 84-86%, demonstrating effectiveness in identifying anomalies in images but required further improvement for practical applications, particularly in autonomous vehicles. The location prediction model showed a reduction in mean squared error from 129,796 to 81,723, indicating progress, but still struggling with certain outlier predictions. The use of both Huber loss and mean squared error provided a balanced evaluation of performance, revealing successes and limitations in bounding box accuracy. To enhance contextual understanding, the integration of a large foundation model, alongside multimodal data, is recommended. Overall, while the project successfully developed a neural network for detecting and locating anomalies, limitations in computing resources and data availability hindered performance, suggesting that further optimization and refinement could significantly improve results.

Works Cited

- Hörl, Sebastian, Francesco Ciari, and Kay W. Axhausen. "Recent perspectives on the impact of autonomous vehicles." *Arbeitsberichte Verkehrs-und Raumplanung* 1216 (2016).
- Bansal, Prateek, and Kara M. Kockelman. "FORECASTING AMERICANS' LONG-TERM ADOPTION OF CONNECTED AND."
- "What Are the Levels of Autonomy in Self-Driving Cars? - Imagination." *Imagination*, 5 Dec. 2023, www.imaginationtech.com/future-of-automotive/when-will-autonomous-cars-be-available/what-are-the-levels-of-autonomy-in-self-driving-cars/.
- Anderson, James M., et al. *Autonomous vehicle technology: A guide for policymakers*. Rand Corporation, 2014.
- "How Motional Built Our Next-Generation Robotaxi in Less than a Year | Motional." *Motional.com*, motional.com/news/how-motional-built-our-next-generation-robotaxi-less-year. Accessed 9 May 2024
- "Learn How Waymo Drives - Waymo Help." *Support.google.com*, support.google.com/waymo/answer/9190838?hl=en.
- "Super Sensors: How Motional's AVs Can "See" Better than a Human Driver | Motional." *Motional.com*, motional.com/news/super-sensors-how-motionals-avs-can-see-better-human-driver.
- Alqahtani, Hamed, and Gulshan Kumar. "Machine learning for enhancing transportation security: A comprehensive analysis of electric and flying vehicle systems." *Engineering Applications of Artificial Intelligence* 129 (2024): 107667.
- "Waymo's Robotaxi Expansion beyond SF Temporarily Suspended as Part of Review." *ABC7 San Francisco*, 21 Feb. 2024, abc7news.com/waymo-self-driving-cars-robotaxi-expansion-suspension-san-mateo-county-los-angeles/14453501/.
- Nuscenes.org, 2021, www.nuscenes.org/nuscenes.
- "Visual Geometry Group - University of Oxford." *Www.robots.ox.ac.uk*, www.robots.ox.ac.uk/~vgg/data/flowers/102/.
- Kumari, Roshan, and Saurabh Kr Srivastava. "Machine learning: A review on binary classification." *International Journal of Computer Applications* 160.7 (2017).
- Ruby, Usha, and Vamsidhar Yendapalli. "Binary cross entropy with deep learning technique for image classification." *Int. J. Adv. Trends Comput. Sci. Eng* 9.10 (2020).
- O'Shea, K. "An introduction to convolutional neural networks." *arXiv preprint arXiv:1511.08458* (2015).
- "Animals-10." *Www.kaggle.com*, www.kaggle.com/datasets/alessiocorrado99/animals10.
- "Domain Adaptation - UC Berkeley." *Gatech.edu*, 2024, faculty.cc.gatech.edu/~judy/domainadapt/.
- Bergmann, Dave. "What Is Fine-Tuning? | IBM." *Www.ibm.com*, 15 Mar. 2024, www.ibm.com/topics/fine-tuning.
- Elhafsi, Amine, et al. "Semantic anomaly detection with large language models." *Autonomous Robots* 47.8 (2023): 1035-1055.

Cai, Mu, et al. "Making large multimodal models understand arbitrary visual prompts." arXiv preprint arXiv:2312.00784 (2023)

Chen, Bohsun. "Understanding Huber Loss Function: Insights from Applications." Medium, 11 June 2024, medium.com/@devcharlie2698619/understanding-huber-loss-function-insights-from-applications-5c1c5145d2c4.

A Greedy Algorithm for Solving the Graph Coloring Problem

By Matej Drenjancevic, Denis Ivanovic, Tomislav Rudec

Abstract.

A greedy algorithm for graph coloring is one of the most well-known algorithms for obtaining an approximate solution in solving the NP-hard coloring problem. The question is how and in which order graph vertices should be processed when coloring. It has been shown that vertices are to be arranged either ascending or descending given the number of edges that contain them.

On a graph with 20 vertices and 100 edges, the efficiency of both types of arranging vertices was tested. The results were compared to the optimal algorithm results.

Keywords NP-hard problem, vertex coloring problem, greedy algorithm

Introduction

In the theory of algorithms, one of the most important issues is the complexity of an individual algorithm. Algorithms are compared with each other and their effectiveness is evaluated by measuring the time complexity, which expresses the program execution duration for some general input and spatial complexity, which tells us how much computer memory our program will occupy during execution depending on some general input [1].

For example, an algorithm in which the user types in a natural number n and then types in n numbers, and the computer prints the smallest of those numbers is of $O(n)$ time complexity because the smallest number can be found using only one loop that performs n checks. The algorithm for multiplying two matrices, which multiplies elements by the definition of multiplying two square matrices, has a complexity of $O(n^3)$ because the program will contain three loops that move from 1 to n and are located one inside the other. A program that prints all permutations of a set of n elements has complexity $O(n!)$ because the total number of permutations of a set of n different elements is equal to $n!$ and so on.

All problems solvable in polynomial time will be put into one set called P . All those problems whose correctness of a given or offered solution cannot be checked, i.e. the set of all such problems, will be called NP .

The famous open problem is whether $P=NP$?

One of the most famous NP problems is the vertex coloring problem (simple graph), which calls for the action [2]:

What is the lowest number of colors we can use to color the vertices of the given graph so that no edge in the image contains vertices of the same color?

In other words, color each of the vertices of the given graph with one color so that the vertices that are connected by an edge are colored with different colors. The lowest number of colors is called the chromatic number of the given graph G and is denoted by $\chi(G)$.

In this problem, only simple graphs will be studied. A simple graph is a graph with no parallel edges (these are edges that connect the same two vertices) and no loops (a loop is an edge that connects a vertex to itself). Parallel edges do not change the difficulty of the vertex coloring problem because the solution we get on a graph containing parallel edges is equal to the one we get on the same graph if we delete "duplicate" edges. Also, if the graph contains a loop, we immediately know that there is no solution because the vertex is connected to itself.

The graph in Figure 1 can be colored with 4 colors as seen in Figure 2.

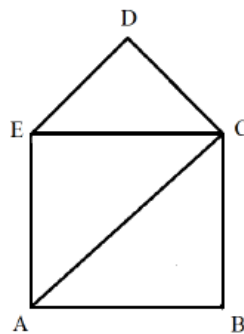


Figure 1: Given graph with 5 vertices and 7 edges.

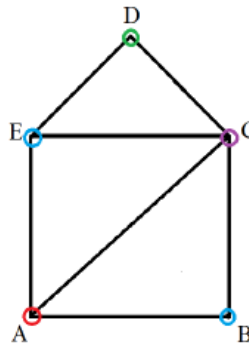


Figure 2: Given graph with 5 vertices and 7 edges – 4-color solution (D is colored green, E and B blue, A red and C purple).

However, the chromatic number of that graph is not 4, but 3 because the given graph can be colored with only 3 colors as shown in Figure 3. Also, the given graph cannot be colored with two colors (because not even its part, "triangle" DEC, cannot be colored with two colors because all vertices are connected to each other).

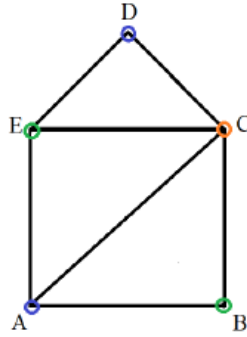


Figure 3: Given graph with 5 vertices and 7 edges – 3-color solution (A and D are colored blue, E and B green, C orange).

For easier calculation, instead of coloring the graph, i.e. assigning a color to each of the vertices, we can assign numbers to the vertices. We assign a natural number to each vertex; however, vertices that are connected by an edge cannot be assigned the same number.

If the given graph is of the form $G=(V, E)$, where V is the set of all vertices of the graph, and E is the set of all edges of the graph, the problem is to find the function $f : V \rightarrow \{1, 2, 3, \dots, k\}$ (f assigns a color to each vertex of the graph) so that $\{u, v\} \in E \rightarrow f(u) \neq f(v)$, and such that for any number smaller than k such a function no longer exists. Then, of course, k is the chromatic number of the given graph.

A greedy algorithm for solving graph vertex coloring problem

The question is whether the chromatic number of the graph can be found, i.e. how we can find the fewest colors to color the vertices of the graph so that no two vertices connected by an edge are colored the same color anywhere in the graph.

We observe the following: a graph with n vertices can certainly be colored with n colors, we just color each vertex with a different color. The chromatic number is always less than or equal to the number of vertices of the given graph.

If the graph contains a fully connected "part" (subgraph) [3] of size k (such a fully connected subgraph is called a clique), i.e. there are k vertices in the graph, each of which is connected to all other vertices of that part (of that subgraph), the chromatic number is certainly greater than or equal to k . In our first example, the graph contained a triangle, so the chromatic number had to be greater than or equal to 3.

So we can start with $n-1$ colors, where n is the number of vertices of the graph. If the graph can be colored with $n-1$ colors, the chromatic number is less than or equal to $n-1$, and if it cannot, the chromatic number is n .

How can we check if a given graph with n vertices is colorable with $n-1$ colors, or in general with k colors?

We have to create a function that will assign one color to each vertex, so that the vertices that are connected by an edge are not colored with the same color.

First, a "brute force" algorithm, i.e. an algorithm that will try all possibilities, comes to mind.

Such an algorithm that generally associates color vertices, assuming that we have n vertices in the graph that need to be associated with one of k colors, will have a complexity of $O(k^n)$. According to the Fundamental principle of counting [4], if we can color the first vertex with a total of k different colors, the second vertex with also k different colors, etc., then we must multiply these numbers in total in order to get the number of all possible combinations coloring.

If we have a graph with only, for example, 10 vertices and if we check whether it is colorable with 5 colors, we need to check $5^{10}=9765625$ different combinations.

In the loop that will do the calculations, we will choose a smart solution that, if we add, for example, color 1 to the first vertex, will prevent any vertex that is connected to that first vertex from getting that color as well.

For example, if we are trying to color the graph from Figure 1 with four colors (we know it is possible), then we do not start checking all combinations with fives (1, 1, 1, 1, 1), (1, 1, 1, 1, 2), (1, 1, 1, 1, 3), etc. but, because we see how the first and second edges are connected, all the fives will start with two different numbers.

Furthermore, as the third edge is connected to the first and second edges, the fives must actually start with three different numbers. Therefore, we start the brute force search with (1, 2, 3, 1, 1) and (due to the specificity of the graph) (1, 2, 3, 1, 2) is already the first solution.

Next, we observe the following acceleration of the "brute force" algorithm that we can use for all graphs: n -tuples of numbers must never start with a number different from 1 because if, for example, the five (2, 3, 2, 1, 4) is the solution to the coloring problem, then it is also the five (1, 3, 1, 2, 4) because we just swapped colors 1 and 2.

In total, with only these first two ideas when coloring vertices, we can start coloring with two vertices that are connected to each other by an edge and then add the number 1 to the first vertex and the number 2 to the second. The desired solution will certainly be of the form (1, 2, ...) which means that we reduced the number of all combinations from k^n to k^{n-2} .

Despite these and other ideas, mathematicians have not come up with an algorithm that would be faster than exponential, so the optimal coloring of vertices for "larger" graphs is still an unsolvable problem.

In practical examples, for graphs with, for example, several hundred or several thousand vertices, we often apply a greedy algorithm, a fast heuristic algorithm which, like other heuristic algorithms, does not always provide an optimal solution, but is fast and "often" provides a solution "close to" optimal [5].

The greedy algorithm for solving the graph vertex coloring problem works as follows:

1. Arrange the vertices of the graph in an arbitrary sequence.
2. Color the first vertex in the row with color 1.
3. If the second vertex is connected to the first, color it with color 2, and if not, with color 1.
4. Coloring other vertices: we color the vertex with the color of the smallest number for which it applies that this vertex is not already connected to another vertex of the same color.
5. We return to step 4 until all vertices are colored.

For the graph in Figure 1, coloring with the greedy algorithm would go like this:

1. Order: A, B, C, D, E.
2. A is colored with color 1.
3. B is colored with color 2 (because it is connected to A).
4. C is colored with color 3 (because the vertices to which C is connected are colored with colors 1 and 2, so 3 is the next free color).
5. D is colored with color 1 (because that color is free and it is the smallest number with which no vertex connected to D is colored).
6. E is colored with color 2 (because the vertices connected to E are colored with color 1).

We managed to color the given graph with three colors. In total, since this algorithm is not optimal, and yet it colored the vertices of the graph in an acceptable way, we conclude that the chromatic number of the given graph is less than or equal to 3.

The 4th step of the aforementioned algorithm is the following - we color the observed vertex with a number that is not in the set of vertex numbers connected to the observed vertex.

Here it is clear that in the first step, we can arrange the vertices in any way we want out of a total of $n!$ ways (in our example, we can color 5 vertices in a total of $5!=120$ ways).

Let us see how to color the vertices for a different initial arrangement of the vertices:

1. Order: B, D, A, C, E.
2. B is colored with color 1.
3. D is colored with color 1.
4. A is colored with color 2.
5. C is colored with color 3.
6. E is colored with color 4.

In total, since this algorithm is not optimal, and yet it colored the vertices of the graph in an acceptable way, we conclude that the chromatic number of the given graph is less than or equal to 4.

In general, a greedy algorithm will never use more than $s+1$ colors, where s is the largest vertex degree that appears in the given graph. The degree of a vertex is equal to the number of edges that "come out" from it, i.e. it is equal to the number of vertices connected to that vertex [6].

For example, in Figure 1, the degrees of vertices A, B, C, D, E are 3, 2, 4, 2, 3, respectively.

Since 4 is the highest degree of the vertex that appears in that graph, the greedy algorithm will be able to color it with $4+1=5$ colors. (The proof of this statement is evident in the figure: if we color a vertex, it can be connected to the most s vertices that can be colored with the most s different colors, so we will then color that vertex with the $s+1$ color.)

Ascending and descending ordering of vertices in the application of the greedy algorithm for a simple graph vertex coloring

There are simple and effective methods of coloring in which the edges are initially arranged in either the ascending or descending order according to their degree.

The coloring of vertices with descending ordering gives (on average) better results than the coloring of vertices with ascending ordering.

Let us see the example of a graph where coloring with a greedy algorithm will use two different colors due to different sets of vertices.

For the graph in Figure 4, let us start by arranging the vertices in the following way: 7, 8, 2, 1, 3, 4, 5, 6. We arranged them in the descending order by the degree of the vertex.

The solution can be found in Figure 5 where the graph is colored with two colors.

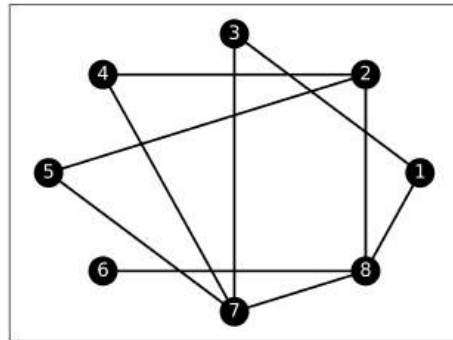


Figure 4: The given graph whose vertices are to be colored with different colors

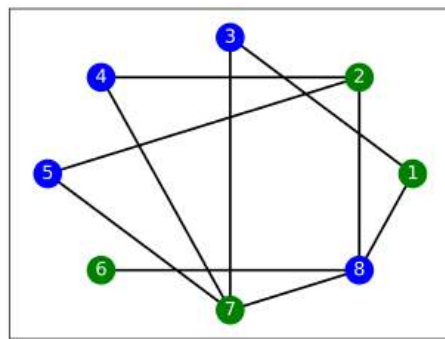


Figure 5: Two-color solution

If we arrange the vertices in the ascending order, i.e. in the following way: 6, 1, 3, 4, 5, 2, 7, 8, we need four colors as shown in Figure 6.

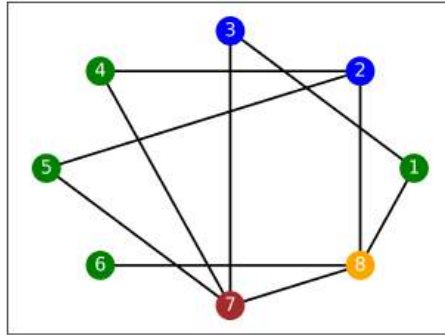


Figure 6: Four-color solution

In order to be able to evaluate the effect of the greedy algorithm on the ascending and the descending set of vertices, let us compare these algorithms on specific cases where a graph consisting of 20 vertices and 100 edges is given in advance, and the edges are distributed in a hundred different instances of the problem randomly. At the same time, let us calculate the results of the optimal algorithm on the same given graph. The results are listed in Table 1.

Table 1: The number of colors needed to color the vertices of a simple graph - comparison of the results of two versions of the greedy algorithm and the optimal coloring of the vertices (with the average value in the last row)

Optimal algorithm	Descending order	Ascending order	Frequency of occurrence
5	5	7	1
5	6	7	4
5	7	7	2
6	6	6	4
6	6	7	14
6	6	8	17
6	6	9	5
6	6	10	1
6	7	6	1
6	7	7	9
6	7	8	30
6	7	9	6
6	8	8	3
7	7	8	2
7	8	8	1
5.99	6.7	7.93	

Conclusion

The graph vertex coloring problem is one of the most famous problems in the graph theory. There is no polynomial optimal algorithm for it.

Instead of the optimal algorithm, we often use various heuristic methods, among which the most famous one is the greedy algorithm.

The results of the greedy algorithm differ from the results of the optimal algorithm as can be seen from the measured data (the averages are 6, 7 and 8, rounded to a whole number in the table above).

In other words, the specified greedy algorithms differ in this case by one color on average. The optimal algorithm is one color better than the greedy algorithm with the descending number of vertices, and the algorithm with the descending number of vertices is one color better than the greedy algorithm with the ascending number of vertices.

However, Table 1 lists one case (out of 100 cases) when the greedy algorithm with the ascending number of vertices is one color better than the greedy algorithm with the descending number of vertices (row 9 – 6 7 6).

It is very important to arrange vertices in the initial step of the greedy algorithm. The results show that this order is more important than the selection of either the optimal or greedy algorithm.

Works Cited

- [1] Sipser, M. (2006). Introduction to the Theory of Computation (2nd ed.), USA: Thomson Course Technology, ISBN 0-534-95097-3A
- [2] Garey, M. R.; Johnson, D. S. (1979). Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman, ISBN 0-7167-1045-5
- [3] Gross, J.L.; Yellen, J. (2004). Handbook of graph theory. CRC Press. p. 335. ISBN 978-1-58488-090-5.
- [4] Rosen, K. H., ed. Handbook of discrete and combinatorial mathematics. CRC press, 1999.
- [5] Kučera, L. (1991). "The greedy coloring is a bad probabilistic algorithm", Journal of Algorithms, 12 (4): 674–684, doi:10.1016/0196-6774(91)90040-6, MR 1130323.
- [6] Brooks, R. L. (1941), "On colouring the nodes of a network", Mathematical Proceedings of the Cambridge Philosophical Society, 37 (2): 194–197, Bibcode: 1941PCPS...37..194B, doi:10.1017/S030500410002168X, S2CID 209835194.

Optimizing Electric Vehicle Charging Station Placement in New York City: A Spatiotemporal Analysis of Traffic Patterns and Range Scenarios

By Aaron Yin

Abstract

In this paper, we aim to develop an optimized spatiotemporal model for the placement of electric vehicle (EV) charging stations in New York City, factoring in high-traffic zones, commuting patterns, and different EV range scenarios. With the increasing adoption of EVs, strategically locating charging stations is crucial for minimizing range anxiety and supporting urban sustainability efforts. Our approach leverages high-resolution spatial data on daily miles traveled and traffic volume during morning and afternoon rush hours to identify optimal locations for EV infrastructure. The methodology involves analyzing traffic data across three EV range scenarios: 100 miles, 250 miles, and 400 miles on a full charge. Using a metric inspired by Ji et al. (2017), we calculated the mileage intensity (MI) for each grid cell in NYC and combined this with the EV range to estimate charging demand (CD). The results were normalized to create a charging demand index (CDN), allowing for a direct comparison between locations across the city. Multiple choropleth maps were produced to visualize car density, rush hour traffic patterns, and charging demand, helping to highlight spatial disparities in infrastructure needs. The results indicate that high traffic areas, such as Manhattan and parts of Brooklyn, consistently exhibit elevated demand for EV charging stations across all range scenarios. Staten Island and Queens show more sporadic demand, correlating with lower traffic volumes and less EV adoption potential. However, limitations in the dataset, particularly in terms of temporal resolution and the lack of detailed socioeconomic data, may affect the precision of station placement. Future studies should consider integrating renewable energy sources and socioeconomic factors to refine the model further and ensure equitable access to charging infrastructure across diverse urban populations.

Keywords Electric vehicles, charging stations, New York City, traffic, spatiotemporal analysis, charging demand

Introduction

The transportation sector is one of the largest contributors to greenhouse gas emissions globally, responsible for 29% of total emissions and over 50% of nitrogen oxide (NOx) pollution (Nyugen et al., 2024). As the effects of climate change worsen, EVs have emerged as a key solution for reducing the carbon footprint of the transportation industry. EVs offer substantial environmental benefits compared to traditional internal combustion engine (ICE) vehicles, with the potential to lower emissions by up to 90% when considering the vehicle's entire lifecycle (GNANAVENDAN et al., 2024). However, despite these advantages, widespread EV adoption faces several barriers—particularly the availability of charging infrastructure. In cities like New York, public EV charging stations are vital for supporting EV ownership, especially given the limited availability of private home charging options. Common locations for EV charging

include city-managed parking garages, airports, and commercial garages, which are critical to reducing range anxiety—the fear that a vehicle will run out of power before reaching a charging station (Ji et al., 2017). Without accessible and reliable charging infrastructure, drivers may be discouraged from switching to EVs due to this range anxiety. Currently, the charging network in New York City is inadequate to meet rising demand, with long charging times and limited availability during peak usage periods (Moghaddam, 2018). This lack of infrastructure not only hinders EV adoption but also risks exacerbating urban congestion as drivers detour to find available stations.

Beyond the practical challenges of charging access, there are significant social equity concerns. Studies show that EV charging infrastructure is often concentrated in wealthier, predominantly white neighborhoods, leaving low-income and marginalized communities underserved (Khan et al., 2021). This inequitable distribution of charging stations means that not all New Yorkers have equal access to the environmental and economic benefits of EVs. Addressing this disparity is critical for ensuring that the city's transition to electric transportation does not perpetuate existing social and environmental injustices.

This study aims to address these challenges by developing a spatiotemporal model for optimizing EV charging station placement in New York City. By integrating traffic patterns, commuting behavior, and existing infrastructure, the model seeks to identify high-demand areas that would benefit most from additional charging stations. The focus is not only on reducing range anxiety but also on ensuring that charging infrastructure is equitably distributed, particularly in underserved neighborhoods. We hypothesize that strategically locating charging stations in areas of high traffic and in communities currently lacking access will increase EV adoption and reduce range anxiety. Furthermore, by improving the availability of charging infrastructure in marginalized areas, this model can help make EVs a more viable option for all New Yorkers, contributing to the city's broader sustainability and social equity goals.

Methods

To estimate the optimal locations for electric vehicle (EV) charging stations, we utilized traffic and vehicle data from multiple sources. The NYC OpenData Traffic Volume Counts dataset provided detailed traffic counts across New York City, divided by borough and time of day (NYC OpenData, 2024). This dataset includes data on daily vehicle volumes during both morning and afternoon rush hours, collected on a recurring basis throughout each year. The traffic volume data spans from 2014 to the present, offering a longitudinal view of traffic patterns that are essential for understanding trends and changes over time. Additionally, annual vehicle miles traveled (VMT) data was collected from the NYC Department of Health and Mental Hygiene's data explorer, covering VMT estimates on a yearly basis, with updates based on new reports (New York City Department of Health and Mental Hygiene, 2024). The VMT data spans multiple years, providing comprehensive coverage of vehicle usage trends across New York City.

The datasets were chosen specifically for their relevance to understanding traffic density and vehicle flow patterns, both critical factors in determining the optimal placement of EV charging stations. By using detailed daily traffic volume data, we can identify high-traffic zones and rush hour patterns, which are likely to influence charging demand. The annual VMT data complements this by offering insights into overall vehicle activity across the city, which helps ensure that charging stations are located in areas that reflect long-term transportation trends rather than short-term fluctuations.

New York City Car Density

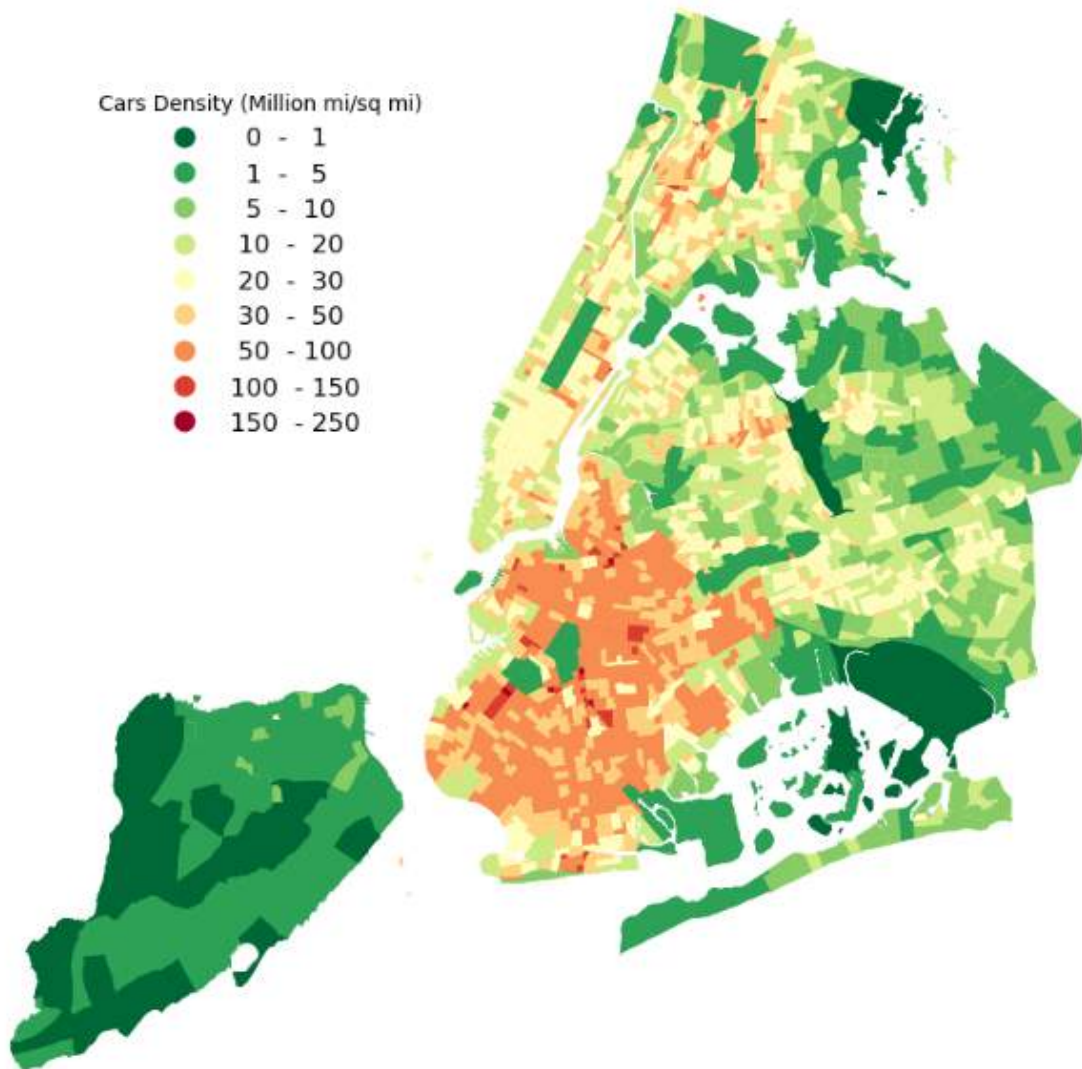


Fig. 1. Car Density Distribution in New York City. I used the NYC Department of Health and Mental Hygiene’s Walking, Driving, and Cycling dataset to analyze and map vehicle density across New York City, as shown in the car density map. This dataset provided comprehensive traffic data, allowing for an accurate identification of areas with high vehicle density, which is crucial for optimizing the placement of EV charging stations in regions with higher traffic volumes.

One way to analyze traffic patterns is to understand the density of vehicles traveled. To calculate traffic density across different geographical regions, we utilized the following approach. Let “T” represent the total vehicle miles traveled per square mile in millions, and “A” represent the area of the geographical region in square miles. The traffic density “D” for each region was calculated using the following formula:

$$D = (T * 10^6)/A \quad (1)$$

In this case, density is based on the total vehicle miles traveled, allowing us to capture the intensity of traffic flow through each region. This calculation takes into account data from multiple years to provide a comprehensive view of vehicle activity patterns across the city. This method ensures that we can accurately estimate car density in relation to both the total miles traveled and the area of the region, forming the foundation for optimizing EV charging station placement. Fig. 1 illustrates the distribution of car density across New York City. This map was critical in understanding how vehicle density varies spatially across the boroughs and helped form the basis for the subsequent analysis of charging demand.

The raw traffic volume data obtained from NYC OpenData underwent a data cleaning process to ensure that only complete and accurate records were used in the analysis. This dataset contained daily traffic volume counts across various locations in New York City. The cleaning process focused specifically on removing null points, as these represented missing or incomplete data entries. No interpolation or replacement of missing values was conducted, and no outliers were removed to maintain the natural variations in traffic patterns. The exclusion of null points ensured the integrity of the dataset without artificially altering the data. After cleaning, the dataset was queried into two key time periods: morning rush hour (6 AM – 10 AM) and afternoon rush hour (4 PM – 8 PM). This query was essential to capture the distinct traffic dynamics at different times of the day. Morning rush hour data reflects the movement of commuters into central areas such as Manhattan, while afternoon rush hour data captures the reverse flow of commuters leaving the city. These two time periods were analyzed separately to identify areas with varying traffic intensities throughout the day, which could influence EV charging demand.

New York City Car Volume During Morning Rush Hour

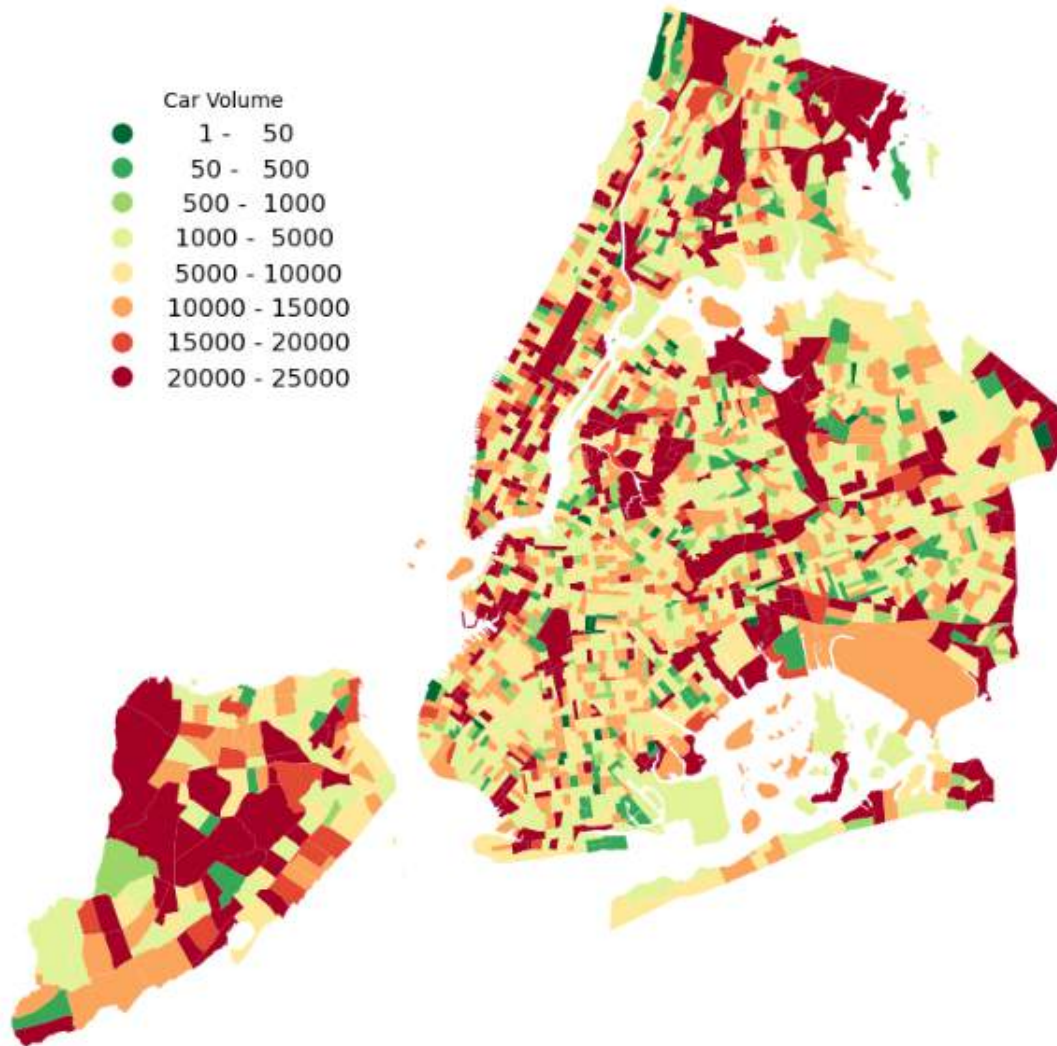


Fig. 2. Morning Rush Hour Traffic Volume in New York City. I used the NYC OpenData Traffic Volume Counts dataset to generate the map of car volumes during the morning rush hour in New York City, as shown in Fig. 2. This dataset allowed for a detailed analysis of traffic flow during peak commuting hours, providing crucial insights into regions with higher car volumes where EV charging infrastructure may be most necessary.

New York City Car Volume During Afternoon Rush Hour

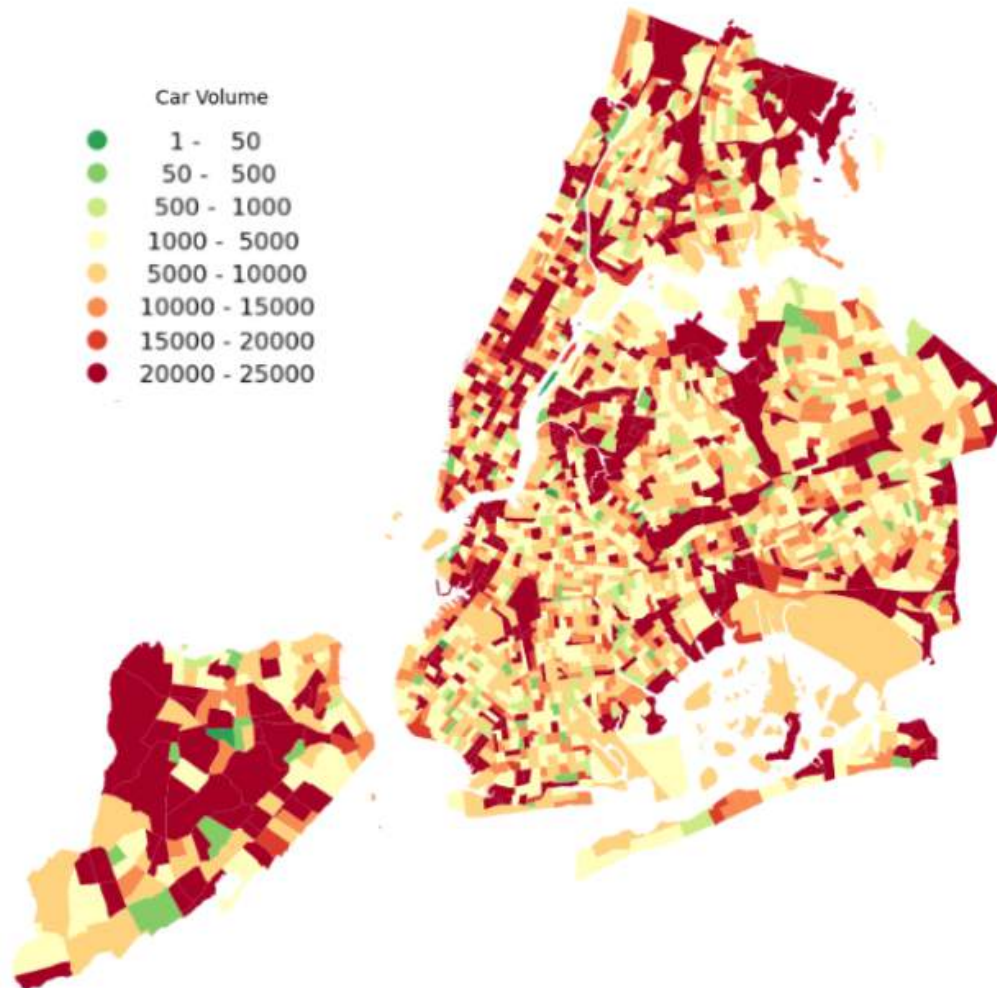


Fig. 3. Afternoon Rush Hour Traffic Volume in New York City. I used the NYC OpenData Traffic Volume Counts dataset to generate the map of car volumes during the afternoon rush hour in New York City, as shown in Fig. 3.

This dataset enabled the visualization of traffic patterns during the evening commute, highlighting areas with increased traffic flow, which is essential for determining optimal EV charging station locations based on peak vehicle usage.

Fig. 2 and Fig. 3 visualize the distribution of car volumes during morning and afternoon rush hours, respectively. These maps highlight spatial patterns in traffic that are crucial for understanding potential hotspots for EV charging infrastructure. For example, Manhattan shows high traffic volumes during the morning rush hour, suggesting a greater need for EV charging stations to accommodate the influx of vehicles.

The next step in data processing involved aggregating traffic volume data at the borough level. We calculated the total vehicle volume for each borough during both the morning and afternoon rush hour periods. Aggregating data at this level enabled us to compare boroughs and

identify regions with the highest concentrations of vehicles, giving us a clear view of broader macro-trends.

However, borough-level analysis provides only a high-level understanding of traffic patterns. By using more detailed shapefiles that break down data into smaller, more specific geographic regions, we can uncover finer-grained spatial trends. This detailed approach allows for a more accurate identification of specific areas within boroughs where traffic is heaviest, which is crucial for optimizing the exact locations for EV charging stations. The combination of borough-level aggregation and detailed shapefile analysis provides a comprehensive perspective, balancing macro-level insights with localized precision.

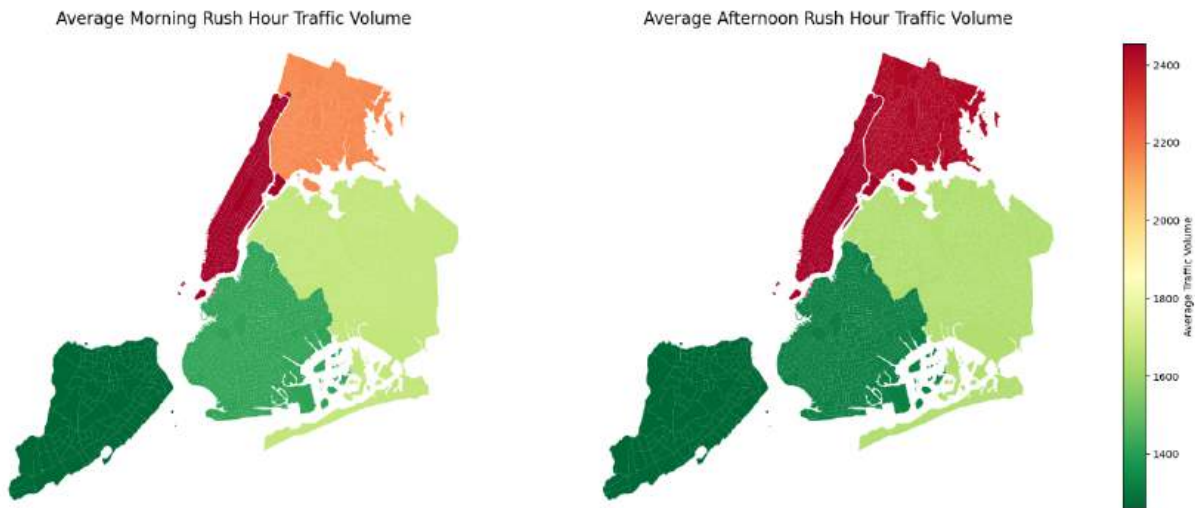


Fig. 4. Daily Average Morning and Afternoon Traffic Volume in NYC During Rush Hour by Borough. I used the NYC OpenData Traffic Volume Counts dataset to generate the maps of daily average morning and afternoon traffic volumes by borough, as shown in Fig. 4. These maps provide a comparative view of traffic intensity across different boroughs during rush hours, identifying Manhattan and Brooklyn as key areas with consistently high traffic volumes, which are crucial for planning the deployment of EV charging infrastructure.

To further refine the analysis, average daily traffic volumes for each borough were calculated. This involved summing the total traffic volume across the entire dataset for each borough and dividing it by the number of valid days with complete data. This step provided a long-term perspective on traffic patterns and allowed us to normalize fluctuations in daily traffic. The results of this calculation are shown in Fig. 4, which illustrate average morning and afternoon traffic volumes for each borough. Manhattan and Brooklyn consistently show the highest volumes, making them key candidates for EV charging infrastructure.

Additionally, traffic data was aggregated spatially using a tract-level shapefile for New York City, allowing for the identification of traffic patterns across different boroughs. This method allowed for the identification of localized traffic hotspots, such as highways and major transit corridors, which experience higher volumes of vehicles and therefore, higher potential demand for EV charging stations. Through the process of removing null points, segmenting data into rush hour periods, aggregating traffic volumes by borough, and creating a high-resolution

grid for localized analysis, we developed a dataset that provided a detailed and accurate representation of New York City’s traffic patterns. This dataset served as the foundation for calculating EV charging demand, enabling a precise estimate of where charging stations would be needed the most.

Range Scenarios

Scenarios	Mileage range (# miles)
1	100
2	250
3	400

Table 1. These are the three range scenarios. The 150-mile and 400-mile ranges represent short- and long-range EVs, while the 250-mile range was calculated as the midpoint between 100 and 400 miles to capture a realistic mid-range scenario (U.S. Department of Energy, 2021).

To determine the electric vehicle (EV) ranges used in our spatiotemporal model for calculating EV charging demand across New York City, we referenced data from the U.S. Department of Energy's Alternative Fuels Data Center (AFDC). In the Electric Drive Vehicles report, the AFDC outlines the typical range of EVs based on various vehicle models and battery capacities. This report provided a comprehensive overview of the three key range scenarios—150 miles, 250 miles, and 400 miles—representing short-, mid-, and long-range EVs, respectively. We derived the 250-mile value by calculating the middle number between the lower and upper bounds of EV ranges, which are 100 miles and 400 miles. This approach allowed us to capture a realistic average for mid-range EVs that can handle both city driving and longer regional trips. Using this midpoint ensures that our model considers a broad spectrum of EV users, optimizing the placement of charging stations for vehicles with varying driving ranges.

To calculate the demand for EV charging stations across different regions of New York City, we employed a spatiotemporal model based on the methodology outlined by Ji et al. (2017). The process began by estimating the daily miles traveled per car (M) for each spatial grid cell. This was done by dividing the annual million miles traveled in each grid cell by the number of data points (N) in that cell and then normalizing by 365.25 days to represent the average number of days in a year. The use of 365.25 accounts for leap years, ensuring accurate daily estimates. This method adjusts for data variation across regions and helps identify areas with high daily travel, crucial for determining EV charging demand:

$$M = \text{Annual million miles per cell} / (N * 365.25) \quad (1)$$

Next, we calculated the mileage intensity (MI) for each grid cell. The mileage intensity is a measure of how much vehicle traffic occurs in each cell relative to the average traffic across all

cells. This was calculated by dividing the daily miles traveled in a cell (M) by the mean of daily miles traveled across all cells:

$$MI = M / \text{mean}(M) \quad (2)$$

This intensity factor (MI) was then multiplied by each EV range scenario (S)—100, 250, and 400 miles—to calculate the charging demand (CD) for each grid cell:

$$CD = MI * S \quad (3)$$

To ensure comparability across all grid cells, the charging demand values were normalized (CDN) using the following normalization formula:

$$CDN = (CD_i - \min(CD)) / (\max(CD) - \min(CD)) \quad (4)$$

The result was a spatially resolved map displaying the normalized charging demand (CDN) for each of the EV range scenarios, offering insight into where charging infrastructure should be prioritized based on vehicle traffic patterns across New York City.



Fig. 5. Spatial Distribution of EV Charging Demand in NYC for 100, 250, and 400 Mile Range Vehicles. I used the NYC OpenData Traffic Volume Counts dataset to create the maps of EV charging demand for 100, 250, and 400-mile range vehicles, as shown in Fig. 5. These maps illustrate the spatial distribution of charging demand across New York City, highlighting how different vehicle ranges influence the areas where charging infrastructure will be most critical, with longer-range vehicles requiring fewer but more strategic station placements in high-demand regions.

Fig. 5 presents the spatial distribution of EV charging demand for the three scenarios—100, 250, and 400 miles—demonstrating how charging demand increases with EV range. These maps highlight areas with the highest anticipated demand for charging infrastructure across New York City.

Results

The spatial distribution of car density across the five boroughs of NYC is shown in Fig. 1. This map provides a foundational understanding of vehicle distribution, which directly influences traffic patterns and subsequent charging demand. The highest density of vehicles is concentrated in Manhattan and Brooklyn, with specific regions in Queens also exhibiting elevated vehicle densities. In contrast, Staten Island and the outer parts of the Bronx demonstrate significantly lower vehicle densities, suggesting lower traffic volumes and potential EV charging demand in those areas. The higher concentration of vehicles in Manhattan can be attributed to its role as a commercial hub, while Brooklyn shows a mix of both residential and commercial traffic activity.

To better understand the temporal distribution of traffic across the city, we analyzed both morning and afternoon rush hour traffic volumes. Fig. 2 visualizes the morning rush hour traffic volume, highlighting significant congestion in Manhattan, parts of Brooklyn, and certain corridors in Queens. The higher traffic volumes in Manhattan during the morning hours align with the centrality of Manhattan's business districts, where commuters from outer boroughs travel to work. Similarly, Fig. 3 shows the afternoon rush hour traffic volume, with patterns indicating a redistribution of traffic as workers leave Manhattan and return to residential boroughs, primarily Brooklyn and Queens. The Bronx and Staten Island show consistently lower traffic volumes during both peak periods, suggesting a relatively low demand for EV infrastructure in these regions based solely on traffic volume metrics. Further summarizing these observations, Fig. 4 represents the average morning and afternoon traffic volumes by borough, respectively. Manhattan exhibits the highest overall average traffic volume, surpassing 2,000 vehicles in both morning and afternoon periods, followed by Brooklyn and Queens. These figures reinforce the notion that Manhattan and Brooklyn should be prioritized when developing EV infrastructure due to the consistent high traffic volumes throughout the day. Conversely, Staten Island's consistently low traffic volume of fewer than 1,400 vehicles indicates that EV infrastructure investment may be less critical in this borough based solely on current traffic volume data.

The charging demand distribution was calculated for three EV range scenarios: 100 miles, 250 miles, and 400 miles. Fig. 5 illustrates the spatial distribution of EV charging demand for these three ranges across NYC. The maps reveal that charging demand is consistently highest in Manhattan and the downtown areas of Brooklyn, irrespective of the vehicle range. This pattern corresponds with the high traffic density and volume seen in the earlier figures. Notably, in all range scenarios, Staten Island and large sections of the Bronx exhibit significantly lower charging demand, further reinforcing the findings from the traffic volume analysis. As the vehicle range increases from 100 miles to 400 miles, the spatial extent of high charging demand slightly expands. However, the central business district in Manhattan remains the focal point of demand in all scenarios. This suggests that longer-range EVs, while slightly reducing the frequency of charging needs, do not significantly alter the spatial distribution of high-demand areas, as Manhattan's traffic concentration remains constant across all vehicle ranges.

Recognizing that proximity to major transportation hubs would influence charging demand, we conducted a focused analysis of areas surrounding LaGuardia and JFK airports. Traffic volumes in these areas are consistently high due to both passenger vehicles and commercial traffic, indicating a potentially higher need for EV charging infrastructure. We integrated airport locations into the spatial model to refine our understanding of charging demand in these high-traffic zones. LaGuardia Airport (LGA) is located at approximately 40.7769° N latitude and 73.8740° W longitude, while John F. Kennedy International Airport (JFK) is situated at approximately 40.6413° N latitude and 73.7781° W longitude. These coordinates were incorporated into the spatial analysis to help determine where EV charging stations may be most necessary in the airport vicinity.

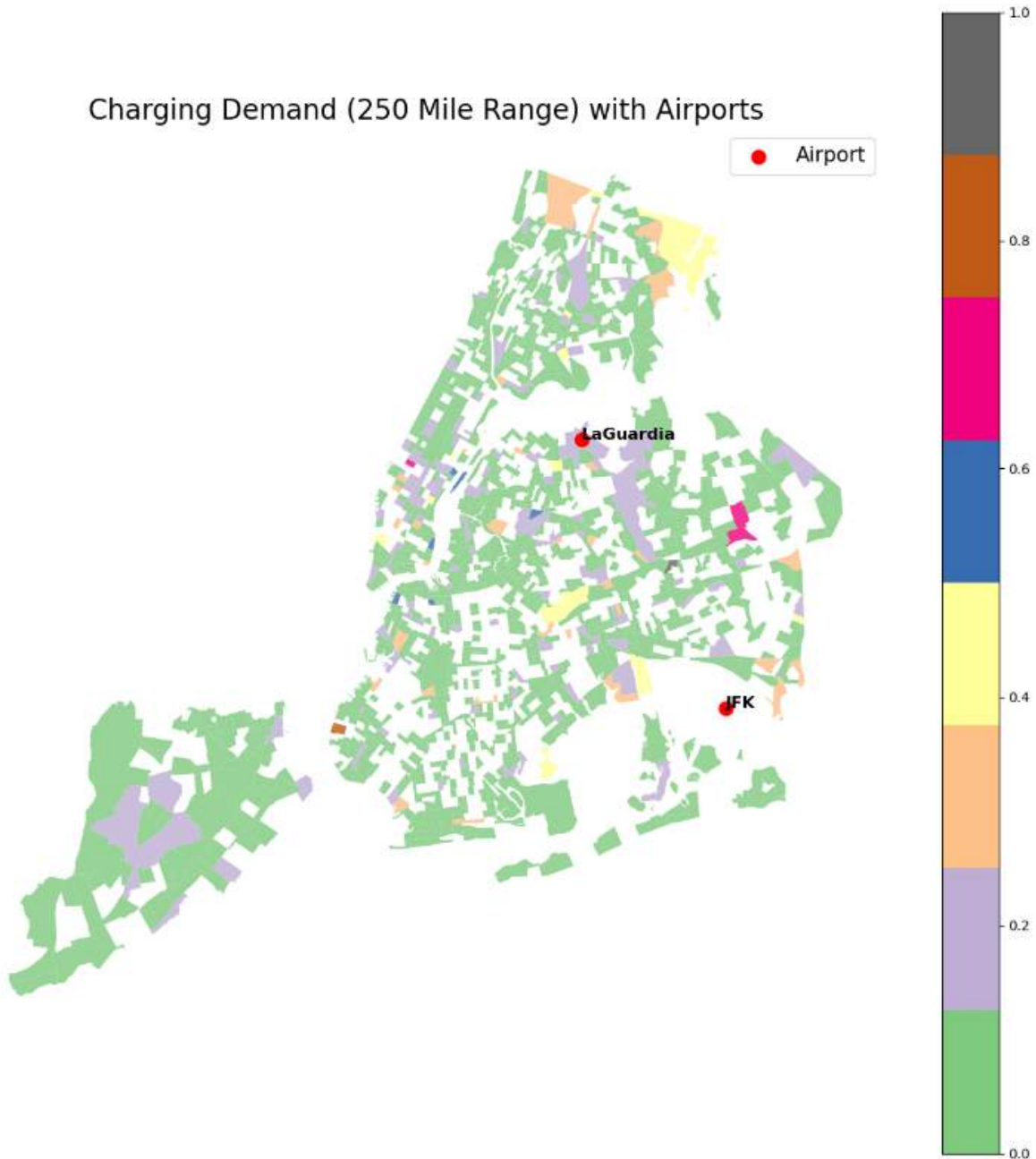


Fig. 6. Spatial Representation of Charging Demand (250 Mile Range) with LaGuardia and JFK Airports in NYC. I used the NYC OpenData Traffic Volume Counts dataset to create the spatial representation of EV charging demand for 250-mile range vehicles, with airports overlaid, as shown in Fig. 6. This map highlights key locations like LaGuardia and JFK airports, where higher demand for charging infrastructure may arise due to the high volume of vehicles and the strategic importance of these transit hubs for long-range EVs.

Fig. 6 shows the charging demand for the 250-mile range scenario, overlaid with the locations of JFK and LaGuardia airports. This figure highlights how major transportation hubs increase charging demand and supports the recommendation for installing additional EV infrastructure near airports.

In addition to the general charging demand maps, Fig. 6 specifically highlights charging demand for EVs with a 250-mile range and incorporates the locations of LaGuardia and JFK airports. This visualization underscores the potential need for charging infrastructure near major transportation hubs, where both passengers and workers may require EV charging services. The proximity of high-demand areas to these airports suggests that integrating EV charging stations near LaGuardia and JFK could serve a dual purpose, accommodating both daily commuters and long-distance travelers.

All traffic volume and charging demand data were analyzed statistically to identify correlations between high-traffic zones and projected EV charging demand. We performed spatial analyses to map traffic volume and EV charging demand across New York City and employed temporal segmentation to account for daily traffic patterns. Historical trends in traffic volume were visualized using bar charts representing rush-hour traffic volumes across boroughs from 2000 to 2023.

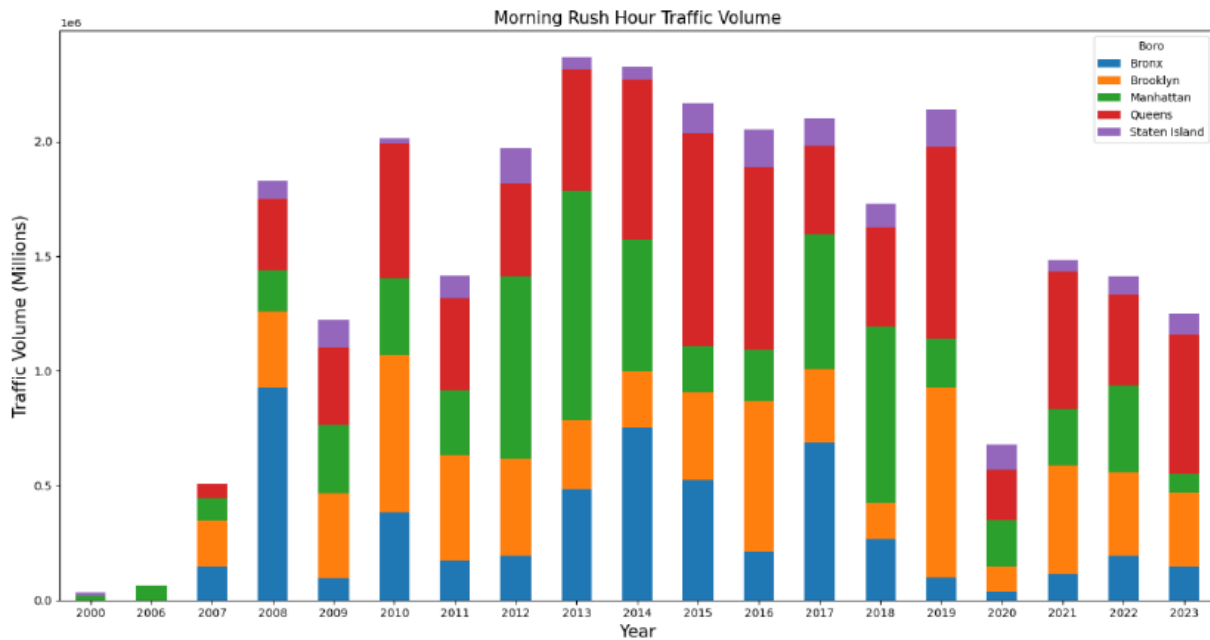


Fig. 7. Morning Rush Hour Traffic Volume by NYC Borough (2000-2023). I used the NYC OpenData Traffic Volume Counts dataset to create the bar chart illustrating morning rush hour traffic volume by borough from 2000 to 2023, as shown in Fig. 7. This chart reveals temporal changes in traffic patterns across New York City's boroughs, with noticeable peaks and declines that provide insight into how traffic volume has evolved over the years, which can inform future EV charging infrastructure planning.

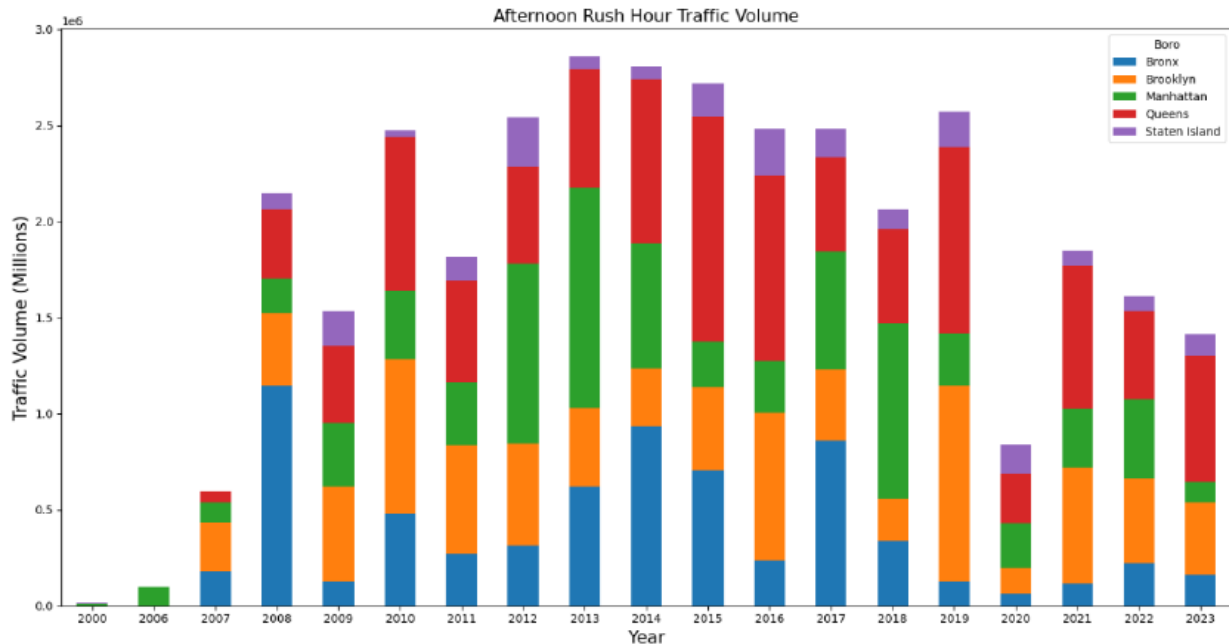


Fig. 8. Afternoon Rush Hour Traffic Volume by NYC Borough (2000-2023). I used the NYC OpenData Traffic Volume Counts dataset to create the bar chart showing afternoon rush hour traffic volume by borough from 2000 to 2023, as depicted in Fig. 8. This chart highlights trends in afternoon traffic across the boroughs over time, reflecting how traffic volumes fluctuate and indicating critical periods for assessing the demand for EV charging stations based on commuting patterns.

Fig. 7 and Fig. 8 show the historical morning and afternoon rush-hour traffic volumes, respectively, broken down by borough from 2000 to 2023. These figures support the long-term view of where EV charging demand is likely to increase in the future based on traffic growth trends. To provide a temporal perspective on traffic patterns and their implications for charging demand, we examined traffic volume trends from 2000 to 2023. Fig. 7 presents the morning rush hour traffic volume for each borough over this 23-year period. Notably, Manhattan consistently exhibits the highest traffic volumes, peaking around 2014. The decline in traffic volume in 2020 and 2021, likely due to the COVID-19 pandemic and associated lockdowns, is evident across all boroughs, though Manhattan's drop is particularly pronounced (Rajput, 2022). Brooklyn and Queens show a gradual increase in traffic volume from 2000 to 2019, indicating growing residential and commercial activity. Staten Island, by contrast, shows minimal fluctuation over the years, further reinforcing its relatively low demand for EV infrastructure. Similarly, Fig. 8 displays the afternoon rush hour traffic volume for the same time period. The traffic patterns during the afternoon rush hour mirror those of the morning, with Manhattan showing the highest volume, followed by Brooklyn and Queens. The year 2014 marks the peak traffic volume for most boroughs, after which a gradual decline is observed. Post-2020, the recovery of traffic volume is slow but steady, particularly in Brooklyn and Queens, where residential and commercial development continues to fuel traffic growth. The data suggests that any future

infrastructure development plans, including the placement of EV charging stations, should take into account the expected return of pre-pandemic traffic volumes in these high-density areas.

Overall, the results indicate that Manhattan and Brooklyn are the primary areas of focus for EV charging infrastructure development, given their high traffic volumes, vehicle density, and consistent demand across all EV range scenarios. These two boroughs exhibit not only the highest current demand but also show resilience in maintaining high traffic volumes even after the disruptions caused by the COVID-19 pandemic. Queens also demonstrates moderate demand for EV infrastructure, particularly in areas with high traffic corridors. In contrast, Staten Island and parts of the Bronx demonstrate significantly lower traffic volumes and EV charging demand, indicating that these areas may not require as urgent investment in EV infrastructure. The inclusion of airports in the charging demand analysis further highlights the potential for placing charging stations near transportation hubs, such as LaGuardia and JFK, where high traffic volumes and proximity to high-demand areas align with the need for charging infrastructure. As the EV market continues to grow, the strategic placement of charging stations in these high-demand areas will be crucial for supporting the transition to electric vehicles in NYC.

Discussion

The analysis of Figures 7 and 8 revealed key insights into how external events, such as the COVID-19 pandemic, impacted EV charging demand in NYC. Charging patterns were noticeably disrupted during the pandemic, with shifts in behavior that can be linked to the decrease in commuter traffic and changes in public transit usage (NREL, 2023). This mirrors findings from other studies, which have also documented how the pandemic led to temporary reductions in traffic congestion and energy demand across various sectors. Our results align with this, showing decreased demand at previously high-traffic areas such as airports and major highways, as fewer people traveled for work or leisure during the pandemic. However, the data also highlights a rapid rebound in charging demand as restrictions ease, indicating that charging infrastructure must be adaptable to such external disruptions.

Our findings align with several studies, which also emphasize the importance of strategically placed EV chargers in urban settings to meet peak demand (Davis, 2018). For example, fast chargers placed along major highways and at destination locations, such as hotels and airports, have been shown to ease range anxiety and encourage EV adoption. In our model, similar patterns were observed, where high-traffic areas, like major highways, saw substantial charging demand, validating our hypothesis that public charging demand in these areas needs to be prioritized. The list of key EV charging locations, including workplaces, public stations at malls, grocery stores, and tourist attractions, echoes the common places identified by other researchers (Shahriar et al, 2020). Our results suggest that these locations are critical in reducing range anxiety and ensuring convenience for EV users across NYC. Airports, in particular, stand out in our study due to their consistently high demand, which aligns with research emphasizing the importance of providing fast charging options at transportation hubs to accommodate long-haul travelers. This also supports the idea that more EV stations should be located in areas

of high transit and tourist activities, as indicated in the spatial distribution data of our results.

Our data strongly supports the hypothesis that a spatiotemporal model can be optimized to predict high-demand zones for EV charging in NYC. By identifying hotspots such as airports, major highways, and shopping centers, our model provides a framework that reflects the patterns of EV user behavior. This corresponds with predictive approaches highlighted in the literature (Shahriar et al, 2020), which focus on energy consumption and session duration in high-demand areas. The strong correlation between our predicted demand areas and actual usage supports the notion that urban planners and policymakers can use this model to allocate resources efficiently.

One limitation of this study is the lack of granular data on renewable energy integration at charging stations. While the focus of our model was primarily on demand forecasting and optimal station placement, future studies should integrate energy sources at a more detailed level to account for power grid impacts. Additionally, the study assumes that travel patterns and charging behavior will remain relatively stable over time, despite the clear disruptions caused by external factors such as the COVID-19 pandemic, as reflected in our data. This could limit the model's accuracy in the face of future, unpredictable events. Another key limitation is the absence of dynamic pricing data and specific user preferences for different types of charging (Shahriar et al, 2020). Incorporating user behavior in real-time, such as willingness to pay for faster charging or preferences for specific locations like grocery stores or airports, would enhance the predictive accuracy of the model.

The next steps in this research include expanding the spatiotemporal model to incorporate renewable energy forecasting at EV charging stations. As noted in the literature, integrating renewable energy, such as solar or wind, could alleviate the strain on the power grid during peak charging times and contribute to a net-zero carbon future (NREL, 2023). In addition, future iterations of the model could explore the impact of dynamic pricing strategies to encourage charging during off-peak hours, as well as incorporate behavioral insights from machine learning models to refine charging station locations. For example, using reinforcement learning to optimize charging schedules at community hubs could reduce both wait times and overall system costs (Moghaddam, 2018). By building on these insights, the model could offer a more robust, flexible, and sustainable approach to EV infrastructure planning in NYC.

In conclusion, this study's results underscore the importance of using predictive models to place EV charging stations strategically in high-traffic areas such as highways, airports, and urban centers. While our hypothesis is supported by the data, further research into renewable energy integration and behavioral dynamics will be crucial to ensuring the sustainability and scalability of EV charging networks.

Conclusion

We explored the optimal placement of electric vehicle (EV) charging stations in New York City using spatial-temporal analysis and machine learning models to predict areas of high demand. Our results highlighted significant patterns in EV charging behavior, such as the increased demand for charging infrastructure in highly traveled zones like airports, workplaces,

and public spaces, as well as the variation in charging needs between different boroughs. By comparing our findings with other research, we demonstrated that smart scheduling and predictive analytics can greatly improve the efficiency of charging networks, reduce grid strain, and provide more equitable access to EV infrastructure.

The broader implications of this work emphasize the importance of integrating data-driven models into urban planning for EV infrastructure. Our findings can guide policymakers and city planners in making informed decisions about where to install new charging stations, ensuring that charging infrastructure meets the growing demand while avoiding overburdening specific areas of the grid. Additionally, understanding consumer behavior through machine learning techniques enables more strategic deployment of resources, which is crucial for scaling EV adoption. Overall, this work contributes to the larger goal of reducing urban carbon footprints and fostering a more sustainable transportation system for the future.

Works Cited

- Bryden, T. S., Hilton, G., Cruden, A., & Holton, T. (2018). Electric vehicle fast charging station usage and power requirements. *Energy*, *152*, 322–332.
<https://doi.org/10.1016/j.energy.2018.03.149>
- Correction to: The driving downturn: A preliminary assessment. (2017). *Journal of the American Planning Association*, *83*(3), 328–328. <https://doi.org/10.1080/01944363.2017.1323527>
- Davis, S. J., Lewis, N. S., Shaner, M., Aggarwal, S., Arent, D., Azevedo, I. L., Benson, S. M., Bradley, T., Brouwer, J., Chiang, Y.-M., Clack, C. T., Cohen, A., Doig, S., Edmonds, J., Fennell, P., Field, C. B., Hannegan, B., Hodge, B.-M., Hoffert, M. I., ... Caldeira, K. (2018). Net-zero emissions energy systems. *Science*, *360*(6396).
<https://doi.org/10.1126/science.aas9793>
- Gnanavendan, S., Selvaraj, S. K., Dev, S. J., Mahato, K. K., Swathish, R. S., Sundaramali, G., Accouche, O., & Azab, M. (2024). Challenges, solutions and future trends in EV-technology: A Review. *IEEE Access*, *12*, 17242–17260.
<https://doi.org/10.1109/access.2024.3353378>
- Hu, L., Dong, J., Lin, Z., & Yang, J. (2018). Analyzing battery electric vehicle feasibility from taxi travel patterns: The case study of new york city, USA. *Transportation Research Part C: Emerging Technologies*, *87*, 91–104. <https://doi.org/10.1016/j.trc.2017.12.017>
- Ji, D., Zhao, Y., Dong, X., Zhao, M., Yang, L., Lv, M., & Chen, G. (2018). A spatial-temporal model for locating electric vehicle charging stations. *Communications in Computer and Information Science*, 89–102. https://doi.org/10.1007/978-981-13-1026-3_7
- June 27, 2023 | By Anna Squires | Contact media relations Share. (n.d.).
Building the 2030 national charging network. NREL.
<https://www.nrel.gov/news/program/2023/building-the-2030-national-charging-network.html>
- Khan, H. A., Price, S., Avraam, C., & Dvorkin, Y. (2022a). Inequitable access to EV charging infrastructure. *The Electricity Journal*, *35*(3), 107096.
<https://doi.org/10.1016/j.tej.2022.107096>
- Moghaddam, Z., Ahmad, I., Habibi, D., & Phung, Q. V. (2018). Smart charging strategy for Electric Vehicle Charging stations. *IEEE Transactions on Transportation Electrification*, *4*(1), 76–88. <https://doi.org/10.1109/tte.2017.2753403>
- New York City Department of Health and Mental Hygiene. (2024). *Walking, driving, and cycling: NYC data explorer*. NYC Health.
<https://a816-dohbsp.nyc.gov/IndicatorPublic/data-explorer/walking-driving-and-cycling/?id=2113#display=summary>
- NYC OpenData. (2024). *Traffic volume counts (2014–present)*. NYC Department of Transportation.
<https://data.cityofnewyork.us/Transportation/Traffic-Volume-Counts-2014-Present-/ertz-hr4r>

- Nguyen, D. M., Kishk, M. A., & Alouini, M.-S. (2024). Dynamic charging as a complementary approach in modern EV charging infrastructure. *Scientific Reports*, *14*(1). <https://doi.org/10.1038/s41598-024-55863-3>
- Rajput, A. A., Li, Q., Gao, X., & Mostafavi, A. (2022). Revealing critical characteristics of mobility patterns in New York City during the onset of covid-19 pandemic. *Frontiers in Built Environment*, *7*. <https://doi.org/10.3389/fbuil.2021.654409>
- U.S. Department of Energy. (2021). Electric drive vehicles (DOE/EE-2121). Alternative Fuels Data Center. https://afdc.energy.gov/files/u/publication/electric-drive_vehicles.pdf
- Shahriar, S., Al-Ali, A. R., Osman, A. H., Dhou, S., & Nijim, M. (2020). Machine learning approaches for EV charging behavior: A Review. *IEEE Access*, *8*, 168980–168993. <https://doi.org/10.1109/access.2020.3023388>

Exploring How Prompt Engineering Can Be Used Alongside Bloom's Taxonomy to Improve Learning Efficiency By Vardhan Madhavarapu

Abstract

This paper explores the potential of prompt engineering and how it can enhance learning efficiency by engineering prompts to fit into Bloom's Taxonomy. Bloom's Taxonomy, which includes six cognitive levels, Remember, Understand, Apply, Analyze, Evaluate, and Create, is a structured educational framework, commonly used to structure school curriculums. By using effective prompt engineering that will specifically target certain levels of the taxonomy, this paper will investigate how exactly it can be used to optimize your learning efficiency, as well as the prompting techniques that will allow you to do so. It will also dig into how to adapt learning techniques to best utilize AI in studying/learning in order to best use AI as a tool rather than as an information dump.

Introduction

The extremely quick advancements in artificial intelligence (AI) we have seen in recent years have already transformed many different niches fundamentally and will continue to alter various more, with education being one of the most heavily impacted so far. One of the highest-impact developments we have seen is generative AI, most notably large language models (LLMs) such as OpenAI's "ChatGPT." They have completely revolutionized information, and how it is used and understood (Cain, 2023). They also offer an incredible amount of opportunities for adaptive/personalized learning experiences (Heston & Khun, 2023). A technique known as prompt engineering is used to effectively harness the power of these LLMs. Prompt engineering involves carefully crafting the inputs, or "prompts," given to AI to produce specific outputs of higher quality than regular prompting (Lo, 2023). Research has shown that the specificity and format of prompts greatly impact the relevance and accuracy of responses generated by LLMs, making prompt engineering vital producing outputs that will allow for us to learn more efficiently (Cain, 2023; Lo, 2023). A framework in education that remains relevant, even in this new AI-dominant landscape is Bloom's Taxonomy. This taxonomy classifies cognitive skills into six different levels: Remember, Understand, Apply, Analyze, Evaluate, and Create. These levels range from basic recall to complex and creative thinking (Krathwohl, 2002). It structures educators' approaches to designing curricula that can target varying levels of cognitive function (Krathwohl, 2002). In the realm of education, given the powerful capabilities of LLMs, a key question arises: How can we use prompt engineering to address each level of Bloom's Taxonomy effectively, and as a result maximize learning efficiency? Addressing this question is important because prompt engineering, which is the process of formulating inputs to produce desired outputs from AI, has proved to be a key skill when using AI (Lo, 2023; Heston & Khun, 2023). Rather than simply asking an AI to summarize a lecture, or write notes for you, we will explore how certain prompt structures can increase our ability to learn said information. When used with cognitive frameworks like Bloom's Taxonomy, prompt engineering can transform our traditional educational practices by enhancing the effectiveness of AI-facilitated

learning across different cognitive levels (Cain, 2023; Heston & Khun, 2023). This paper will not only argue that prompt engineering can be used alongside Bloom's Taxonomy to improve educational products greatly but also reveal techniques for how to do so effectively. We will explore different methods to facilitate different types of learning, and how they can help us learn. Learners can optimize AI tools like LLMs to improve learning efficiency by tailoring prompts to address each cognitive level. This approach not only maximizes the effectiveness of AI-facilitated learning but also assures that AI interactions are educationally sound, raising deeper cognitive development across all learning stages (Krathwohl, 2002). The methodology of research used was to ask the same question to an LLM (ChatGPT) and to analyze the accuracy and the presentation of “learnable” information.

What is Prompt Engineering and Why is it Important?

Prompt engineering is a method that is used to maximize the effectiveness of LLMs. This method involves carefully creating inputs that are then given to LLMs to produce outputs that are more contextually sound (Lo, 2023). Unlike simple queries, these engineered prompts are generally designed with specific parameters in mind that will help get a more desired result. Some of these parameters can include the desired format, diction, target audience, etc. (Cain, 2023). This approach to creating prompts allows people to take advantage of AI at a higher level than someone who only understands regular prompting strategies. Due to its abilities, prompt engineering can be very helpful in educational settings to help people grasp information faster and better (Lo, 2023; Heston & Khun, 2023).

Prompt engineering is important because of its ability to lead outputs towards more specific objectives. By changing the structure of prompts, we can tailor outputs to help in different types of cognitive tasks, from basic recall to complex problem-solving, depending on the type of information that we are trying to learn (Cain, 2023; Heston & Khun, 2023). This technique doesn't only improve the quality of LLM outputs but it can also structure them to be more personalized, and when paired with the structural framework of Bloom's Taxonomy, it can help us reach our learning goals more efficiently and effectively (Krathwohl, 2002; Lo, 2023).

What is Bloom's Taxonomy?

Bloom's Taxonomy is a widely known framework for education that categorizes human cognitive skills into six different levels: Remember, Understand, Apply, Analyze, Evaluate, and Create as seen in Figure 1. Originally developed by Benjamin Bloom and later revised by many educational psychologists, this taxonomy is used as a guide for educators to create curricula that help develop a structured progression of learning in students that ranges from simple memorization to complex, critical thinking and creativity (Krathwohl, 2002). Each level builds on the previous one, encouraging learners to move from foundational knowledge acquisition to higher-order thinking skills. For example, as students move up levels in their schooling, from junior high to college to PhD learning, they will find themselves using more sophisticated levels of Bloom's taxonomy more often (Anderson & Krathwohl, 2001).

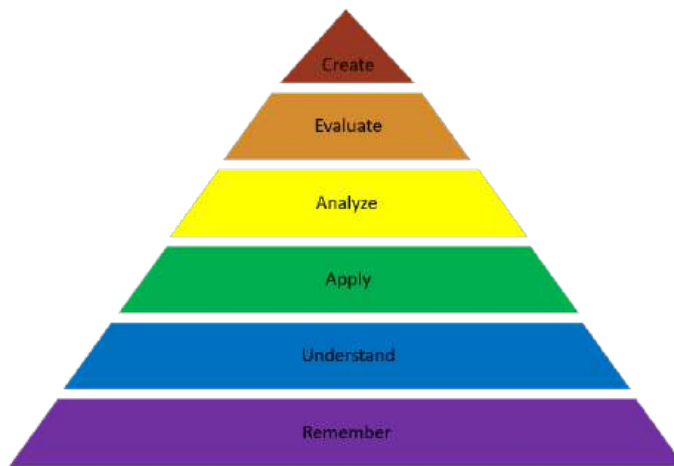


Figure 1: Bloom's Taxonomy Levels (revised)

The strength of Bloom's Taxonomy is mostly in its ability to provide a clear roadmap for education. By clearly defining different types of cognitive tasks, it allows educators to create curricula that are targeted at certain cognitive levels, and therefore different learning outcomes (Krathwohl, 2002). For instance, at the "Remember" level, activities may involve a basic recall of facts, while at the "Analyze" or "Create" levels, students could be asked to simplify complex concepts or create entirely new ideas (Anderson & Krathwohl, 2001). This structured approach doesn't only help in the organization of teaching, but also in improving the depth, as well as the retention of learning. Unbeknownst to many students, this structure encourages students to engage with certain material at a desired cognitive level, leading to students understanding certain curricula at just the level they need to be at (Krathwohl, 2002; Anderson & Krathwohl, 2001).

Why Prompt Engineering is Essential When We Want to Actually Learn Things

Let us use mathematics as an example for the core of our learning objective. Let's put you in a scenario: you are an 11th-grade student who wants to use ChatGPT to help you ace your next AP Calculus AB test on integrals. To start, we will use an obviously vague and ineffective prompt: "Teach me calculus." Similar to the input, the output is incredibly vague, and briefly summarizes the two main concepts in calculus: differentials and integrals. While this approach is a start, there are many ways to further improve this prompt for ChatGPT to return a more viable response, and all of these improvements are considered prompt engineering (Lo, 2023). Rather than taking many tiny little steps up to reach our final approach to prompting, we will take a much bigger step, and take a look at a prompt that an 11th grader with some LLM experience would input: "Help me pass my AP Calculus AB integrals test." Although much more specific than the original prompt, this input still gives an extremely brief summary of the integrals unit, summarizing everything in just 4 headers and 11 bullet points, paired with some very vague

studying and exam strategies. Although this response isn't very helpful, it is still much more effective than the starting prompt (Heston & Khun, 2023).

In order to use AI to the fullest potential that we can, we have to look at it as a tool, rather than a robot that can do everything for you (Lo, 2023). It is best used when paired with a human, rather than when it is told to do something all alone. Instead of making an LLM blindly try to help us pass a test, we must provide it with more specific goals. Our new prompt will give it a more specific range and goal to work with: “Summarize unit 6 of AP Calculus AB. Specifically from 6.3 - 6.9.” This prompt looks to work much better. It gives a small description of each unit, which can be made longer with a small adjustment in the prompt, as well as what skills are learned, to show why each topic is useful. However, when the output is examined further, we notice a big problem. The topics in the unit are not accurate as shown in the figures below..

Figure 2: Screenshot of Unit 6.3-6.9 from the AP Calculus AB Textbook. *Source*: College Board. (2021). *AP Calculus AB & BC Course and Exam Description* (p. 29).

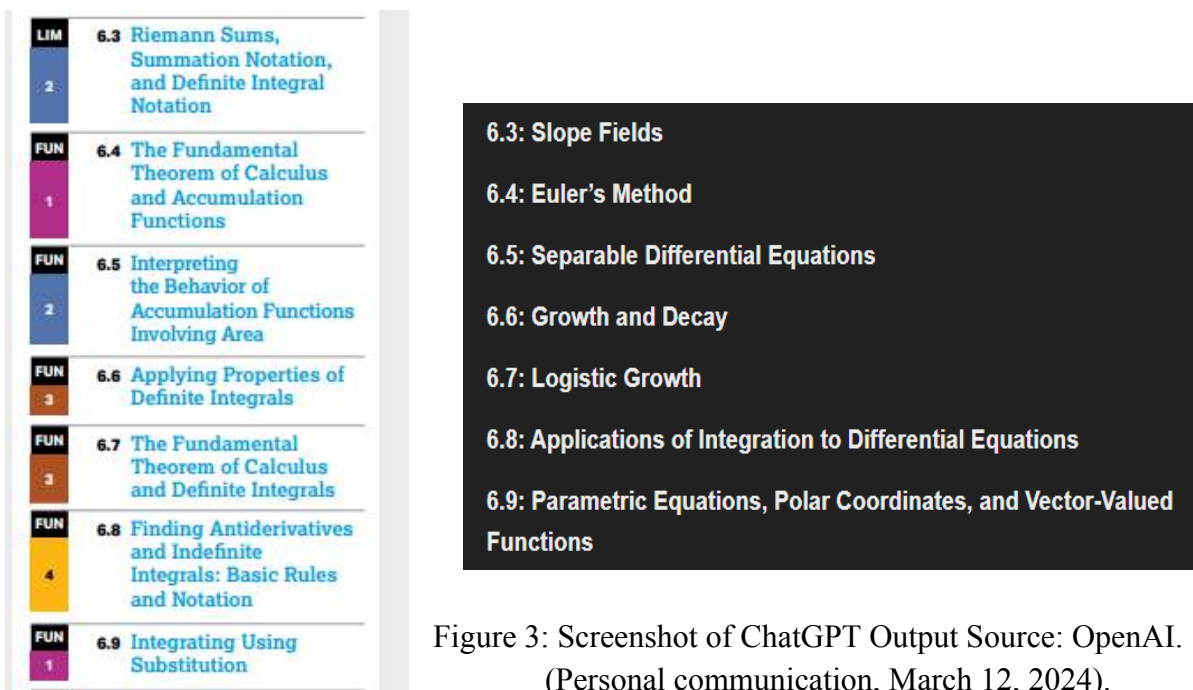


Figure 3: Screenshot of ChatGPT Output *Source*: OpenAI. (Personal communication, March 12, 2024).

What Makes a Good Prompt?

A good prompt is essential to use an LLM effectively. Understanding how these models function is crucial in creating prompts that lead to accurate responses (Lo, 2023). LLMs create their outputs by predicting what is most likely the next chain of information that would follow the input they received based on patterns learned from their training data (Cain, 2023). This not only means the quality of the input directly influences the quality of the output but also that the better the context that an LLM is given, the more accurate the response (Cain, 2023).

Providing as much context as possible is arguably more important than the quality of the prompt (Lo, 2023). Since LLMs will try to show the most likely continuance of an input that is given to them based on their training data, context can help them narrow their search to the correct sectors of their information (Heston & Khun, 2023). Although context provided in a question within the prompt indubitably helps an LLM generate a more specific and contextually sound response, there is one way to provide context that beats all.

Near the end of 2023, OpenAI implemented two features that would change prompt engineering forever. The former is the ability to upload PDF files which at the time, just seemed like a quality of life (QoL) update, as many people were accustomed to copying and pasting documents into their prompts. The ability to upload many PDFs to our prompts helped us work around the character limit of prompts as well as greatly improve the formatting, as some documents are too massive to fit into a prompt. However, this ability paired with the “memory” feature, forever changed prompt engineering. This update added a session-based memory for ChatGPT. By adding files to the memory base, ChatGPT will look through this specified data first to influence its outputs as shown in the screenshot from OpenAI’s website (OpenAI 2024). This completely revolutionized the way we can provide context, and taking advantage of these features can help us massively in our quest to use prompt engineering to learn more effectively (Heston & Khun, 2023).

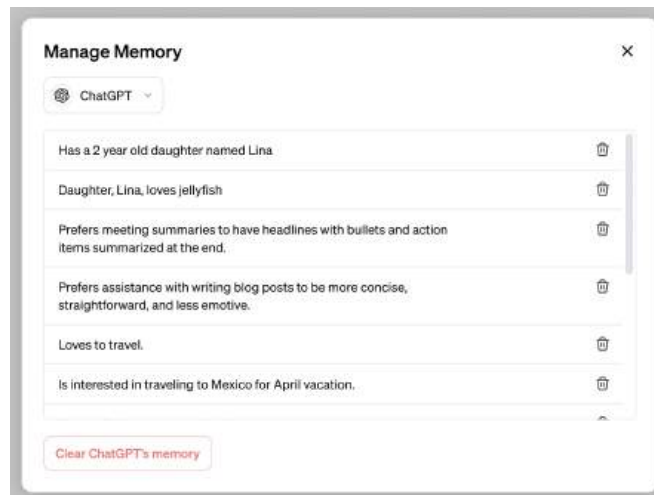


Figure 4: Screenshot of ChatGPT’s Memory Feature.

How Can We Use Bloom’s Taxonomy to Help Us Learn More Effectively?

AI is only as powerful as the person using it. The problem with how AI is commonly used among students is that they use it to look up things that they are too lazy to figure out. However, being spoonfed just answers does not equal learning (Lo, 2023). So, instead of using AI to replace our thinking, we must look at it as an extension of our thinking. So instead of

expecting AI to somehow teach us everything we need to know, we must use it to guide our thinking, which can help lower our mental load (Anderson & Krathwohl, 2001).

There are 3 distinct phases of understanding information. We will take a look at them one at a time. The first phase of learning is the gathering of new information. There is a sequence that many people go through every single time they learn something new. Think about learning programming with Python. It would be most effective to learn what the basic statements like “print” and “while” do first. That can help set the foundation for learning how to handle and manipulate data, and together those could then help you learn how to program a game. We could always start by trying to code a game first and work backward, but this would increase our mental load which can lower our learning ability. The order of learning is very important, not just for comprehension, but also to save time trying to learn something before we are ready for it (Anderson & Krathwohl, 2001). However, figuring out this order can be stressful and difficult, which is where we can use prompt engineering to ease our load (Heston & Khun, 2023).

A great technique when prompting an AI is to tell it that it is a certain type of person. Let’s take our Python programming example again. In order to get a framework of what order we should learn Python in, we can give ChatGPT this prompt: “You are an expert programmer in the language of Python. How would you recommend a beginner to learn Python? Create a syllabus and lesson plan that spans 4 weeks for me to learn this information effectively.” You can modify this prompt to fit whatever you are trying to learn, whether it’s something generic that you are interested in or even a certain unit that you are learning in school. For example, you could also add “Give me a list in bullets of other topics and concepts I should explore to understand programming Python more completely, and also tell me why.” This addition can prove to be very important, as just like LLMs, humans can understand concepts much more effectively when they truly understand the context of what they are learning. You have probably experienced the effects of truly grasping the context of something you were learning before. Usually, this is what causes that “click” where everything suddenly makes sense (Krathwohl, 2002). When we understand how a small topic fits into the bigger picture, it helps us understand a topic better, and this context is vital to our quest to learn more effectively. The way we are using ChatGPT with prompt engineering is to create a sequence to approach learning as well as guide us in critically thinking about information to create an outline of how information goes together in the big picture (Anderson & Krathwohl, 2001).

This is where the lower two levels of Bloom’s taxonomy start to show. When we are learning using these methods mentioned above, we are reading summaries, and understanding ways that certain information can fit in the bigger picture. We can memorize certain facts and define terms just by reading these summaries given by ChatGPT when we ask for details about a certain topic, which all fits the “Remember” level. When we add to our prompts and ask for context to understand the bigger picture, we start to be able to discuss our topic better, as well as explain/teach it to other people, which fits the “Understand” level (Anderson & Krathwohl, 2001). By just using ChatGPT to learn this way, we can already get a basic understanding of our

desired topic more efficiently and effectively, as we save time and energy by avoiding learning in the wrong way (Lo, 2023).

The next phase of learning is to take what we learn and connect it all to understand how something works. We want to find out how one concept relates to other concepts. To do this we must analyze and compare relationships, as when some concept seems less relevant, we are less likely to understand why it is important, which can hurt our learning as we are more inclined to forget it. We can use AI to help us immensely in this step because we don't have to spend time searching for those relationships anymore as it can create a great starting point within seconds. To start, we can ask ChatGPT to “create a table to compare and contrast certain different concepts and explain why the relationship between these concepts is significant in understanding the bigger picture.” This works well because it helps establish context better and helps us connect the dots to understand why a certain topic matters and how one topic can affect another (Lo, 2023). But let's imagine we are completely stuck, and don't even know what concepts to compare and contrast. With a simple prompt, we can get ChatGPT to give us some concepts to compare: “Help me explore additional concepts related to *insert major concept here*.” We can use this prompt alongside the last one to strengthen our connection with the big topic. Being able to compare and contrast certain topics is where the “Analyze” level of Bloom’s Taxonomy is shown (Anderson & Krathwohl, 2001). Being able to use this higher-order thinking can help massively when trying to make connections between different sectors of an idea, and can overall immensely improve our understanding of our desired topic. At this point in the learning process, we can start to reinforce topics that we have learned about as well as their contexts in the bigger picture (Krathwohl, 2002).

Finally, we want to put what we have learned to the test. Doing practice problems has always been a very good way to study a topic effectively, as it challenges whether something has stuck with you or not. You can utilize ChatGPT to get practice questions about your desired topic. After giving it context, you can ask ChatGPT to give you a practice test to test what you have learned. These questions can be given in a specific format, like multiple choice or short answer, and ChatGPT can also explain any problem and doubts to you. However, while these questions are not necessarily bad, we can do better by utilizing Bloom’s Taxonomy in your prompt, as it will enhance the response to be better for your learning. Depending on the level of understanding of a concept you wish/need to attain, you can use certain levels of Bloom's Taxonomy in your prompts to help you. For example, you can ask ChatGPT to give you a practice test of the AP Biology Kinetics unit, and it will give you a decent practice test that is styled like the AP test, however, to gain a better understanding of the topic, we can use a level of Bloom's Taxonomy. In this scenario, we will use the fourth level: Analyze. This time we will ask the same question, however, we will specify that these questions should target the “Analyze” level of Bloom’s Taxonomy (Krathwohl, 2002). If these questions are too hard, you can start at a lower stage and work your way up. Being able to answer questions from as many levels of Bloom’s Taxonomy will guarantee you have a good grasp of a certain concept, and if you haven't

reached a certain level of understanding, you can always start at a lower level (Anderson & Krathwohl, 2001).

Comparison to Existing Literature

While several studies, such as Cain (2023) and Lo (2023), delve into the use of prompt engineering, even in learning, they mostly focus on its general applications in learning environments or information literacy. For example, Cain's research dives into the potential of prompt engineering to transform fields but doesn't completely connect it to learning, especially not into structured frameworks like Bloom's Taxonomy. Lo's work talks about the CLEAR framework for enhancing information literacy with prompt engineering, but also connects it to learning strategies and how to implement it into something like Bloom's Taxonomy. Our research is unique due to our focus on how to integrate Bloom's Taxonomy into our learning methods in order to improve efficiency. By doing so, this research shows the potential of prompt engineering as a tool as well as a means to optimize our learning efficiency by specifically targeting cognitive skills.

Next Steps

While Bloom's taxonomy has traditionally been applied to more traditional education methods like within school curriculums, its effectiveness in working alongside artificial intelligence could potentially show massive benefits to learning outcomes by improving cognitive engagement. Future research can involve controlled experiments comparing and contrasting the effectiveness of learning with this method (or similar methods) using AI with traditional learning done within a classroom setting. Another major future topic can include studies on how prompt engineering can be used to understand personal learning patterns, and how it can be implemented into AI-assisted learning in order to further improve learning outcomes. The method explored within this paper is quite general among humans, so finding methods where learners would be able to use prompt engineering to take a deep dive into their own brains could be a substantial finding and could potentially improve learning and comprehension exponentially.

Conclusion

Ultimately, the use of prompt engineering with Bloom's Taxonomy gives us an improved approach to learning with better efficiency. By blending Bloom's Taxonomy with prompt engineering in ways similar to what was discussed in this paper, learners can better target specific cognitive skills in order to retain information faster and for longer. Being able to use AI as a just tool is a crucial step to being able to use it to its full potential, in any situation. This framework could help shape future developments of AI-assisted education, further improving information retention and adaptive learning methods. Prompt engineering is evidently becoming more and more important as our world starts to get more dependent on AI, and being able to use it effectively, looks to be an extremely promising skill to hav

Works Cited

- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of educational objectives. Allyn & Bacon.
- Cain, W. (2023). Prompting Change: Exploring Prompt Engineering in Large Language Model AI and Its Potential to Transform Education. *TechTrends*, 68(1), 47-57.
<https://doi.org/10.1007/s11528-023-00896-0>
- College Board. (2021). AP Calculus AB & BC Course and Exam Description. College Board.
<https://apcentral.collegeboard.org/pdf/ap-calculus-ab-and-bc-course-and-exam-description.pdf>
- Heston, S., & Khun, J. (2023). Prompt Engineering in Medical Education: A Pathway to Enhanced Learning and Clinical Practice. *Innovative Medical Education*, 2(1), 19-34.
<https://doi.org/10.1016/j.ime.2023.02.019>
- Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, 41(4), 212-218. https://doi.org/10.1207/s15430421tip4104_2
- Lo, L. S. (2023). The CLEAR Path: A Framework for Enhancing Information Literacy through Prompt Engineering. *The Journal of Academic Librarianship*, 49(1), 102720.
<https://doi.org/10.1016/j.acalib.2023.102720>
- OpenAI. (2024). *ChatGPT Output Screenshot*. OpenAI. (Personal communication, March 12, 2024).
- OpenAI. (2023). Screenshot of memory feature description. Retrieved from <https://openai.com/index/memory-and-new-controls-for-chatgpt/>

Climate Change Has No Significant Negative Effects on the Suitable Habitats for Granular Poison Frog (*Oophaga granulifera*) According to Species Distribution Modeling

By Juliana Careaga

Abstract

Almost half of amphibian species globally are experiencing population declines. Future climate change effects, such as prolonged droughts and increased temperatures, are expected to contribute to the decline of these species. The *Oophaga granulifera* (granular poison frog) located in Costa Rica in particular is an endangered frog species that is experiencing such declines. The *O. granulifera* controls the populations of small invertebrates, such as ants and mites, as their predators and are used as biodiversity indicators due to their sensitivity toward environmental changes. This study aims to determine potential distribution and habitat suitability of the *O. granulifera* by developing a species distribution model (SDM) to analyze the effects of climate change on the species between the present and the future (years 2061-2080). I collected occurrence records of *O. granulifera* from the Global Diversity Information Facility (GBIF) database and 19 climatic variables from the WorldClim database to produce a generalized linear model (GLM)—the model aimed to predict the effects of present and future climatic variables on the species distribution of *O. granulifera*. The SDM showed increased areas of suitable habitat from the present to the future, indicating that climate change may have no significant negative effects on the species distribution. Conservationists aiming to increase the *O. granulifera* population should focus on other factors that may contribute to their endangerment, such as habitat loss and diseases, since gradual climate change proved to not be a significant threats to population decline.

Keywords *Oophaga granulifera*, Granular poison frog, Climate change, Species distribution, modeling, Suitable habitat, Temperature

Introduction

A large proportion (43%) of amphibian species are experiencing population declines globally which only continue to worsen (Wake & Koo, 2018). Factors that have led to this population declines include habitat destruction, drainage of wetlands, the rise of urbanization, and a pathogenic chytrid fungus called *Batrachochytrium dendrobatidis* (Bd). Bd was discovered in the late 20th century and is the forefront of the investigation for determining amphibian decline (Wake & Koo, 2018). The fungus invades the amphibian's permeable skin and disrupts their osmotic balance, therefore inducing heart failure and killing the organism infected (Wake & Koo, 2018). Climate change causes frogs to be more susceptible to Bd. Bd shows significantly higher infection rates during cool-dry season than during the warm-wet season (Longo & Zamudio, 2016). This seasonal trend shows that certain environmental conditions are more favorable to the pathogen than others. Bd maintains a strong threat to amphibian species but especially frogs.

Future climate change is predicted to worsen the situation (Corn, 2005). With their sensitivity to temperature and precipitation, amphibians are expected to suffer under future climate change conditions such as reduced soil moisture which could reduce prey species and eliminate habitat (Corn, 2005). Reduced snowfall and increased summer evaporation also poses an issue as it may affect the occurrence of seasonal wetlands, a crucial and popular habitat for amphibians (Corn, 2005). Prolonged droughts caused by increased temperatures, for example, can significantly reduce the number of wetlands therefore eliminating the habitats available for amphibians (Walls et al., 2013). Extreme climate change effects like droughts contribute to amphibian habitat loss, consequently leading to lower rates of survival and population decrease.

Despite their decline, amphibians are essential to ecosystems and are the most abundant vertebrate species (Wake & Koo, 2018). One of the most well-known amphibians are frogs. Frogs play vital roles in wet and dry ecosystems as they are both prey and predators to several organisms, affecting the food chains in many ecosystems (Dorcas & Gibbons, 2011). They are also bioindicators—indicators of an ecosystems' health—because of their heightened sensitivity to environmental problems and changes (Dorcas & Gibbons, 2011). Their permeable skin makes them more susceptible to toxins in their environment (Dorcas & Gibbons, 2011). Frogs are found in terrestrial and freshwater aquatic habitats including, but not limited to, areas in and around lakes, ponds, swamps, wetlands, and mountain streams (Dorcas & Gibbons, 2011). Lakes, in particular, are increasing in temperature due to climate change. Lake warming decreases the population of zooplankton in the body of water, limiting the food supply of frogs' prey and disrupting the food web (Havens & Jeppesen, 2018). Climate change effects on aquatic areas therefore affect their ability to thrive in their habitat.

One frog species in particular that is listed as a vulnerable species on the IUCN Red List is the *Oophaga granulifera*, also known as the granular poison frog. The granular poison frog is common in the Pacific portion of Costa Rica and is declining due to changes in their habitats (IUCN, 2020). This species specifically is threatened by expanding agriculture through plantations of oil palm, banana, and pineapple and the spraying of pesticides and fungicides, and is usually found in small streams within humid lowland forests (IUCN, 2020). The granular poison frog is a species of dart frog that control the populations of small invertebrates as their predators, and are used as biodiversity indicators due to their sensitivity toward environmental changes (McGugan et al., 2016).

To determine the habitat suitability and potential distribution for the *O. granulifera*, I developed a species distribution model (SDM) for the species. My objectives are as follows: (a) to identify the environmental factors in Costa Rica associated with Granular Poison Frogs; (b) to predict present and future habitat distribution for Granular Poison Frogs using available occurrence records and literature records.

Methods

Data collections

I obtained occurrence records of the *O. granulifera* from Global Biodiversity Information Facility (GBIF) GBIF.org (13 June 2024) GBIF Occurrence Download <https://doi.org/10.15468/dl.bvfdgq>. The GBIF is a database and data infrastructure supported by multiple national governments, aimed at providing open-access biodiversity data for the public and the science community. The total number of occurrences for *O. granulifera* downloaded is 434. These occurrences are found mainly in Costa Rica. The timeline of the sightings for this species ranges from 1958 to 2024. I filtered to keep only occurrences with coordinates. The latitudes range from 8 to 10 and longitudes range from -84 to -83.

Environmental Variables

I acquired environmental variables representing the current climatic conditions from the WorldClim database (Fick and Hijmans, 2017) at the resolution of 2.5 arc min. The environmental variables included: Bio1 = Annual mean temperature, Bio2 = Mean diurnal range (max temp – min temp) (monthly average), Bio3 = Isothermality (Bio1/Bio7) * 100, Bio4 = Temperature Seasonality (Coefficient of Variation), Bio5 = Max Temperature of Warmest Month, Bio6 = Min Temperature of Coldest Month, Bio7 = Temperature Annual Range (Bio5-Bio6), Bio8 = Mean Temperature of Wettest Quarter, Bio9 = Mean Temperature of Driest Quarter, Bio10 = Mean Temperature of Warmest Quarter, Bio11 = Mean Temperature of Coldest Quarter, Bio12 = Annual Precipitation, Bio13 = Precipitation of Wettest Month, Bio14 = Precipitation of Driest Month, Bio15 = Precipitation Seasonality (Coefficient of Variation), Bio16 = Precipitation of Wettest Quarter, Bio17 = Precipitation of Driest Quarter, Bio18 = Precipitation of Warmest Quarter, and Bio19 = Precipitation of Coldest Quarter. Using R software (R Core Team, 2023) and raster package (Hijmans, 2024), I trimmed the environmental variables so that the geographic range of environmental variables was contained with 0 and 15 degree latitudes and -90 and -75 degree longitudes. The extent of the environmental variables covers and extends beyond the latitudinal and longitudinal ranges of *O. granulifera*.

In addition, to project species future suitable areas, I acquired environmental variables representing future climatic conditions for the years 2061- 2080 under the model MPI-ESM1-2-HR and CMIP6 (Gutjahr et al., 2019) The future climatic raster layers were also trimmed using the sampling extent mentioned above.

Modeling Strategy

To investigate the suitable habitat areas of *O. granulifera*, we performed a multivariate generalized linear model (GLM). I randomly selected 10,000 pseudo-absence, or “background” points, within the range of the environmental variables (see section above). The number of 10,000 pseudo-absence points followed the recommendation by Barbet-Massin et al. (2012). After running the GLM model, I converted the continuous probability of habitat suitability to a binary response output (suitable = TRUE or FALSE). The suitability cut-off was equal to the threshold at which the sum of the model sensitivity (true positive rate) and specificity (true negative rate) is highest (Field et al. 1997; Liu et al. 2011). I then used five-fold cross validation

to evaluate the trained GLM; in each cross validation, 80% of the occurrence data were training data, leaving the remaining 20% as testing data. We evaluated the performance of GLM by calculating the Area Under the ROC Curve (AUC). We then projected future suitable habitat areas for *O. granulifera* in 2061-2080.

Results

The final GBIF dataset of *O. granulifera* contains 310 records with coordinates. The GLM species distribution model shows an AUC value of 0.995, which indicates that it is an accurate classifier between positive and negative cases (Fig 1). The GLM model output shows that Bio2, 5, 9, 10, 12, 13, and 14 show significant positive correlation to predicting species presence, and Bio1, 3, 4, 6, 7, 8, 11, 15, 16, 17, 18 and 19 show significant negative correlation to species presence (Table 1). The comparison between current and future suitable habitat area reveals that areas of suitable habitat will increase (Fig 2).

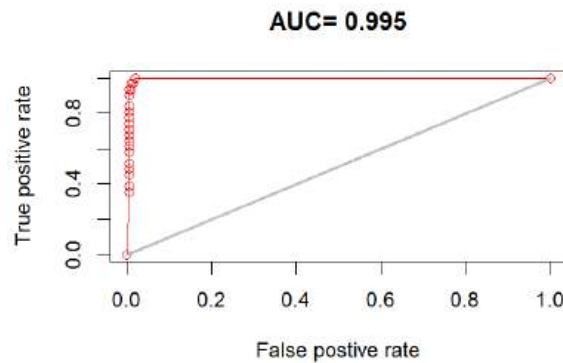


Fig 1: ROC (Area under the receiver operator) curve, demonstrates the accuracy of the SDM model of the *O. granulifera* in distinguishing positive and negative cases.

	A	B	C	D	E
	Estimate	Std. Error	z value	Pr(> z)	
1					
2	(Intercept)	396.3651614	473.659128	-0.8368591986	0.4025717172
3	wc2.1_2.5m_bio_1	-13.42437703	57.50620962	-0.2334423163	0.8154180347
4	wc2.1_2.5m_bio_2	55.4760011	47.26674346	1.173881138	0.2405227683
5	wc2.1_2.5m_bio_3	-5.899591978	6.417186444	-0.9347697827	0.3495069553
6	wc2.1_2.5m_bio_4	-5.701797404	2.156438478	-2.644280721	0.008191311908
7	wc2.1_2.5m_bio_5	5204126.975	19113545.5	0.2722743965	0.7854110491
8	wc2.1_2.5m_bio_6	-6204178.003	19113546.59	-0.2722769413	0.7854080848
9	wc2.1_2.5m_bio_7	-6204189.975	19113545.76	-0.2722775794	0.7854085942
10	wc2.1_2.5m_bio_8	-4.86143251	5.846544412	-0.8609571014	0.3892616705
11	wc2.1_2.5m_bio_9	13.69535267	0.437673933	1.621804269	0.1048174165
12	wc2.1_2.5m_bio_10	257.563234	101.5243421	2.534469976	0.01126195059
13	wc2.1_2.5m_bio_11	-203.15915	78.70152996	-2.581174967	0.0098460783672
14	wc2.1_2.5m_bio_12	0.002185796422	0.01037139306	0.2107524427	0.8330904502
15	wc2.1_2.5m_bio_13	0.0651730697	0.02826818403	2.305527985	0.02113702096
16	wc2.1_2.5m_bio_14	0.06191391796	0.07696465055	0.8044461848	0.4211393482
17	wc2.1_2.5m_bio_15	-0.7879855141	0.3885930154	-2.021816313	0.04319533182
18	wc2.1_2.5m_bio_16	-0.054898443441	0.02487577344	-1.990598315	0.0429676606
19	wc2.1_2.5m_bio_17	-0.07243904609	0.04330722282	-1.672678606	0.09439068309
20	wc2.1_2.5m_bio_18	-0.0001678410036	0.01089032944	-0.01070392070	0.99145907
21	wc2.1_2.5m_bio_19	-0.011869908	0.005897980948	-1.181029484	0.2375870493

Table 1: Generalized linear model (GLM) output for predicting the effects of climatic variables on *O. granulifera* species presence.

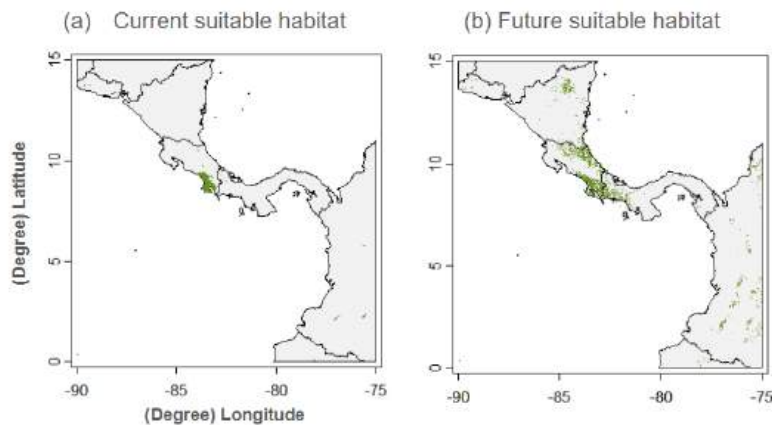


Fig 2: Species Distribution Model (SDM) for showing areas of current and future suitable habitat for the *O. granulifera* species.

Discussion

The Generalized Linear Model (GLM) showed that bioclimatic variables 2, 5, 9, 10, 12, 13, and 14 had a significant positive correlation with the distribution of *O. granulifera*. Variables 2, 5, 9, and 10 correspond to temperature being mean diurnal range, max temperature of warmest month, mean temperature of driest quarter, and mean temperature of warmest quarter, respectively. As these factors increase, so does habitat suitability for the species. Because they are directly proportional, if the factors decrease so does habitat suitability. This shows that overall higher temperatures make areas more environmentally advantageous for the species and will therefore lead to more areas of suitable habitat. These variables have a positive correlation to species distribution because granular poison frogs are already adapted to warmer temperatures due to their location close to the equator in Costa Rica. Increased temperatures are therefore not a threat to the species but can help the frogs maintain their optimal body temperature and enhance their activity and breeding behavior. The GLM also showed that bioclimatic variables 1, 3, 4, 6, 7, 8, 11, 15, 16, 17, 18, and 19 had a significant negative correlation with the distribution of *O. granulifera*. Variables 1, 3, 4, 6, 7, 8, and 11 correspond to temperature being annual mean temperature, isothermality, temperature seasonality, min temperature of coldest month, temperature annual range, mean temperature of wettest quarter, and mean temperature of coldest quarter, respectively. Variables 15, 16, 17, 18, and 19 correspond to precipitation being precipitation seasonality, precipitation of wettest quarter, precipitation of warmest quarter, and precipitation of coldest quarter, respectively. The SDM model is fairly accurate but it is important to recognize its limitations. It is difficult to interpret the variables of a species distribution model (SDM) because of collinearity of the predictor variables. Because environmental variables like temperature and precipitation are highly correlated, it is difficult to distinguish their individual effects on the species distribution and compare their importances.

In the species distribution model, the area of suitable habitats for the *O. granulifera* in Costa Rica from present day to the years 2061-2080 increased. Based on these results, climate

change does not negatively impact habitat suitability and is therefore not a threat towards the species nor contributing to their endangered state. However, it is essential to recognize that although the model demonstrates an increase in suitable habitat for the species, the model only accounts for variables that correspond to climate change, not any other factors. So although climate change is not a disadvantage for this species, it doesn't mean that the number of suitable habitats won't decrease or their endangered state won't worsen in the future due to other factors that weren't accounted for through these results. The SDM model is also based on variables that account for gradual climate change. Gradual environmental changes like a slow increase in temperature may not affect the species since they can most likely adapt over time to warmer climates. More extreme and rapid environmental effects caused by climate change are not accounted for by the SDM, however, these are factors that would potentially impact suitable habitats.

Other studies have shown that future climate change is a strong threat to amphibian species (Corn, 2005). Prolonged droughts caused by a rise in temperatures, for example, can significantly reduce the number of wetlands therefore eliminating the habitats available for amphibians (Walls et al., 2013). Droughts caused by El Nino, a more extreme and short term climate change, also cause breeding ponds to dry, killing the eggs and tadpoles of anurans (Corn, 2005). However, the data collected from the SDM shows otherwise. Bioclimatic variables 2, 5, 9, and 10 have a direct relationship with species distribution of *O. granulifera* which demonstrates that as temperatures increase, so does habitat suitability. While it was determined through other studies that increased temperatures reduce the number of suitable habitats for amphibians, the SDM proves the opposite for this specific amphibian species. It is important to recognize that the other study is analyzing amphibian species as a whole instead of one specific species through the SDM, which may explain the conflicting results. It was also determined by other studies that lake warming caused by higher temperatures from climate change kills the Zooplankton population which would serve as food for frogs (Havens & Jeppesen, 2018). Lake warming caused by climate change would therefore decrease the number of habitats suitable for frogs. However, variables of temperature in the SDM showed positive correlation with species distribution, showing that increasing temperature actually increases the number of suitable habitats. It is important to note that the study only focuses on frog species in general while the SDM is analyzing only one specific frog species. The conflicting results between the study and SDM results may also be explained by the fact that lake warming primarily affects aquatic frog species while the granular poison frog in Costa Rica is more terrestrial. The *O. granulifera* forages for food on land instead and is more likely to feed on invertebrates such as ants or mites (McGugan et al., 2016).

The SDM results showed an increased habitat suitability from the present to 50 years in future which suggests that climate change has no significant negative effect on the *O. granulifera*. Gradual trends of climate change may have even contributed to the increase in areas of suitable habitat. It is important to recognize, however, that the SDM only accounts for bioclimatic variables associated with climate change. While gradual climate change proved to

not be a threat to the *O. granulifera*, there are several other factors not taken into account that are the reason for endangering the species. Therefore, species conservation for the granular poison frog should put its efforts more on prevalent issues and factors that actually contribute to its endangerment, such as extreme climate events, the rise of urbanization, the spread of pathogenic fungi such as *Bd*, and the spraying of pesticides and fungicides, instead of focusing on gradual climate change.

Works Cited

- Corn, Paul S. "Climate Change and Amphibians." *Animal Biodiversity and Conservation*, vol. 28, no. 1, 2005, pp. 59–67. Accessed October 10, 2024.
<https://doi.org/10.32800/abc.2005.28.0059>.
- Dorcas, Michael E., and Whit Gibbons. *Frogs: The Animal Answer Guide*. Johns Hopkins University Press, 2011. Accessed October 10, 2024.
- Havens, Karl, and Eric Jeppesen. "Ecological Responses of Lakes to Climate Change." *Water*, vol. 10, no. 7, 2018, article 917. Accessed October 10, 2024.
<https://doi.org/10.3390/w10070917>.
- International Union for Conservation of Nature (IUCN). "Oophaga Granulifera." *IUCN Red List of Threatened Species*, 2020. Accessed October 10, 2024.
<https://www.iucnredlist.org/species/55186/3025447>.
- Longo, Amanda V., and Karen R. Zamudio. "Environmental Fluctuations and Host Skin Bacteria Shift Survival Advantage Between Frogs and Their Fungal Pathogen." *The ISME Journal*, vol. 11, no. 2, 2016, pp. 349–361. Accessed October 10, 2024.
<https://doi.org/10.1038/ismej.2016.138>.
- McGugan, Jessica R., et al. "Ant and Mite Diversity Drives Toxin Variation in the Little Devil Poison Frog." *Journal of Chemical Ecology*, vol. 42, no. 6, 2016, pp. 537–551. Accessed October 10, 2024.
<https://doi.org/10.1007/s10886-016-0715-x>.
- Wake, David B., and Michael S. Koo. "Amphibians." *Current Biology*, vol. 28, no. 21, 2018. Accessed October 10, 2024.
<https://doi.org/10.1016/j.cub.2018.09.028>.
- Walls, Stephen, William Barichivich, and Matthew Brown. "Drought, Deluge and Declines: The Impact of Precipitation Extremes on Amphibians in a Changing Climate." *Biology*, vol. 2, no. 1, 2013, pp. 399–418. Accessed October 10, 2024.
<https://doi.org/10.3390/biology2010399>.

When will advancing technology make most of us unemployable? What should we do about this? By Davin Miao

So many hardworking individuals wake up before sunrise and arrive back home just so they can put food on the table, but what if they were at risk of being automated? In less than a year, technological progress has created new job opportunities and industries. Moore's Law states that computed power doubles approximately every two years, dangling the prospect of mass unemployment over us. The current wave of AI is predicted to cause mass automation once issues like AI hallucinations are resolved, allowing machines to handle tasks with near-perfect accuracy. The question is what measures we, governments, and big technology companies can take to mitigate impacts and ensure a future where technology benefits society rather than increasing unemployment.

The appearance of AGI (Artificial General Intelligence) can have profound implications for mass automation because of its unprecedented capabilities to outperform the future workforce. Currently, Narrow AI, including AlphaGo, excels at specific tasks like detecting qualities and relationships, but its greatest weakness is the lack of generalization, or the ability to apply skills learned in one activity to different situations. This fundamental barrier means that while an AI might excel at chess, it cannot transfer its abilities to similar tasks. AGI's ability to learn, understand, and apply knowledge across many activities reduces the need for human monitoring. Max Tegmark has noted that AGI's ability to learn from each interaction provides an opportunity for exponential growth in its problem-solving abilities. AGI's rapid improvement in handling complicated tasks leads to a higher rate of unemployment because it will surpass human capacity.

Solving hallucinations, or inaccurate information, will lead to mass unemployment as it increases AI's reliability. Hallucinations can occur unpredictably in Large Language Models as they may have "data voids" or "information gaps", which potentially cause risks including the spread of misinformation, the generation of errors in decision-making, and mistrust in AI applications. These errors have been a serious barrier to fully relying on AI, thus hindering automation. For example, Google's new AI generative generated "odd" or "inaccurate" summaries such as telling pregnant people to smoke 2-3 cigarettes a day and telling depressed people to jump off the Golden Gate Bridge. These are partially due to AI's inability to discern information from inputs. Until hallucinations can be solved, it will be hard to fully trust AI in times when a user does not have the skill or time to fact-check or ask the appropriate question. This indeed would take a notable amount of time and money, making the task daunting. However, AI will become more dependable and could potentially lead to mass unemployment.

The Age of Anti-ambition will encourage mass unemployment, for the Gen-Z and Gen-Alpha may no longer see the value in striving for academic success due to a shortened attention span. This is reflected in their poor performance in problem-solving and communication in the international PISA study in 2022. It shows that scores for reading and math fell by nearly 10 and 15 points respectively from 2018 to 2022, indicating that students lose

essential mathematical, reading, and science skills. The inevitable consequence of this is mass unemployment because the workforce will lack the skills to analyse.

The other factor contributing to Anti-ambition results from Big Tech company's distribution of Universal Basic Income, which leads to a decrease in people's motivation to work. In the 1970s, the government's incentive to give out UBI was to reduce poverty and boost economic growth and stability. From this standpoint, every American was entitled to \$500-1200 per month in addition to their income, regardless of their salary. Theoretically, UBI can give people the freedom to pursue their hobbies or interests, start businesses, as well as upskill and reskill without the fear of financial instability. A noteworthy experiment in Stockton, California reported improvement in mental health, greater job security, and even a significant increase in full-time employment as 125 residents received \$500 a month for 24 months.

However, UBI in the form of cryptocurrency to compensate for automation and promote their digital platforms may lead to mass unemployment. Worldcoin, co-founded by Sam Altman, is a cryptocurrency distributed in exchange for a user's iris scan. While this aims to have a global distribution of wealth, this can accelerate the trend of technological displacement as people may no longer find value in working and would rather stay home to browse online. People's proneness to seek constant entertainment without worrying about income can lead to addiction to gaming and social media. This is supported by the fact that a small dopamine hit from immediate rewards or achievements can create a powerful dependency, driving people to experience these sensations again instead of craving social mobility. The introduction of UBI from Big Tech companies could change the way people think and introduce a loss of purpose and motivation.

The current rapid advancement of automation is outpacing the creation of new jobs, potentially resulting in mass unemployment. As companies adopt these technologies and drive costs down, there simply won't be enough positions created to balance the speed of automation. Entire sectors like manufacturing and retail could become automated, so the displacement of employees could occur faster than our ability to create new ones. Statistics demonstrate that the overall unemployment rate in Canada rose to 6.2% as of May of 2024, up from 5.1% a year earlier. This increase in unemployment rate signals changes in the workforce or major adjustments in particular sectors. As automation advances speedily, the pace of job creation lags behind, causing mass displacement and economic disproportion.

Moreover, the global chip shortage will also significantly influence automation. When the chip shortage is resolved, the availability of AI products may lead to unemployment in vulnerable sectors. Currently, the main factor that contributes to chip shortage is geopolitical conflicts including sanctions on China. However, since the development of any technology essentially requires global cooperation, it is reasonable to be optimistic about the termination of their crisis in the near future. By the time when the issue of chip shortage is resolved, the production of everything from personal gadgets to industrial equipment will be greatly influenced by AI. This fundamentally alters the current job situations, for workers would no longer be heavily reliant on manual labour.

To address this risk of mass unemployment, the government should establish regulatory frameworks, international collaboration, transparency and accountability. Furthermore, the government should ensure their nation remains competitive in AI by enhancing security, and ethical standards, and contributing to the advancement of society. These aim to balance the benefits and risks of AI, making sure that “bad actors” do not threaten the route to life, liberty and property. An example of this would be the 2023 Bletchley AI Safety Summit which pooled 80 million dollars to go towards research on a new “AI for Development” program. It has started setting frameworks and regulations for the responsible development of AI. Key regulations include ethical standards, and attempting to uphold the social contract by ensuring the protection of individual rights. By fulfilling these duties, governments ensure accountability to the principles of the social contract.

Furthermore, the Global Partnership on AI (GPAI) established in 2020 is an international initiative that brings together leading AI researchers from countries around the world such as Canada, France, Germany, the UK, and the USA to promote responsible AI development in policy and awareness. It also promotes workforce reskilling that focuses on helping workers gain the necessary for the adapting job market. As John Lock states, individuals’ right to livelihood cannot be impoverished by the government, for the government’s power is based on the mutual consent between it and its citizens. If the government allows Big Tech to deploy AI without accountability and make unethical decisions, that would violate the idea of governmental power coming from the approval of its citizens. Upholding the social contract during the advancements of AI protects individual rights, allowing people to be better off than in “the state of nature”.

The government should focus on monitoring companies to increase transparency towards civilians and employees. Specifically, upholding AI ethics under John Locke’s social contract requires respect for some fundamental rights such as privacy and non-discrimination. Individuals enter into civil society and consent to be governed because they wish to protect their natural inalienable rights. In the 2019 controversy of Apple’s Siri, their contractors are allowed to listen to a small number of conversations to improve their AI. This lack of transparency sparked public outcry because it trespasses into the private sphere of people’s lives. If the government decides to ignore the potential side effects of AI interfering with privacy, it would be a direct violation of the social contract.

The government should also increase transparency to establish clear boundaries for respecting privacy. Take the example of Worldcoin which aims to distribute wealth to everyone despite their background. The Privacy Act of Canada is key in this instance as it aims to protect the privacy of individuals concerning personal information about themselves. Scanning irises may raise significant concerns as it involves collecting sensitive data without clear consent mechanisms nor safeguards against misuse. Here, it should be the government’s responsibility to safeguard personal information.

In this rapidly evolving market of AI, companies often compete for their share of technological advancements. However, there always remains a critical need for common rules and guidelines. While these businesses may be competitors, they should prioritize these common

rules in AI algorithms, decision-making processes, and respect for privacy rights. In this context, businesses agreeing to adhere to common rules reflects a commitment to respecting these rights. In 2017, attended by prominent figures including Nick Bostrom and Elon Musk, industry leaders and AI experts gathered at the Asilomar Conference Grounds in California and developed 23 principles to guide the future of AI, thereby establishing a clear guideline of what tasks can and cannot be completed by AI.

Establishing this can potentially solve unemployment, as it demonstrates that there are almost always tasks that cannot be done with AI. The logic of AI points to the fact that an individual must give their private information into the wide dataset to facilitate AI to generate the best solution, yet that would fail to keep that individual's privacy. Many people's reluctance to share their personal information highlights the irreplaceability of human labour. For example, it is common for people to find personal mentors and therapists to discuss questions they have in life. By doing so, people wish to have personal advice, and so they are unwilling to share their privacy for exposure. Therefore, these roles are not in a position to be replaced by technology that risks privacy exposure.

What this example shows is that, once we establish a clear boundary of what AI should and should not interfere with or what it should and should not see, it becomes easy to see that there are always positions that cannot and should not be replaced by AI.

Ultimately, as we navigate the ever-changing world of AI and technology, it is imperative to ensure all the fundamentals and ethical boundaries are in place, taking measures needed to mitigate the effects of automation and safeguard civil rights, ultimately fostering a balanced and equitable future where humans and technology can safely coexist.

Works Cited

- Myers, Joe. "OECD PISA results: Maths and reading skills in 'unprecedented drop'. Here's why that matters" World Economic Forum (Dec 21 2023)
<https://www.weforum.org/agenda/2023/12/oecd-pisa-results-maths-reading-skills-education/>
- "The truth behind Olena Zelenska's \$1.1m Cartier haul". The Economist (May 4 2024)
<https://www.economist.com/interactive/science-and-technology/2024/05/01/the-truth-behind-olena-zelenskas-cartier-haul>
- Ross, Casey Ike, Swetlitz. "IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show". Stat (July 25 2018)
<https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/>
- "Labour Force Survey, May 2024". Statistics Canada (June 7 2024)
<https://www150.statcan.gc.ca/n1/daily-quotidien/240607/dq240607a-eng.htm>
- "Fact Sheet: Treasury Actions to Enhance Financial Transparency and Combat Illicit Finance". U.S. Department Of The Treasury (Feb 14 2024)
(<https://home.treasury.gov/news/press-releases/jy2097>)
- "World leaders, top AI companies set out plan for safety testing of frontier as first global AI Safety Summit concludes". GOV.UK (Nov 2 2024)
<https://www.gov.uk/government/news/world-leaders-top-ai-companies-set-out-plan-for-safety-testing-of-frontier-as-first-global-ai-safety-summit-concludes>
- "G7 Ministers Reaffirm Commitment to Advancing Generative AI Policy Through Multilateral Collaboration, including GPAI" GPAI (Sep 7 2023) <https://www.gpai.ai>
- Haselton, Todd. "Apple apologizes for listening to Siri conversations" CNBC (Aug 28 2019)
<https://www.cnn.com/2019/08/28/apple-apologizes-for-listening-to-siri-conversations.html>
- Wolford, Ben. "What is GDPR, the EU's new data protection law?" GDPR.EU
<https://gdpr.eu/what-is-gdpr/>
- Hart, Robert. "Google Restricts AI Search Tool After 'Nonsensical' Answers Told People To Eat Rocks And Put Glue On Pizza" (May 31 2024)
<https://www.forbes.com/sites/roberthart/2024/05/31/google-restricts-ai-search-tool-after-nonsensical-answers-told-people-to-eat-rocks-and-put-glue-on-pizza/>
- Huttenlocher, Daniel. "The Age of AI: And Our Human Future" Little, Brown and Company (Sept 14 2021)
- McMorrow, Ryan Nian, liu Qianer, liu Eleanor, Olcott. "Baidu's bet on AI could make or break China's fallen tech group" Financial Times (Jan 4 2024)
<https://www.ft.com/content/94ea3e9b-fc2e-4454-bc41-62b9b29ccd47>
- Murgia, Madhumita, Anne-Sylvaine C. "Mistral becomes the talk of Davos as business leader seeks AI gains" Financial Times (Jan 19 2024)
<https://www.ft.com/content/9e7ca55c-6987-4064-a181-47905eeb4662>

- Mance, Henry. "AI keeps going wrong. What if it can't be fixed?" Financial Times (Apr 5 2024)
<https://www.ft.com/content/648228e7-11eb-4e1a-b0d5-e65a638e6135>
- J. Kamps, Haje. "Nvidia's Jensen Huang says AI hallucinations are solvable, artificial general Intelligence is 5 years away" Tech Crunch (Mar 19, 2024)
<https://techcrunch.com/2024/03/19/agi-and-hallucinations/>
- Malone, Noreen. "The Age of Anti-ambition". The New York Times Magazine (Feb 15 2022)
<https://www.nytimes.com/2022/02/15/magazine/anti-ambition-age.html>
- Morris, Stephen and Madhumita Murgia. "Google's AI search tool tells users to 'eat rocks' for your health" The Financial Times (May 24 2024)
<https://www.ft.com/content/13b5b637-f2bb-4208-bed4-2fa760adfb7f>
- M. Solow, Robert. "Problems That Don't Worry Me". Massachusetts Institute of Technology (July, 1962) 52 Years Ago | MIT Technology Review
- S. Smith, Craig. "Is A.I. the Future of Test Prep?". The New York Times (Dec 27 2022)
<https://www.nytimes.com/2022/12/27/business/ai-education-app-riiid.html>
- "What will humans do if technology solves everything?". The Economist (Apr 9 2024)
<https://www.economist.com/finance-and-economics/2024/04/09/what-will-humans-do-if-technology-solves-everything>
- M. Keynes, John. "Economic Possibilities for our Grandchildren" Yale (1930)
<http://www.econ.yale.edu/smith/econ116a/keynes1.pdf>
- "New research shows the robots are coming for jobs—but stealthily" . The Economist (Jan 16th 2021)
<https://www.economist.com/finance-and-economics/2021/01/16/new-research-shows-the-robots-are-coming-for-jobs-but-stealthily>
- Jacobs, Emma. "The rise of the chief AI officer". The Financial Times (Apr 13 2024)
<https://www.ft.com/content/4c5d823e-004b-4f00-9536-9fea93330453>
- Tegmark, Max. "Life 3.0 Being Human In The Age Of Artificial Intelligence" Knopf (Aug 23 2017)
- "Asilomar AI Principles" Future of Life Institute (Aug 11 2017)
<https://futureoflife.org/open-letter/ai-principles/>

Making Classical Music Relevant: Applying Cultural Entrepreneurship to Contemporary Innovations in Classical Music Performance By Jiayi Sun

Abstract

This paper explores the role of cultural entrepreneurship in revitalizing classical music by integrating innovation and creativity to expand its appeal among modern audiences. Using observational qualitative data from contemporary classical music performances and marketing strategies, this research identifies three entrepreneurial approaches for modernizing the genre: fusing information technology into classical music settings, incorporating the comedy genre into classical music performance, and combining multiple music sub-genres. Through case studies, the paper demonstrates that adopting a more inclusive and interactive model can help classical musicians engage a broader audience, deepen the public's appreciation of its heritage, and strengthen the connection between the art form and its listeners.

Introduction

On April 1st, 2014, at London's Royal Festival Hall, the audience experienced classical music in a dramatically new way, as seen in one of YouTube's most popular classical music videos. Through an unconventional blend of musical virtuosity and comedic flair, the irreverent and uproarious Rainer Hersch took the classical music world by storm, exemplifying how innovation can transform traditional art forms. His YouTube videos, which have amassed over thirty million views, showcased his unique ability to infuse humor into classical performances. With a staggering number of viewers, ranging from classical music enthusiasts to comedy lovers, Hersch's unique performances obviously struck a chord with audiences worldwide. Comments on his videos often highlight how Hersch's humor breathes new life into the genre, making it both enjoyable and accessible without sacrificing musical integrity.

Classical music, with its rich tradition and timeless repertoire, holds a revered place in the world of music. From the mid-18th to the early 19th centuries, performance was governed by entrenched norms and structures that were faithful to the composer's original intentions. The performer was expected to be a faithful servant of the composer, with this devotion determining everything from interpretation to the conception of the musical work. Classical music scores typically include precise directions regarding dynamics, articulation, and tempo, which musicians are expected to follow closely. However, in the contemporary era marked by rapid societal and cultural changes, these conventions are increasingly being questioned. The meaning and accessibility of classical music are facing new challenges, calling for innovation to keep the genre relevant. A growing body of research and debate in the field of cultural entrepreneurship is required to support the autonomy, creativity, and innovation of performers by advocating for new ways to engage diverse audiences and breathe new life into this traditional art form.

This research paper delves into how contemporary innovative approaches in cultural entrepreneurship can rejuvenate classical music performance. By embracing a more inclusive and interactive model, classical music can attract a broader audience, encourage a more profound

appreciation of its rich heritage, and foster a deeper connection between the art form and its audience. Key research questions include: How can classical music performance evolve to incorporate greater performer agency, creativity, and innovation? How can it stay relevant to a diverse audience? How can it challenge established norms and acknowledge performers as vital contributors to musical creation and entrepreneurship? By examining case studies of artists who have pioneered changes in this field, the paper highlights three main strategies for revitalizing classical music: (1) fusing information technology into classical music settings, (2) incorporating the comedy genre into classical music performance, and (3) combining multiple music sub-genres. To stay pertinent and captivating in today's rapidly evolving cultural landscape, classical music must adapt by recognizing performers as co-creators, fostering diversity and innovation, and utilizing modern technological advancements.

Background and Context

Classical music, a genre with roots stretching back to ancient civilizations, evolved significantly through the centuries and reached a revered status. The early foundations of classical music trace back to ancient Greece and Rome, where music played a crucial role in religious and social rituals. However, it wasn't until the medieval period that Western classical music began to take shape with the development of written notation. This era saw the emergence of Gregorian chant, which laid the groundwork for more complex forms of music. The Renaissance period brought about significant changes, with a focus on harmony and the expansion of musical forms. This evolution paved the way for the Baroque era's intricate compositions and the birth of opera. During the Classical era, composers like Joseph Haydn, Wolfgang Amadeus Mozart, Ludwig van Beethoven refined and formalized musical structures, emphasizing clarity, balance, and form. This period marked the transition from the ornate Baroque style to a more accessible, elegant style. The Classical era's influence extended into the Romantic period, characterized by emotional expression and individualism. Composers such as Franz Schubert, Frédéric Chopin, Richard Wagner pushed the boundaries of classical music, reflecting societal changes and personal sentiments through their works. The Romantic era's innovations expanded the possibilities for a diverse range of styles and genres.

Classical music innovation faces modern challenges in several aspects. Classical music, historically seen as a masterpiece of high culture, often reaches a limited audience due to its association with a rigid and dogmatic environment rooted in historical standards and traditions of “work-concept.” This perception is rooted in long-standing conventions and expectations. For centuries, the genre has been closely tied to Western European traditions and is often seen as catering to a specific demographic: people with a certain level of cultural capital, knowledge, or economic power. These traditions dictate the way classical music is performed, interpreted, and experienced, which can make it seem overly formal and exclusionary. The focus on strict adherence to established norms—such as the formality of concert settings, specific dress codes, and the serious demeanor expected of both performers and audiences—reinforces stereotypes that classical music is inaccessible and elitist. As a result, classical music often appears

incompatible with modern, freedom-seeking values, which prioritize openness, experimentation, and inclusivity. The rise of popular music genres and changing cultural values have challenged classical music's position, leading to debates about its relevance and accessibility. At the same time, the field has seen declining ticket sales, shrinking donor bases, and limited funding opportunities. Without new strategic innovation to attract diverse audiences and engage with contemporary culture, classical music risks losing its position as a vital and living art form. Another significant challenge facing classical music is its reliance on a single, traditional channel of communication: the concert hall. While live performances are an integral aspect of classical music, they limit the genre's reach compared to the diverse media platforms utilized by today's popular music, such as television, video streaming, live broadcasts, and social media. These modern channels provide instant access to a global audience, while classical music remains confined to formal and often geographically limited settings. Furthermore, in a traditional concert hall setting, the audience only experiences the final performance, which represents just one aspect of classical music. Hidden from the audience, moments such as backstage preparations, rehearsals, and the creative processes of musicians often remain unseen and unexplored. Furthermore, when classical music appears in movies, it is often used passively as background music, failing to capture the audience's attention or convey the genre's depth and richness. In sum, classical music that emphasizes faithful historical interpretation leaves little room for innovative reinterpretation or contemporary expression, further alienating potential new audiences. The lack of cultural diversity can create a sense of alienation for listeners who do not see their own stories, traditions, or experiences reflected in the music. As a result, classical music has limited its appeal and ability to foster a sense of shared cultural identity. The challenge is to create a more inclusive and globally resonant musical culture that speaks to people from all walks of life.

This paper proposes that the approach of cultural entrepreneurship offers one potential path forward to counter these challenges by fostering innovation and experimentation within the field of classical music. "Cultural entrepreneurship" can be defined as the specific activity of marketing creative products with cultural or artistic value for the potential to generate financial revenues. It involves applying entrepreneurial principles to culture and art, such as identifying market opportunities, securing funding sources, and producing new forms of culture and art from a business perspective. Cultural entrepreneurs often seek to introduce new ideas, forms or experiences to engage audiences and create sustainable models for arts enterprises. Classical musicians are a subgroup of impactful cultural entrepreneurs today.

Approach 1: Fusing Information Technology into Classical Music Settings

Combining classical music with other artistic forms involves integrating visual elements, contemporary media, and various information technologies into classical music settings. This fusion seeks to create a more immersive experience for audiences by appealing to multiple senses—sight, sound, and even movement. The addition of visual projections, lighting design,

multimedia displays, and collaborations with other art forms like film, dance, or live painting enhance the performance, turning a traditional concert into a dynamic, multi-sensory event.

More specifically, visual projections during classical music performances offer an innovative way to enhance the emotional and thematic aspects of the music. By adding artistic visuals or video narratives, these projections create a multi-sensory experience that deepens the audience's connection to the music. One example is the use of large-scale video projections in performances of Gustav Holst's *The Planets*. Projections of astronomical images and visual representations of the planets are synchronized with the performance to illustrate the mood and narrative of each musical movement. For example, stormy visuals accompany "Mars, the War Bringer," while serene imagery enhances "Venus, the Bringer of Peace." These projections not only captivate the audience but also reinforce the thematic content of the music, creating a cohesive and emotionally engaging experience.

Similarly, podcast discussions on classical music performances provide an innovative way to communicate the emotional depth and thematic richness of the music. By providing detailed analysis and context, these discussions deepen people's understanding of the music and bring listeners closer to the heart of the composition. One example is the Mahler Foundation's podcast discussing Gustav Mahler's *Symphony No. 2* ("Resurrection Symphony"). In some episodes, hosts explain the origins, beginning with the initial concept of "Tottenfeier" (Funeral Ceremony) shortly after Mahler completed his *First Symphony*. Listeners gain a deeper understanding of the contrasting movements, from the solemnity of the Scherzo to the hopeful chorale that signifies a journey towards renewal. This exploration emphasizes how Mahler's musical themes interconnect, ultimately culminating in the triumphant "Resurrection" theme. Such discussions enrich appreciation for the symphony, allowing audiences to connect more profoundly with its narrative and emotional landscape.

Performing film scores live with synchronized screenings merges the emotional power of music with the visual intensity of cinema, creating a richer, more immersive experience for the audience. In this format, musicians perform the film's score in real time, matching the mood, pace, and emotion of the scenes unfolding on screen. One notable example is the live performance of John Williams' score for *Schindler's List*. In this production, the haunting solo violin melodies, performed live, deepen the emotional gravity of the film's narrative, heightening the impact of pivotal scenes. The live performance emphasizes the historical significance and emotional depth of the Holocaust, offering audiences an intensified connection between the visual and auditory elements. By experiencing the music performed live, viewers gain a deeper appreciation of how the score elevates the emotional intensity of the film. These live performances draw in both fans of the films and classical music enthusiasts, bridging the two audiences. In this case, the immersive power of live orchestration allows the audience to experience the sweeping musical themes more vividly, connecting them more directly to the film's narrative.

Using multimedia screens to display explanatory videos or live demonstrations has proven to be a highly effective method for enhancing audience understanding of classical music.

One prominent example is *The Young Person's Guide to the Orchestra* by Benjamin Britten, which was designed specifically to introduce young audiences to the different sections and instruments of an orchestra. In this work, Britten provides a musical tour of the orchestra, showcasing the unique timbres and roles of each instrument. Accompanying multimedia presentations often include visual diagrams, instrument close-ups, and performance excerpts that illustrate how each section contributes to the overall sound of the ensemble. This approach helps audiences, particularly those unfamiliar with classical music, to appreciate the complexity and beauty of orchestral compositions by providing clear, engaging explanations of the music's structure and components.

In summary, the role of musicians and organizations as cultural entrepreneurs involves embracing information technologies. The fusion of information technology into classical music performances makes it possible to collaborate across multiple artistic forms. Similarly, incorporating certain educational elements, such as multimedia presentations explaining musical techniques or instruments, helps demystify the production process of classical music and make it more accessible. These technological advancements reflect a proactive approach toward audience engagement. The entrepreneurial role of musicians and organizations also extends to exploring new economic opportunities. By adopting modern marketing strategies, such as live streaming performances or offering interactive digital content, they tap into new revenue streams and broaden their reach. This approach underscores how cultural entrepreneurship not only enriches the artistic landscape but also contributes to the economic viability of classical music.

Approach 2: Incorporating the Comedy Genre into Classical Music Performance

The purpose of incorporating humor into classical music is to make the genre more engaging, accessible, and enjoyable for contemporary audiences. By breaking down the traditional barriers of formality and perceived elitism, humor helps demystify classical music, inviting audiences to engage with classical music in a relaxed, enjoyable manner, making it more accessible for people of all ages and backgrounds. Humor incorporation can range from exaggerated expressions, incorporating theatrical gestures or slapstick comedy in live concerts, depictions of "accidental incidents" during performances, to inserting unexpected musical phrases or "noises". With these approaches, classical music performances will show that they are not just serious or academic pursuits, but also dynamic and flexible art forms that can adapt to contemporary tastes, breaking down the perceived formality and rigidity often associated with classical concerts.

Victor Borge (1909-2000) was a Danish-American comedian, conductor, and pianist, renowned for his pioneering fusion of classical music and comedy. Through humorous depictions of "accidents" that could occur during classical performances, Borge proved that classical music does not need to be performed with strict seriousness and "perfection," allowing audiences to enjoy and appreciate classical music in a new way. In his iconic "Page Turner" act, Borge humorously emphasized the awkwardness and unpredictability of live performances through comedic interactions with the person assigned to turn the pages. For example, when the

page-turner flipped pages too early, Borge would comically rush back to correct the error; when the page-turner was slow to turn the page, he would exaggerate by shaking his head, stomping his feet, and jumping up and down to signal the page-turner to act. This playful interaction added an unexpected layer of comedy to the performance, resonating with both the audience and fellow musicians. Through this clever humor, Borge turned even the smallest aspects of a classical concert into entertaining moments, endearing his performances to the public. Through these performances, Borge skillfully infused humor into classical music, proving that the genre can be vibrant, engaging, and, most importantly, enjoyable for everyone.

Subsequently, Rainer Hersch (b. 1962), a comedic pianist and entertainer, has brilliantly woven humor into classical music performances. By playfully reimagining well-known classical compositions with a comedic twist and combining them with his engaging stage persona, Hersch transforms the often-formal concert environment into one that is relaxed and enjoyable. For instance, he humorously replays a well-known segment from Gioachino Rossini's *William Tell Overture*, creating musical interruptions and comical repetitions that mimic technical errors and exaggerate the tension within the music. This technique elicits empathy from the audience for the distorted audio while achieving a humorous effect through its unexpected melody breaks. By drawing listeners into these playful moments, Hersch makes classical music more relatable, transforming it from an untouchable art form into something that resonates with everyday experiences and surprises. Hersch also turns everyday concert interludes, such as coughs and sneezes from the audience, into humorous moments, showing that even disturbances can become part of the musical experience. For example, he led the orchestra in depicting a performance interrupted by exaggerated coughs or sneezes, and then skillfully incorporated these "noises" into the music, from the violins' coughing to the violas' spitting, and ending with a humorous "shush" by the cellists. These humorous portrayals of real-life distractions break the otherwise rigid atmosphere of classical concerts, encouraging the audience to feel more at ease and to immerse themselves fully in the performance. Through such inventive and light-hearted approaches, Hersch shows that the fusion of humor and classical music can break down barriers and allow more people to experience this art form in new and engaging ways.

Aleksey Igudesman and Hyung-ki Joo, known as IGUDESMAN & JOO, are a musical comedy duo renowned for blending virtuosic performance with comedic sketches. In their performance of the Youtube video titled "Rachmaninov Had Big Hands," Igudesman and Joo exaggeratedly highlight the notorious difficulty of Sergei Rachmaninov's piano pieces, which are known for their complex fingerings and expansive stretches. The vignette humorously exaggerates the physical challenges a pianist faces when playing Rachmaninoff's work, and IGUDESMAN & JOO have created a wooden block device that helps the pianist to play spanning chords, and uses it to help play, laboriously, the notes, emphasizing the impossibility of the task in a grotesque way. Their performance cleverly blends virtuosic piano playing with slapstick humor, turning a technically challenging piece into an entertaining comedy act. By playfully mocking the intimidating reputation of Rachmaninov's music, Igudesman and Joo

demystify the complexity of classical compositions, inviting the audience to laugh and engage with the music without fear of elitism or inaccessibility.

As these examples show, classical musicians and their representative organizations act as cultural entrepreneurs by leveraging humor to innovate programming and to engage with diverse audiences. By blending the comedy genre with classical performance, these artists not only enhance audience experiences but also demonstrate the genre's capacity for growth and adaptation, ensuring its continued vitality and connection to modern audiences.

Approach 3: Combining Multiple Music Sub-genres

Combining multiple sub-genres in classical music refers to the intentional blending of diverse musical styles—such as jazz, rock, pop, and world music—with the foundational elements of classical music. This involves integrating rhythms, harmonies, instrumentation, and performance techniques from other sub-genres into classical compositions, creating a unique and hybrid music style. By merging these different musical traditions, composers expand the creative vocabulary of classical music, allowing for new expressions that go beyond the boundaries of traditional Western art music.

Such a fusion approach may involve creativity in the areas of rhythm and harmony. Leonard Bernstein's *Symphony No. 2: The Age of Anxiety* (1949) offers one powerful example of sub-genre fusion, blending classical forms with jazz idioms to explore complex psychological themes. Inspired by W. H. Auden's poem of the same name, Bernstein's symphony combines a traditional symphonic structure with a solo piano part that reflects the introspective and fragmented nature of the modern human experience. Throughout the work, Bernstein weaves jazz rhythms, bluesy harmonies, and even boogie-woogie styles into the fabric of a classical symphony, creating a piece that is both accessible and intellectually challenging. One of the standout movements, "The Masque," was written specifically for solo jazz piano and percussion, highlighting Bernstein's ability to seamlessly integrate jazz into a classical framework. This movement features jazz-like improvisations and a rhythmic, almost dance-like quality reminiscent of a jazz club performance, showcasing syncopated rhythms and percussive accents that evoke a lively, improvisatory spirit, which underscores the composer's intent to bridge different musical subgenres. Bernstein's ability to navigate between these styles creates a symphony that is fresh, inventive, and reflective of contemporary American culture.

Fusion may also happen when blending classical music with modern electronic instruments. Max Richter's *Vivaldi: The Four Seasons – Recomposed* (2012) represents a more modern approach to sub-genre fusion, reimagining Vivaldi's famous baroque concertos through the lens of minimalism and electronic music. Richter retains much of the original melodic material of Vivaldi's Four Seasons but transforms it by layering repetitive minimalist motifs, altering tempos, and adding new harmonies that reflect contemporary sensibilities. He also integrates subtle electronic elements, such as synthesized sounds and digitally altered textures, to create an innovative soundscape that blurs the boundaries between classical and modern electronic sub-genres. By maintaining the recognizable elements of Vivaldi's original

composition while simultaneously injecting minimalist and electronic influences, Richter's work appeals to both classical purists and modern listeners. This reinvention not only pays homage to the baroque tradition but also revitalizes it, demonstrating the timelessness of classical music when fused with current stylistic trends.

Fusion may occur when classical music is integrated with world music. Kevin Volans' *WhiteMan Sleeps* (1987) is a notable example of combining classical composition with traditional African rhythms and textures. The work draws on the minimalist techniques of classical music while incorporating rhythmic patterns and instrumental timbres inspired by the music of South Africa, African thumb piano and percussion, exploring rhythmic patterns and harmonic structures derived from African musical traditions. Volans combines the repetitive structures characteristic of minimalist music with the complex rhythms and tonal qualities of traditional African music. Another compelling example of the fusion of world music sub-genres is Tan Dun's *Crouching Tiger Concerto* (2000), which blends traditional Chinese music with elements of Western classical music. The *Crouching Tiger Concerto*, composed for the film score of the same name, combines traditional Chinese instruments, the erhu (a two-stringed bowed string instrument) and guqin (a seven-stringed zither), with Western orchestral music. Ultimately, these pieces bridge the gap between two very different musical traditions, demonstrating how classical music can become richer and more diverse through cross-cultural influences, and appealing to people from diverse cultural backgrounds, increasing the spread of classical music.

Composers who blend multiple sub-genres in their works often assume the role of cultural entrepreneurs, driving innovation within the classical music landscape, while maintaining artistic integrity. By borrowing the elements of rhythms, harmonies, instrumentations, and performance techniques, these composers show conscious efforts to break away from traditional boundaries and embrace a more open, experimental approach to music-making. By drawing on diverse influences, composers help transform classical music from a fixed, historical genre into a living, breathing art form that speaks directly to the hearts and minds of today's global community. This versatility not only attracts new listeners but also helps classical music retain its artistic depth while moving beyond its perceived elitism, making it more inclusive and reflective of a global culture that increasingly values cross-genre exploration.

Conclusion

This paper explores how innovative cultural entrepreneurship can revitalize classical music performance. By adopting a more inclusive and interactive model, classical music has the potential to engage a wider audience, deepen appreciation of its rich heritage, and strengthen the connection between the art form and its listeners. Through case studies of artists who have pioneered changes in the field, the paper identifies three strategies for revitalizing classical music: (1) fusing information technology into classical music settings, (2) incorporating the comedy genre into classical music performance, and (3) combining multiple music sub-genres.

Musicians and organizations, acting as cultural entrepreneurs, embrace these innovations to expand classical music's economic, social, and cultural reach. These approaches require risk-taking, a deep understanding of multiple cultures and artistic forms, and mastery of diverse musical sub-genres.

The research predominantly relies on qualitative methods, drawing on anecdotal and observational data from contemporary classical music performances and marketing strategies. While this provides valuable insights into cultural shifts within the industry, the lack of quantitative data limits the ability to precisely assess the impact on audience engagement, revenue growth, and the sustainability of classical music organizations. Additionally, the case studies focus on specific innovations, such as visual projections and live film scores, which may not represent broader trends within the global classical music landscape. Furthermore, the research assumes that integrating technology and interdisciplinary collaborations are universally beneficial, without considering the financial or logistical challenges that smaller ensembles may face. Future research should incorporate more statistical analyses of contemporary classical music performances and marketing strategies. Gathering data such as audience surveys in diverse cultural and educational contexts, along with financial performance metrics, would provide clearer insights into how these entrepreneurial efforts lead to measurable outcomes.

Works Cited

- Byseb. "IGUDESMAN & JOO - Rachmaninov Had Big Hands." YouTube, 26 Dec. 2006, youtu.be/ifKKlhYF53w?si=Bi9CVK7BUC9RZzbu. Accessed 11 Oct. 2024.
- KlassikAkzente. "Recomposed by Max Richter: Vivaldi, the Four Seasons - Doku." YouTube, 19 July 2012, youtu.be/qTapNp-31rU?si=yKxFfa8jT7mgdYAN. Accessed 11 Oct. 2024.
- New Jersey Symphony. "The Young Person's Guide to the Orchestra." [www.youtube.com](https://www.youtube.com/watch?v=rbUbx9cJPX0), 31 May 2022, www.youtube.com/watch?v=rbUbx9cJPX0.
- Barbara. "Why Is Classical Music Dying? A Fully Exploration." Our Music World, OurMusicWorld.com, 15 July 2024, www.ourmusicworld.com/archives/13385.
- Bricio, Francisco. "Listening Guide - Symphony No. 2." Mahler Foundation, 29 Mar. 2021, mahlerfoundation.org/mahler/compositions/symphony-no-2/listening-guide-symphony-no-2/.
- Brown, Steven, and Ulrik Volgsten. *Music and Manipulation*. Berghahn Books, 2005.
- Bukofzer, Manfred F. *Music in the Baroque Era: From Monteverdi to Bach*. Read Books Ltd, 2013.
- Burkholder, J. Peter, et al. *A History of Western Music*. 10th ed., W.W. Norton And Company, 2014.
- Caldwell, John. *Medieval Music*. Routledge, 2020.
- Deutsche Grammophon - DG. "John Williams & Saito Kinen Orchestra - Schindler's List (Live at Suntory Hall, 2023)." YouTube, 7 Mar. 2024, www.youtube.com/watch?v=rQIz8Calg9o. Accessed 11 Oct. 2024.
- Dobrev, Nevena, and Stanislav Hristov Ivanov. "Cultural Entrepreneurship: A Review of the Literature." *SSRN Electronic Journal*, vol. 16, no. 4, 2020, <https://doi.org/10.2139/ssrn.3824265>.
- Dotsey, Calvin. "The Age of Anxiety: Leonard Bernstein's Symphony No. 2." *Houston Symphony*, 20 Mar. 2018, houstonsymphony.org/age-anxiety-bernstein-symphony-2/.
- Freedman, Richard. *Music in the Renaissance*. W. W. Norton, Cop, 2013.
- Goehr, Lydia. *The Imaginary Museum of Musical Works : An Essay in the Philosophy of Music*. Clarendon Press, 1992.
- Hakvoort, Laurien, and Avi Gilboa. "Untangling Strings." *Journal of Trial and Error*, vol. 3, no. 2, Dec. 2023, pp. 1–4, <https://doi.org/10.36850/tegd-e423>. Accessed 11 Oct. 2024.
- Heartz, Daniel. *Music in European Capitals : The Galant Style, 1720-1780*. W.W. Norton, 2003.
- Hersch, Rainer. "About Rainer Hersch - Rainer Hersch." Rainer Hersch, 26 Nov. 2021, www.rainerhersch.com/about-2/. Accessed 11 Oct. 2024.
- Hiley, David. *Western Plainchant : A Handbook*. Clarendon Press, 1997.
- Igudesman & Joo. "About – Igudesman & Joo." igudesmanandjoo.com, 2019, igudesmanandjoo.com/about-igudesman-joo/. Accessed 11 Oct. 2024.
- Johanna. "Victor Borge - 'Page-Turner.'" Youtube, 20 Oct. 2006, youtu.be/LWqFaGwNCMU?si=-8ePfdqd4iEmF9E. Accessed 11 Oct. 2024.

- Johnson, Julian. *Who Needs Classical Music? : Cultural Choice and Musical Values*. Oxford University Press, 2002.
- Johnson, Stephen. "Discovering Music - Volans: White Man Sleeps, String Quartet No. 1 - BBC Sounds." BBC, 2014, www.bbc.co.uk/sounds/play/p020cnkk. Accessed 11 Oct. 2024.
- Leech-Wilkinson, Daniel. "Challenging Performance: Classical Music Performance Norms and How to Escape Them." *Challenging Performance*, 2020, challengingperformance.com/the-book/. Accessed 11 Oct. 2024.
- Monastra, Peggy. "Crouching Tiger Concerto | Tan Dun." www.wisemusicclassical.com, 2000, www.wisemusicclassical.com/work/33553/Crouching-Tiger-Concerto--Tan-Dun/.
- Morgan, Robert P. *Twentieth-Century Music : A History of Musical Style in Modern Europe and America*. W W Norton & Company, 1991.
- Rainer Hersch Fan Channel. "Funny! Orchestra Plays Microsoft Windows™ - the Waltz." YouTube, 12 May 2014, www.youtube.com/watch?v=dGKwx-BFO0E.
- Rosen, Charles. *The Classical Style : Haydn, Mozart, Beethoven*. Faber And Faber, 2005.
- Taruskin, Richard. *The Oxford History of Western Music: The Earliest Notations to the Sixteenth Century*. Oxford University Press, USA, 2005.
- Temperley, Nicholas. *The Athlone History of Music in Britain: The Romantic Age 1800-1914*. Athlone Press, 1981.
- The Editors of Encyclopaedia Britannica. "Victor Borge | American Comedian and Musician." *Encyclopedia Britannica*, 31 Aug. 2024, www.britannica.com/biography/Victor-Borge.
- Thornton, Sarah. *Club Cultures Music, Media and Subcultural Capital*. 1995, awareness-akademie.de/wp-content/uploads/sites/14/2021/05/Sarah_Thornton_Club_Cultures_Music_Media_and_Subcultural_Capital-2.pdf.
- Utz, Christian, and Laurence Sinclair Willis. *Musical Composition in the Context of Globalization : New Perspectives on Music History in the 20th and 21st Century*. Transcript-Verlag, 2021.
- Wyard, Adrian. "The Planets Online." Theplanetsonline.com, 2021, theplanetsonline.com/.

Early Detection & Intervention in Bipolar Disorder: Role of Neuroimaging Biomarkers

By Anika Denny

Abstract

Neuroimaging offers a window to observe brain structure and function, both of which when abnormal can be associated with psychiatric disorders like bipolar disorder (BD). In particular, with the ability to reveal abnormalities in brain activity, neuroimaging provides a promising method of identifying tell-tale biomarkers associated with certain disorders, generating extreme interest in the scientific community. Here, we review current neuroimaging techniques used to identify biomarkers for the early detection and treatment of BD, one of the top ten leading causes of disability worldwide. The prospect of early detection entails improved treatment responses and clinical outcomes in those at risk of developing BD.

Introduction

Bipolar disorder (BD) is a psychiatric condition associated with mood swings including emotional highs and lows, typically emerging in early life (Grande et al.). The repeated manic/hypomanic and depressive episodes persist throughout the lifespan. Affecting more than 1% of the world's population, BD is one of the primary causes of disability among youth, giving rise to cognitive and functional impairments as well as increased mortality. Common signs of BD include irritability and aggressiveness, sleep disturbances, depression and mania, hyperactivity, anxiety, and mood swings (Skjelstad et al.). As the onset of BD advances, the symptoms are amplified in strength and prevalence, making it a chronic and debilitating illness. BD is associated with a high suicide rate and a poor clinical and functional result (Phillips and Vieta).

Often misdiagnosed as major depressive disorder, around 60% of BD patients end up waiting five to ten years before receiving an appropriate diagnosis (Colombo et al.). The presence of mood episodes with psychotic aspects in more than 30% of cases makes the differential diagnosis between BD and other psychotic disorders more difficult (Colombo et al.). Untreated BD leads to episodes that are increased in number and severity, which may complicate effective treatment (Skjelstad et al.). Increased accuracy in the diagnosis of BD is therefore a significant aspect of improving the health and treatment of patients (Phillips and Vieta). This objective can be attained by identifying biomarkers that represent the pathophysiologic processes in BD.

Biomarkers for BD are significant because they serve as objective measures required to inform diagnosis, prognosis, and treatment decisions (Abi-Dargham et al.). A biomarker is ““a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention”” (Aylward). Biomarkers can aid in identifying the specific characteristics associated with the diagnosis of BD and illness activity related to mood episodes (Pinto et al.). Biomarkers may also help distinguish between unipolar and bipolar depression, map the progression and prognosis of

the disorder, identify at-risk individuals via trait markers, detect the illness relapse before clinical symptoms even start, and direct the development of new therapies (Muneer).

Studying neuroanatomy through imaging techniques provides significant information about how brains function, are impacted by genes and environment, and change with development and illness (Lerch et al.). Magnetic resonance imaging (MRI) is a noninvasive technique used to obtain quantitative and anatomical information from an individual in parts of the body at high resolution (Holdsworth and Bammer). Structural imaging aids in visualizing and analyzing anatomical properties of the brain, especially useful for identifying brain damage and abnormalities, such as in BD (Hirsch et al.). This neuroimaging technique can quantify structural characteristics like the size and volume of brain structures, and the thickness of a cortical area. Advantages of structural MRI include good spatial resolution and relatively easy accessibility, but it cannot be used in subjects who have metallic implants (e.g. pacemakers) and may be difficult for those uncomfortable in enclosed spaces.

Another form of MRI is diffusion-based MRI, such as diffusion tensor imaging (DTI), which provides detailed information about brain integrity, microstructure, and white matter connections. It can quantify the brain's structural integrity while also describing white matter pathways connecting different areas of the brain. Drawbacks include requiring complex image analyses and relatively long acquisition times and being sensitive to patient movement. Functional MRI (fMRI) is currently the central component of neuroimaging in cognitive neuroscience (Logothetis). It is used to identify brain regions and underlying brain processes that are related to performing a certain cognitive or behavioral task (Hirsch et al.). Inferences between the location of brain activity and brain function can be made based on the kind of signal being studied. Advantages include having good spatial resolution and being noninvasive, even though it involves exposure to high-intensity magnetic fields. However, fMRI has limited temporal resolution ability and is an indirect measure of brain activity. Nonetheless, neuroimaging technology has been crucial for discovering neurological biomarkers for BD.

Structural neuroimaging studies indicate the presence of abnormalities in the brains of bipolar disorder patients, predominantly observed in total cortical volume, cerebral white matter, cortical and prefrontal gray matter, and ventricular volumes (van der Schot et al.). Recent meta-analyses suggest that the prefrontal lobe and cortical-cognitive network components like the dorsolateral prefrontal cortex (DLPFC) are reportedly smaller in BD patients (Houenou et al.). There have been mixed findings, however, regarding ventral-limbic structures with increased amygdala volume, normal size of the hippocampus, and smaller than normal gray matter content in the orbitofrontal cortex, the subgenual cingulate cortex, and the rostral anterior cingulate cortex. Overall, the findings from functional and structural neuroimaging studies of BD appear to support the neurobiological model, and the abnormalities reported might serve as biomarkers of BD, which can aid the early detection and intervention of bipolar disorder.

The mortality rate of BD is two to three times greater than that of the general population, with around 10-20% of patients taking their own lives and almost one-third of individuals admitting to at least one suicide attempt (Müller-Oerlinghausen et al.). Delayed treatment of BD

is associated with “an impaired age-appropriate development, worse social adjustment and functional outcome, more hospitalizations, higher comorbidity rates, elevated risk of suicide, and reduced response to mood stabilizing drug treatment” (Martini et al.). Therefore, identifying at-risk individuals and early BD before the onset of the disorder and at early stages can improve the outcomes of the illness. It is expected that validated biomarkers will be useful diagnostic tools and aid in the provision of individually tailored treatment to BD patients (Muneeer). Therefore, this literature review aims to present an overview of candidate neuroimaging biomarkers for BD, focusing on structural and functional trends as well as discussing inconsistencies that hinder the process of ascertaining accurate biomarkers for diagnosing BD.

Candidate Structural Biomarkers: Cortical Thickness & Cortical Surface Area

An increasing number of studies have been conducted on the anatomical and functional aspects of brain regions via neuroimaging techniques for researching bipolar disorder (Abé, Liberg, Klahn, et al.). Cortical thickness is among the most common brain morphometric measures used in these studies. Cortical thickness is related to the size, number, and density of cells and dendrites in one cortical column, while cortical surface area is mainly determined by the number of cortical columns. Cortical thickness is related to cognitive functions and serves as a metric used to study brain-behaviour associations. A cross-sectional study found lower cortical thickness in the left and right frontal and temporal regions, insula, pre- and postcentral regions, and medial occipital lobe (including visual areas) in BD patients compared to healthy controls (Abé, Ekman, Sellgren, et al.). Longitudinal case-control studies discovered noticeable structural brain alterations along the course of bipolar disorder, with numerous studies indicating abnormalities in the prefrontal and temporal cortices, as well as the subcortical structures, specifically the amygdala (Abé, Liberg, Klahn, et al.). In a multi-center ENIGMA study involving 207 bipolar patients and 925 controls from 14 international locations, the thickness of the right fusiform gyrus and right parahippocampal regions decreased over time in healthy controls as expected, while bipolar patients showed less or no decline (Abé, Ching, Liberg, et al.). Likewise, results from a longitudinal BD study involving 123 BD patients and 83 healthy control subjects indicated abnormal cortical thinning of temporal cortices coupled with thickness increases in visual-somatosensory brain areas, of which the latter was related to genetic risk and lithium usage (Abé, Liberg, Song, et al.). The study found a significant increase in medial occipital thickness in patients who used lithium, while patients who did not use lithium showed no change in cortical thickness. Multiple regression analyses showed that lithium usage significantly and uniquely predicts changes in thickness in the medial occipital and central gyrus. However, it can also be speculated that thickness increases in sensory regions could be part of the natural course of BD. Additionally, the largest study on BD to date, a cross-sectional study involving 6503 individuals, including 1837 unrelated adults with BD and 2582 unrelated healthy adults, discovered the most significant cortical thickness abnormalities in the pars opercularis, rostral middle frontal cortex (rMFC), and fusiform gyrus, but no abnormalities in surface area (Hibar, Westlye, Doan, et al.). This study evaluated patients with varying durations of illness,

finding that reduced cortical thickness is associated with longer duration of illness in adults with BD in the occipital cortex, left parietal cortex, and right frontal cortex. BD groups in other studies showed significantly thinner cortical thickness in the left inferior temporal cortices (ITC) compared to the controls (Niu et al.). The ITC belongs to the ventral stream of visual processing, associated with representing features of complex objects, including face perception and number recognition. Additionally, less cortical thickness was found in the dorsal prefrontal cortex (PFC), including the superior frontal cortex (SFC), pars triangularis (PTRI), and rMFC. Cortical thinning was observed in the ventral prefrontal cortex (VPFC) as well, including the frontal pole (FPO) and lateral orbitofrontal cortex (IOFC). Reduced glial cell density and neuronal size in the caudal orbitofrontal cortex (OFC) of BD patients suggested that tissue loss in the OFC is characteristic of BD. BD patients reportedly had a reduced CT in the left FPO, a region in which the CT was significantly positively correlated with the onset age; early disorder-onset might be related to more significant alterations in cortical morphology in the FPO relative to late disorder-onset. Overall, large-scale analysis indicates a widespread reduction of cortical thickness in manifest bipolar disorder.

Significant relationships have been found between regional cortical surface area and neurocognitive test scores (Hartberg et al.). A larger cortical surface area in the left inferior temporal region is related to better processing speed performance in BD groups studied, while a larger cortical surface area in the right caudal middle frontal and left fusiform regions was significantly associated with poorer verbal learning in healthy control groups. Another study found no significant differences in cortical surface area regions of interest for the sex-by-diagnosis but did find evidence of an age-by-diagnosis interaction depicting reduced surface area of the left posterior cingulate cortex as age increased (Hibar, Westlye, Doan, et al.). Surface area abnormalities associated with BD are mixed, with some studies failing to identify differences in surface area between BD and healthy controls. Largely, patients with BD reportedly had lower cortical surface area than healthy controls (Abé, Ekman, Sellgren, et al.). Regarding effects associated with commonly prescribed medications, studies found an increased surface area in the left paracentral lobule in patients taking lithium, reduced cortical surface area in those taking atypical antipsychotics in the right rostral middle frontal gyrus and right superior frontal gyrus, and increased cortical surface area in those taking typical antipsychotics in the left middle temporal gyrus, left inferior parietal gyrus, and right temporal pole (Hibar, Westlye, Doan, et al.).

Candidate Structural Biomarkers: White Matter Abnormalities & Gray Matter Volume

Alongside volume and thickness abnormalities, white matter abnormalities and gray matter volume also serve as key biomarkers for BD. White matter, found throughout the central nervous system, makes up around 50% of human brain volume, playing a vital role in subserving cognition and emotion (Filley). The most important neuroanatomical feature of white matter is myelin, the fatty insulation coating axons in the brain and increasing neuronal conduction velocity. Myelin is primarily what allows white matter to significantly accelerate the transfer of

information in the brain. White matter deficits in the prefrontal, frontotemporal, and cingulate regions are consistently found in BD studies (Teixeira et al.). Because white matter reorganization might contribute to mood dysregulation, inability to control speech, and issues with attention and memory, all of which are characteristic of BD, the integrity of the white matter tracts connecting the network of frontal, striatal, limbic, and parietal brain regions is of great interest (Manelis et al.).

The most noteworthy novel technique for studying white matter-behavior relationships is diffusion tensor imaging (DTI) (Filley). DTI is an advanced, noninvasive, magnetic resonance imaging technique that measures the direction and magnitude of water diffusion along the white matter tracts in the brain, enabling the visualization of neural pathways and connectivity (Ranzenberger et al.). FA, derived from DTI tensor imaging, is a quantitative biomarker of white matter integrity and decreased FA in the major pathways is generally associated with neurodegenerative processes (Figley et al.). FA is a highly sensitive measure for microstructural changes in the brain, reflecting the diameter and density of fibers. Through DTI, white matter reductions have been indicated in the anterior corpus callosum, anterior cingulum, and the uncinate and superior longitudinal fasciculus (Teixeira et al.). These white matter tract abnormalities were found in areas that control emotions, including the limbic-striatal, cingulate, thalamus, corpus callosum, and prefrontal regions, in both pediatric and adult BD. Postmortem studies have revealed a decrease in glial cells, a key cell type found in white matter, in the subgenual prefrontal cortex, and T2-weighted MRI evidence has shown BD patients to have increased white matter hyperintensities compared to healthy controls (Sussmann et al.). A large DTI analysis found no association “between white matter integrity and any class of medication, measures of current or lifetime psychotic symptoms, or mania rating scale scores in either patient group” (Sussmann et al.). Depressive symptoms were reportedly negatively associated with anterior thalamic radiation (ATR)/callosal fractional anisotropy (FA) in the BD group, however, which provided evidence that reductions in the anterior limb of the internal capsule (ALIC) and its associated tracts might play a role in the pathophysiology of affective symptoms.

A decrease in gray matter in the bilateral superior frontal gyri (SFG), left rostral anterior cingulate cortex, and right fronto-insular cortex in patients with multiple episodes has been consistently identified (Muneer). These limbic areas are associated with decision-making and emotion processing, both of which are extremely impacted in BD. A meta-analysis of cross-sectional studies of gray matter volume alterations in patients with BD found that patients in the euthymic phase, or a period of emotional stability without mood disturbances, had smaller gray matter volumes than controls in the right inferior parietal gyri and the right thalamus, as well as larger volumes in the left insula (X. Wang et al.). Despite hypotheses that measures of brain volume in BD patients might be impacted by heterogeneity in variables like age, illness duration, mood state, and psychotropic medication use. Studies found no significant associations between gray matter alterations and the aforementioned features (Hallahan et al.; X. Wang et al.; Wise et al.). Those in the depressive phase reportedly had smaller gray matter volumes in a large cluster in the right insula, extending into the right superior temporal gyrus (STG), and in the right

median cingulate/paracingulate gyri. The STG is a key region in speech and phonological processing, which means it may play a role in the genesis of neurocognitive deficits in patients with BD (X. Wang et al.).

Patients with psychotic symptoms appeared to have significantly smaller gray matter volumes in several regions, with no regions having larger gray matter volumes compared to healthy controls (X. Wang et al.). Studies found that the percentage of patients with BD taking lithium was positively associated with gray matter volumes in the right insula and the left anterior cingulate/paracingulate gyri (ACPG), while the percentage of patients taking antipsychotic medications was negatively associated with gray matter volume in the left ACPG. Those taking lithium had larger gray matter volume in the aforementioned regions compared to healthy controls, while BD patients taking antipsychotic medications found greater volume deficits relative to healthy controls. The influence of medication on changes in gray matter volume is controversial, however. A follow-up study of patients with BD found that insula volume was not impacted by lithium or valproate; there was no evidence suggesting a relationship between antidepressants, antipsychotics, or mood stabilizers and gray matter volume alterations in the anterior insula (Takahashi et al.; X. Wang et al.). To a large extent, various studies suggest gray matter abnormalities in BD patients primarily in the prefrontal and temporal cortex, the ACC, and the insula, with the patients' mood states possibly influencing the gray matter changes. In summary, white and gray matter abnormalities, especially in the prefrontal, temporal, and limbic regions, are key biomarkers in BD, providing evidence for the neurobiological foundations of mood dysregulation, cognitive impairments, and the effects of psychotropic medications.

Candidate Structural Biomarkers: Ventricular Volumes

Enlarged ventricles have also been observed in BD, with patients showing an accelerated increase in ventricular volumes compared to healthy controls (Abé, Liberg, Klahn, et al.; Hibar, Westlye, van Erp, et al.; McDonald et al.; Woods et al.; Roy et al.; Lim et al.). MRI has been used to measure the volumes of the lateral and third ventricles and periventricular structures, including the caudate, putamen, thalamus, and hippocampus (Strakowski et al.). In this study, patients with DSM-IV BD, 18 who were having a first episode and 17 with multiple episodes, were observed relative to 32 healthy controls. The lateral ventricles were larger in patients with multiple-episode BD than in the first-episode patients or the healthy controls. Larger lateral ventricles were correlated with a higher number of prior manic episodes. The first-episode BD patients had significantly greater putamen volumes than the healthy subjects, but the hippocampal volume did not differ between the multiple-episode patients and the healthy subjects. Patients with ventriculomegaly, or enlarged ventricles, at illness onset have a worse illness course. Multiple-episode patients had a mean lateral ventricular volume 122% greater than that of first-episode BD patients with an 8-year difference in duration of the illness in this study; these findings estimate a change of 15% per year. Such findings indicate that prospective

studies of neuroanatomic alterations along the course of BD are practically possible and may aid in the clarification of the functional neuroanatomy of BD.

Table 1: Overview of the main studies reviewed that identified structural abnormalities in the brain in patients with bipolar disorder.

Reference	Biomarker	Method	Main Result	Brain Area Affected
Abé et al., 2016	Cortical thickness	MRI	Decreased cortical thickness in BD patients	Frontal, temporal, insular, pre-/post-central, and medial occipital regions
Abé et al., 2022	Cortical thickness	MRI	Decreased thickness in fusiform gyrus and parahippocampal regions	Fusiform gyrus, parahippocampal regions
Abé et al., 2020	Cortical thickness	MRI	Abnormal cortical thinning in temporal cortices; thickness increases in visual-somatosensory areas	Temporal cortices, visual-somatosensory areas
Niu et al., 2017	Cortical thickness	MRI	Reduced cortical thickness in BD patients	Pars opercularis, rostral middle frontal, fusiform gyrus
Hibar et al., 2018	Cortical surface area	MRI	Reduced surface area in the left posterior cingulate cortex	Left posterior cingulate cortex
Teixeira et al., 2019	White matter abnormalities	DTI	White matter reductions in BD patients	Anterior corpus callosum, anterior cingulum, uncinate and superior longitudinal fasciculus
Sussmann et al., 2009	White matter abnormalities	DTI	Decreased fractional anisotropy (FA) in major pathways	Anterior thalamic radiation (ATR)/callosal FA
Muneer, 2020	Gray matter volume	MRI	Decreased gray matter volume in BD patients	Bilateral superior frontal gyri (SFG), left rostral anterior cingulate cortex, right fronto-insular cortex
X. Wang et al., 2019	Gray matter volume	MRI	Gray matter volume alterations in BD patients depending on mood state	Right inferior parietal gyri, right thalamus, left insula
Strakowski et al., 2002	Ventricular volumes	MRI	Enlarged lateral ventricles in BD patients	Lateral ventricles

Candidate Functional Biomarkers: Resting State

Various studies have been conducted to investigate the brain's functional connectivity networks in a resting state in patients with psychiatric disorders like BD (Vargas et al.). Patients are not assigned a cognitive task and are told to rest quietly but awake with their eyes closed during the fMRI scan. fMRI studies have discovered abnormal resting-state functional connectivity (rsFC) in the default-mode network (DMN), the frontoparietal network (FPN), the salience network (SN), and the affective network (AN) or limbic network in patients with BD compared to healthy controls (Y. Wang et al.). Mood played a role in such abnormalities, however. Although abnormal functional connectivity in the DMN is often reported during acute episodes in BD patients, for example, it is not apparent during remission, which is when BD patients have minimal or no symptoms of mania or depression. Wang et al. conducted a meta-analysis of resting-state functional connectivity found hyperconnectivity, or increased functional connectivity between different regions of the brain, between the affective network (AN) seeds and regions of the ventromedial prefrontal cortex (VMPFC) and the dorsal medial prefrontal cortex (DMPFC) in the DMN in BD patients experiencing acute episodes and those in remission compared to HCs. Hypoconnectivity, or a loss of functional connectivity, was identified between the AN seeds and areas of the supplementary motor area (SMA) in the SMN in the BDR (BD, remission) group.

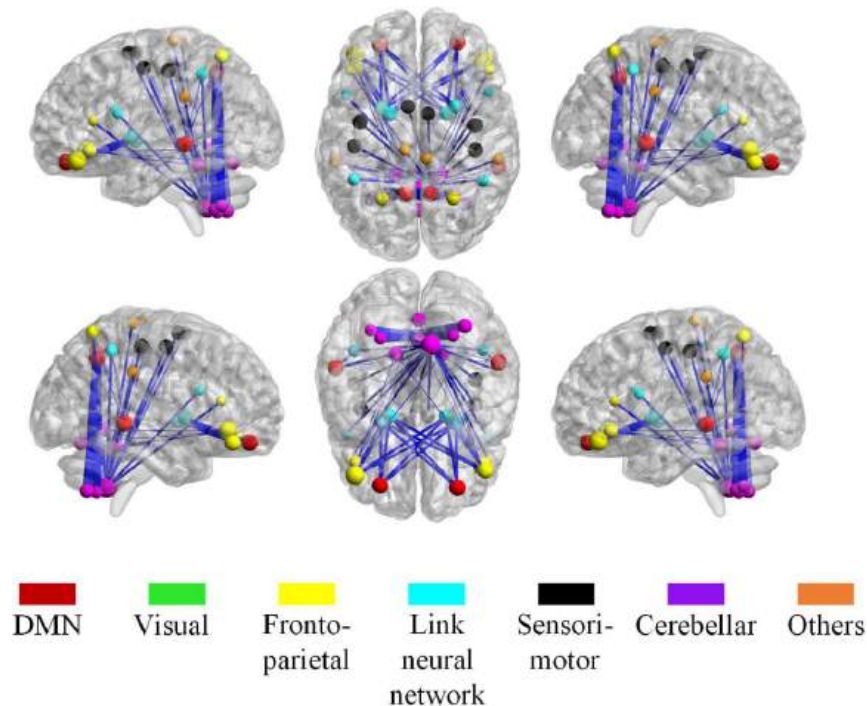


Fig 1: Brain networks and nodes in the whole brain view (Pan et al.).

Candidate Functional Biomarkers: Task-Based Activity

Facial expression identification tasks, which are key social signals of emotional states, are very commonly used in the examination of emotion-processing capabilities in healthy and

psychiatric populations (Phillips and Vieta). Emotional scenes, words, and material are presented to subjects in various sensory methods. Mood instability, related to hindered processing of emotional information, is a common symptom of BD and has been tested through displays of different facial expressions. Studies examining neurological responses to emotional stimuli have found increased amygdala and ventral striatal activity to mild happy and intense fearful expressions in remitted individuals with BD, as well as increased amygdala activity to happy expressions in a mixed group of remitted and unwell individuals with BD. DLPFC activity also reduced in remitted BD patients compared to healthy controls when they were faced with fearful expressions. There was no significant association found between subsyndromal depression severity and amygdala responses to happy and fearful facial expressions in remitted BD subjects. However, other studies conducted have indicated widespread reductions in prefrontal cortical and subcortical neural activity to emotional words in remitted individuals, which insinuates the possibility that emotional facial expressions are processed as particularly significant. Additionally, impaired cognitive control and executive dysfunction are other BD symptoms, alluding to dysfunction in a lateral prefrontal cortical system containing the DLPFC and ventrolateral prefrontal cortex (VLPFC), and the hippocampus (Phillips and Vieta). The digit-sorting task, reliably associated with DLPFC activity in healthy individuals, involves sorting digits into numerical order and memorizing the digit with the middle value. Other attentional tasks, like the Stroop interference task, in which subjects identify the color of the word presented while ignoring the meaning of the word itself, have also supported evidence pointing to the DLPFC, dorsal regions of the anterior cingulate gyrus, and VPFC during task performance in healthy individuals. While performing such attentional tasks, remitted, euthymic BD patients reportedly had reduced activity in dorsal and ventral prefrontal cortical regions and within dorsal regions of the anterior cingulate gyrus. They had increased DLPFC activity during a Stroop interference task, however. Various other studies identified reduced DLPFC activity in euthymic individuals during working memory and verbal encoding tasks as well. While completing a continuous performance task and a working memory task, remitted, euthymic BD patients had increased activity within subcortical regions associated with emotion processing instead of working memory or attention. Decreased absolute prefrontal cortical, increased subcortical metabolism, and depression severity have been found during sustained attention tasks in bipolar depressed patients compared to healthy controls. Thus, despite certain inconsistencies, studies have largely identified increased amygdala and subcortical activity but decreased DLPFC activity during emotion-processing and cognitive control tasks in BD.

Table 2: Overview of the main studies reviewed that identified functional candidate biomarkers in patients with bipolar disorder.

Reference	Biomarker	Method	Main Result	Brain Area Affected
Y. Wang et al., 2020	Resting-state functional connectivity (rsFC)	fMRI	Abnormal rsFC in DMN, FPN, SN, and AN in BD patients	Default-mode network (DMN), frontoparietal network (FPN), salience network (SN), affective network (AN)
Phillips & Vieta, 2007	Task-based activity	fMRI	Increased amygdala and ventral striatal activity in BD patients	Amygdala, ventral striatum
			Reduced DLPFC activity in BD patients	DLPFC

Limitations & Future Directions

Despite the improvements in technology and neuroimaging findings associated with bipolar disorder, there is still a long way to go to accurately detect and treat BD early. BD involves heterogeneous symptomatology and various genetic and environmental factors, greatly complicating its diagnosis (Scaini et al.). BD also overlaps with many other disorders in terms of pathophysiology; the detection of a clear boundary between these disorders will aid in BD diagnosis and the identification of biological targets for the development of new treatments. Additionally, the overall variability in results in the literature hinders researchers' ability to establish the generalizability of any recent findings. However, this present literature review highlights commonalities in specific brain alterations across numerous studies that focus on a wide range of biomarkers, including abnormalities in gray and white matter, cortical surface area, cortical thickness, ventricular volumes, and functional connectivity.

To overcome such limitations, machine learning (ML) and artificial intelligence (AI) methodologies have been used in scientific literature to differentiate BD from healthy controls and other neurological disorders (Colombo et al.). AI mainly develops the algorithms and methods used to determine whether a system's behavior is accurate in disease diagnosis (Ahsan et al.). As opposed to the time-consuming, costly nature of traditional diagnosis processes, ML-based systems are inexpensive and time-efficient. The application of ML to objective markers might overcome prognosis uncertainty because of the subjectivity involved in diagnosing BD (Colombo et al.). Results from systematic reviews and meta-analyses about ML indicate that researchers must pick the best ML algorithm based on a specific type of marker. It has been found that ML models can diagnose BD through clinical and nonclinical data (Jan et

al.). However, future research should better evaluate the performance of ML models beyond common metrics like accuracy and compare them with clinical techniques for diagnosing BD, such as the Affective Disorder Evaluation (ADE) scale and Structured Clinical Interview for DSM-IV. Finally, researchers should study the effectiveness of diagnosing models in clinical settings and develop models that can predict disorders like BD.

Conclusion

In this literature review, an effort has been made to highlight the latest findings related to the detection of candidate neuroimaging biomarkers for bipolar disorder, which is a severe psychiatric condition. The results indicate that BD patients experience a reduction in cortical thickness and lower cortical surface area compared to healthy controls. Evidence also suggests that BD is associated with white matter deficits in various regions, possibly contributing to mood dysregulation, inability to control speech, and issues with attention and memory. Gray matter abnormalities in BD patients may be influenced by the patients' mood states and medication usage, with some reporting smaller gray matter volumes and others having larger volumes compared to healthy controls. Studies found enlarged ventricular volumes in BD patients as well, who consequently experience worse illness courses. Regarding functional biomarkers, abnormal functional connectivity in the DMN has been reported during acute episodes in BD patients and not during remission. During cognitive control and emotional processing tasks, however, BD patients have reportedly increased amygdala and subcortical activity, with decreased DLPFC activity. With the proliferation of modern advanced computational methods, these biomarkers may help improve diagnostic accuracy in individuals with earlier stages of BD, aiding in the timely intervention of this disorder.

Works Cited

- Abé, Christoph, Carl-Johan Ekman, Carl Sellgren, et al. "Cortical Thickness, Volume and Surface Area in Patients with Bipolar Disorder Types I and II." *Journal of Psychiatry and Neuroscience*, vol. 41, no. 4, July 2016, pp. 240–50. DOI.org (Crossref), <https://doi.org/10.1503/jpn.150093>.
- Abé, Christoph, Benny Liberg, Jie Song, et al. "Longitudinal Cortical Thickness Changes in Bipolar Disorder and the Relationship to Genetic Risk, Mania, and Lithium Use." *Biological Psychiatry*, vol. 87, no. 3, Feb. 2020, pp. 271–81. www.biologicalpsychiatryjournal.com, <https://doi.org/10.1016/j.biopsych.2019.08.015>.
- Abé, Christoph, Christopher R. K. Ching, Benny Liberg, et al. "Longitudinal Structural Brain Changes in Bipolar Disorder: A Multicenter Neuroimaging Study of 1232 Individuals by the ENIGMA Bipolar Disorder Working Group." *Biological Psychiatry*, vol. 91, no. 6, Mar. 2022, pp. 582–92. www.biologicalpsychiatryjournal.com, <https://doi.org/10.1016/j.biopsych.2021.09.008>.
- Abé, Christoph, Benny Liberg, Anna Luisa Klahn, et al. "Mania-Related Effects on Structural Brain Changes in Bipolar Disorder – a Narrative Review of the Evidence." *Molecular Psychiatry*, vol. 28, no. 7, July 2023, pp. 2674–82. www.nature.com, <https://doi.org/10.1038/s41380-023-02073-4>.
- Abi-Dargham, Anissa, et al. "Candidate Biomarkers in Psychiatric Disorders: State of the Field." *World Psychiatry*, vol. 22, no. 2, 2023, pp. 236–62. *Wiley Online Library*, <https://doi.org/10.1002/wps.21078>.
- Ahsan, Md Manjurul, et al. "Machine-Learning-Based Disease Diagnosis: A Comprehensive Review." *Healthcare*, vol. 10, no. 3, Mar. 2022, p. 541. *PubMed Central*, <https://doi.org/10.3390/healthcare10030541>.
- Aylward, Elizabeth H. "Change in MRI Striatal Volumes as a Biomarker in Preclinical Huntington's Disease." *Brain Research Bulletin*, vol. 72, no. 2, Apr. 2007, pp. 152–58. *ScienceDirect*, <https://doi.org/10.1016/j.brainresbull.2006.10.028>.
- Colombo, Federica, et al. "Machine Learning Approaches for Prediction of Bipolar Disorder Based on Biological, Clinical and Neuropsychological Markers: A Systematic Review and Meta-Analysis." *Neuroscience & Biobehavioral Reviews*, vol. 135, Apr. 2022, p. 104552. *ScienceDirect*, <https://doi.org/10.1016/j.neubiorev.2022.104552>.
- Figley, Chase R., et al. "Potential Pitfalls of Using Fractional Anisotropy, Axial Diffusivity, and Radial Diffusivity as Biomarkers of Cerebral White Matter Microstructure." *Frontiers in Neuroscience*, vol. 15, Jan. 2022. *Frontiers*, <https://doi.org/10.3389/fnins.2021.799576>.
- Filley, Christopher M. "White Matter: Organization and Functional Relevance." *Neuropsychology Review*, vol. 20, no. 2, June 2010, pp. 158–73. DOI.org (Crossref), <https://doi.org/10.1007/s11065-010-9127-9>.
- Grande, Iria, et al. "Bipolar Disorder." *The Lancet*, vol. 387, no. 10027, Apr. 2016, pp. 1561–72. www.thelancet.com, [https://doi.org/10.1016/S0140-6736\(15\)00241-X](https://doi.org/10.1016/S0140-6736(15)00241-X).
- Hallahan, Brian, et al. "Structural Magnetic Resonance Imaging in Bipolar Disorder: An

- International Collaborative Mega-Analysis of Individual Adult Patient Data.” *Biological Psychiatry*, vol. 69, no. 4, Feb. 2011, pp. 326–35. *ScienceDirect*, <https://doi.org/10.1016/j.biopsych.2010.08.029>.
- Hartberg, Cecilie, et al. “Brain Cortical Thickness and Surface Area Correlates of Neurocognitive Performance in Patients with Schizophrenia, Bipolar Disorder, and Healthy Adults.” *Journal of the International Neuropsychological Society : JINS*, vol. 17, Nov. 2011, pp. 1080–93. *ResearchGate*, <https://doi.org/10.1017/S1355617711001081>.
- Hibar, D. P., L. T. Westlye, N. T. Doan, et al. “Cortical Abnormalities in Bipolar Disorder: An MRI Analysis of 6503 Individuals from the ENIGMA Bipolar Disorder Working Group.” *Molecular Psychiatry*, vol. 23, no. 4, Apr. 2018, pp. 932–42. *www.nature.com*, <https://doi.org/10.1038/mp.2017.73>.
- Hibar, D. P., L. T. Westlye, T. G. M. van Erp, et al. “Subcortical Volumetric Abnormalities in Bipolar Disorder.” *Molecular Psychiatry*, vol. 21, no. 12, Dec. 2016, pp. 1710–16. *www.nature.com*, <https://doi.org/10.1038/mp.2015.227>.
- Hirsch, Gabriella V., et al. “Using Structural and Functional Brain Imaging to Uncover How the Brain Adapts to Blindness.” *Annals of Neuroscience and Psychology*, vol. 2, 2015, p. 5.
- Holdsworth, Samantha J., and Roland Bammer. “Magnetic Resonance Imaging Techniques: fMRI, DWI, and PWI.” *Seminars in Neurology*, vol. 28, no. 4, Sept. 2008, pp. 395–406. *PubMed*, <https://doi.org/10.1055/s-0028-1083697>.
- Houenou, Josselin, et al. “Neuroimaging-Based Markers of Bipolar Disorder: Evidence from Two Meta-Analyses.” *Journal of Affective Disorders*, vol. 132, no. 3, Aug. 2011, pp. 344–55. *ScienceDirect*, <https://doi.org/10.1016/j.jad.2011.03.016>.
- Jan, Zainab, et al. “The Role of Machine Learning in Diagnosing Bipolar Disorder: Scoping Review.” *Journal of Medical Internet Research*, vol. 23, no. 11, Nov. 2021, p. e29749. *PubMed Central*, <https://doi.org/10.2196/29749>.
- Lerch, Jason P., et al. “Studying Neuroanatomy Using MRI.” *Nature Neuroscience*, vol. 20, no. 3, Mar. 2017, pp. 314–26. *www.nature.com*, <https://doi.org/10.1038/nn.4501>.
- Lim, K. O., et al. “Cortical Gray Matter Deficit in Patients with Bipolar Disorder.” *Schizophrenia Research*, vol. 40, no. 3, Dec. 1999, pp. 219–27. *ScienceDirect*, [https://doi.org/10.1016/S0920-9964\(99\)00063-8](https://doi.org/10.1016/S0920-9964(99)00063-8).
- Logothetis, Nikos K. “What We Can Do and What We Cannot Do with fMRI.” *Nature*, vol. 453, no. 7197, June 2008, pp. 869–78. *www.nature.com*, <https://doi.org/10.1038/nature06976>.
- Manelis, Anna, et al. “White Matter Abnormalities in Adults with Bipolar Disorder Type-II and Unipolar Depression.” *Scientific Reports*, vol. 11, no. 1, Apr. 2021, p. 7541. *www.nature.com*, <https://doi.org/10.1038/s41598-021-87069-2>.
- Martini, Julia, et al. “Early Detection of Bipolar Disorders and Treatment Recommendations for Help-Seeking Adolescents and Young Adults: Findings of the Early Detection and Intervention Center Dresden.” *International Journal of Bipolar Disorders*, vol. 9, July 2021, p. 23. *PubMed Central*, <https://doi.org/10.1186/s40345-021-00227-3>.
- McDonald, Colm, et al. “Regional Brain Morphometry in Patients With Schizophrenia or Bipolar

- Disorder and Their Unaffected Relatives.” *American Journal of Psychiatry*, vol. 163, no. 3, Mar. 2006, pp. 478–87. *DOI.org (Crossref)*, <https://doi.org/10.1176/appi.ajp.163.3.478>.
- Müller-Oerlinghausen, Bruno, et al. “Bipolar Disorder.” *The Lancet*, vol. 359, no. 9302, Jan. 2002, pp. 241–47. *www.thelancet.com*, [https://doi.org/10.1016/S0140-6736\(02\)07450-0](https://doi.org/10.1016/S0140-6736(02)07450-0).
- Muneer, Ather. “The Discovery of Clinically Applicable Biomarkers for Bipolar Disorder: A Review of Candidate and Proteomic Approaches.” *Chonnam Medical Journal*, vol. 56, no. 3, Sept. 2020, pp. 166–79. *PubMed Central*, <https://doi.org/10.4068/cmj.2020.56.3.166>.
- Niu, Meiqi, et al. “Common and Specific Abnormalities in Cortical Thickness in Patients with Major Depressive and Bipolar Disorders.” *eBioMedicine*, vol. 16, Feb. 2017, pp. 162–71. *www.thelancet.com*, <https://doi.org/10.1016/j.ebiom.2017.01.010>.
- Pan, Cong, et al. “Temporal-Spatial Dynamic Functional Connectivity Analysis in Schizophrenia Classification.” *Frontiers in Neuroscience*, vol. 16, Aug. 2022. *Frontiers*, <https://doi.org/10.3389/fnins.2022.965937>.
- Phillips, Mary L., and Eduard Vieta. “Identifying Functional Neuroimaging Biomarkers of Bipolar Disorder: Toward DSM-V.” *Schizophrenia Bulletin*, vol. 33, no. 4, July 2007, pp. 893–904. *Silverchair*, <https://doi.org/10.1093/schbul/sbm060>.
- Pinto, Jairo Vinicius, et al. “Chapter 30 - Biomarkers for Bipolar Disorder.” *Neurobiology of Bipolar Disorder*, edited by Joao Quevedo et al., Academic Press, 2021, pp. 347–56. *ScienceDirect*, <https://doi.org/10.1016/B978-0-12-819182-8.00032-6>.
- Ranzenberger, Logan R., et al. “Diffusion Tensor Imaging.” *StatPearls*, StatPearls Publishing, 2024. *PubMed*, <http://www.ncbi.nlm.nih.gov/books/NBK537361/>.
- Roy, Paul D., et al. “Temporal Horn Enlargement Is Present in Schizophrenia and Bipolar Disorder.” *Biological Psychiatry*, vol. 44, no. 6, Sept. 1998, pp. 418–22. *DOI.org (Crossref)*, [https://doi.org/10.1016/S0006-3223\(98\)00105-X](https://doi.org/10.1016/S0006-3223(98)00105-X).
- Scaini, Giselli, et al. “Neurobiology of Bipolar Disorders: A Review of Genetic Components, Signaling Pathways, Biochemical Changes, and Neuroimaging Findings.” *Brazilian Journal of Psychiatry*, vol. 42, no. 5, Apr. 2020, pp. 536–51. *PubMed Central*, <https://doi.org/10.1590/1516-4446-2019-0732>.
- Skjelstad, Dag V., et al. “Symptoms and Signs of the Initial Prodrome of Bipolar Disorder: A Systematic Review.” *Journal of Affective Disorders*, vol. 126, no. 1, Oct. 2010, pp. 1–13. *ScienceDirect*, <https://doi.org/10.1016/j.jad.2009.10.003>.
- Strakowski, Stephen M., et al. “Ventricular and Periventricular Structural Volumes in First-Versus Multiple-Episode Bipolar Disorder.” *American Journal of Psychiatry*, vol. 159, no. 11, Nov. 2002, pp. 1841–47. *psychiatryonline.org (Atypon)*, <https://doi.org/10.1176/appi.ajp.159.11.1841>.
- Sussmann, Jessika E., et al. “White Matter Abnormalities in Bipolar Disorder and Schizophrenia Detected Using Diffusion Tensor Magnetic Resonance Imaging.” *Bipolar Disorders*, vol. 11, no. 1, 2009, pp. 11–18. *Wiley Online Library*, <https://doi.org/10.1111/j.1399-5618.2008.00646.x>.

- Takahashi, Tsutomu, et al. “Insular Cortex Volume in Established Bipolar Affective Disorder: A Preliminary MRI Study.” *Psychiatry Research: Neuroimaging*, vol. 182, no. 2, May 2010, pp. 187–90. *ScienceDirect*, <https://doi.org/10.1016/j.pscychresns.2010.01.006>.
- Teixeira, Antonio L., et al. “Biomarkers for Bipolar Disorder: Current Status and Challenges Ahead.” *Expert Review of Neurotherapeutics*, vol. 19, no. 1, Jan. 2019, pp. 67–81. *DOI.org (Crossref)*, <https://doi.org/10.1080/14737175.2019.1550361>.
- van der Schot, Astrid C., et al. “Influence of Genes and Environment on Brain Volumes in Twin Pairs Concordant and Discordant for Bipolar Disorder.” *Archives of General Psychiatry*, vol. 66, no. 2, Feb. 2009, pp. 142–51. *Silverchair*, <https://doi.org/10.1001/archgenpsychiatry.2008.541>.
- Vargas, Cristian, et al. “A Systematic Literature Review of Resting State Network—Functional MRI in Bipolar Disorder.” *Journal of Affective Disorders*, vol. 150, no. 3, Sept. 2013, pp. 727–35. *ScienceDirect*, <https://doi.org/10.1016/j.jad.2013.05.083>.
- Wang, Xiuli, et al. “Brain Grey-Matter Volume Alteration in Adult Patients with Bipolar Disorder under Different Conditions: A Voxel-Based Meta-Analysis.” *Journal of Psychiatry & Neuroscience : JPN*, vol. 44, no. 2, Mar. 2019, pp. 89–101. *PubMed Central*, <https://doi.org/10.1503/jpn.180002>.
- Wang, Yanlin, et al. “Large-Scale Network Dysfunction in the Acute State Compared to the Remitted State of Bipolar Disorder: A Meta-Analysis of Resting-State Functional Connectivity.” *eBioMedicine*, vol. 54, Apr. 2020. *www.thelancet.com*, <https://doi.org/10.1016/j.ebiom.2020.102742>.
- Wise, T., et al. “Common and Distinct Patterns of Grey-Matter Volume Alteration in Major Depression and Bipolar Disorder: Evidence from Voxel-Based Meta-Analysis.” *Molecular Psychiatry*, vol. 22, no. 10, Oct. 2017, pp. 1455–63. *www.nature.com*, <https://doi.org/10.1038/mp.2016.72>.
- Woods, Bryan T., et al. “Progressive Ventricular Enlargement in Schizophrenia: Comparison to Bipolar Affective Disorder and Correlation with Clinical Course.” *Biological Psychiatry*, vol. 27, no. 3, Feb. 1990, pp. 341–52. *www.biologicalpsychiatryjournal.com*, [https://doi.org/10.1016/0006-3223\(90\)90008-P](https://doi.org/10.1016/0006-3223(90)90008-P).

Analyzing the Prominent Environmental Factors that Cause Forest Fires in the Algerian and Montesinho Forests By Sneha Narayan

Abstract

Forest fires are becoming increasingly common due to rising temperatures from global warming. According to the National Interagency Fire Center (NIFC), in 2022, 66,225 fires in the United States burned 7,534,403 acres of land. Data from the Global Forest Watch (GFW) indicates that in the last 10 years, around 82 million hectares of forests have been destroyed by wildfires worldwide. In this paper, I use a random forest classifier to predict the occurrence of a forest fire given a set of environmental conditions (Fine Fuel Moisture Code, Duff Moisture Code, Drought Code, Initial Spread Index, temperature, relative humidity, wind speed, rain, and the month in which the data was collected). A random forest classifier uses an ensemble of decision trees—which ask a series of binary questions to split the data—to determine whether a fire is likely, and displays which factors are most important in making this decision. My model achieves 100% accuracy in predicting whether a forest fire occurs given a set of environmental conditions from the Algerian Forest Fires dataset, and achieves 55.78% accuracy when given the Montesinho Forest Fires dataset. Knowing the prominent factors responsible for wildfire formation is useful for devising measures to offset the conditions and prevent harm. In the future, we can compare environmental data from more forests around the world to form a holistic view of the environmental factors that cause forest fires.

Introduction

In the last 10 years, around 82 million hectares of forests have been destroyed by wildfires worldwide [1]. Forest fires are common in hot, dry, and windy areas. They pose great harm to infrastructure in surrounding areas and can release harmful contaminants, sediments, and heavy metal that leach into nearby water sources [2]. Machine learning models can be used to analyze the occurrence of these forest fires and the factors that influence their formation. Specifically, this paper uses a random forest classifier to identify the factors which most affect wildfire ignition in the Algerian and Montesinho forests.

Literature Review: Forest Fires

Forest fires are large, uncontrolled fires that burn in vegetation more than six feet in height. Ground fires typically ignite in soil with thick organic matter and can last a full season. Surface fires burn what is on the ground, such as dead leaves, parched grass, and other types of dry vegetation. Crown fires are the biggest type of forest fire and burn tree canopies [3]. Forest fires are greatly influenced by temperature, humidity, precipitation, and wind [4]. High temperatures cause vegetation to dry out, which provides more fuel for fires [5]. Wind increases the supply of oxygen to the fires, which accelerates the ignition and spread of the forest fire. Precipitation and humidity can prevent fire ignition because it dampens the fuel—the air exchanges moisture with the dry vegetation, causing it to become more moist [6].

Forest fires have a damaging effect on wildlife and ecosystems. High severity fires can burn tree canopies and scorch the soil and tree roots [7]. However, the effect is even greater than destroying ecosystems. Smoke from forest fires include a mixture of toxic pollutants such as PM2.5 (fine particulate matter), nitrogen dioxide, ozone, and lead—which contaminate the air. In addition, forest fires release carbon dioxide and other greenhouse gasses into the air, increasing the impact of climate change. With global warming leading to higher temperatures, the risk of forest fires is increased, leading to a vicious cycle [8]. Therefore, it is important to accurately identify the factors that lead to forest fire ignition in order to be better prepared to combat the fires early on and devise mitigation strategies.

According to the National Interagency Fire Center, in 2022, 66,225 fires in the U.S. burned 7,534,403 acres of land [9]. Data from the Global Forest Watch indicates that in the last 10 years, around 82 million hectares of forests have been destroyed by wildfires worldwide [10]. Climate change, which results in global warming and extended drought, has increased the risk of forest fires in the United States over the last two decades [11]. Data from researchers at the University of Maryland shows that between the years 2001 and 2023, the area burned by forest fires increased by around 5.4% per year. Now, forest fires result in nearly 6 million more hectares of tree cover loss per year than they did in 2001—an area roughly the size of Croatia [12]. Data from the Global Fire Watch shows that fires were responsible for 74% of tree cover loss in Algeria between 2001 and 2023 [13]. In contrast, fires were responsible for 34% of tree cover loss in Portugal between 2001 and 2023 [14].

Literature Review: Machine Learning and Random Forest Classifier

Machine learning is the ability to train a machine to imitate human behavior. It is a subset of artificial intelligence. There are two major types of machine learning methods—supervised learning and unsupervised learning. Supervised learning uses labeled data to train the model. On the other hand, learning is unsupervised when the dataset does not have output labels. This means that the model forms clusters based on patterns identified in the data. Within supervised learning, there can be a classification model (where the predictive variable is categorical) or a regression model (where the predictive variable is continuous and numerical).

There are two stages when creating a supervised machine learning model—training and testing. In the training stage, the model creates a function for which the value from that function most closely relates to the true label of the data. In the testing stage, the function previously created is used to predict the labels of new data. Overfitting occurs when the model too closely adapts to the intricacies of the training data without generalizing results. The data is split into two categories, training and testing. Some common training/testing splits are 70% training and 30% testing, 80% training and 20% testing, and 90% training and 10% testing.

A decision tree uses a series of binary questions to split the data. The questions at each node of the decision tree are carefully selected during the training phase to split the data in the best way possible. These same questions are then used during the testing phase on a new set of data. A random forest is an ensemble of decision trees, all of which are used to make the final

decision. In a decision tree, each point where a decision is made is called a “decision node”. Each terminal node is called a “leaf node”. When using a random forest algorithm, the importance of each feature in making the final decision can be shown. Some of the benefits of using a random forest algorithm include: accuracy for small datasets, ability to identify feature importance, ability to handle numerical and categorical data, and ease to interpret and visualize results. Some downsides are that random forest algorithms can create biased trees, overfit the data, and have greedy algorithms. However, these negatives can be mitigated by using an ensemble of trees to make the decision.

Methods

I created a Random Forest Classifier to predict the occurrence of a forest fire. I used methods from the pandas library in Python to clean the data, scikit-learn methods to split the data and create the model, and matplotlib for data visualization.

I ran my model on two different labeled datasets— “Algerian Forest Fires Dataset”¹ and “Montesinho Forest Fire Prediction Dataset”². Each dataset recorded the environmental conditions during times of a fire and normal conditions in a different area—the Bejaia region and Sidi Bel-abbes region in Algeria, and the Montesinho forest in Portugal. The Montesinho forest fires dataset recorded the burn area, so I cleaned the data and sorted it into two classes—whether a forest fire occurred or not. I used the burn area to classify the data into these two classes and made the assumption that a “0” burn area indicated that there was no forest fire.

I further cleaned the datasets so that they included the same environmental factors, making it an apples-to-apples comparison. Each dataset has the following features: Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC), Initial Spread Index (ISI), temperature, relative humidity, wind speed, rain, and the month in which the data was collected. The Fine Fuel Moisture Code is a numeric rating of the moisture content of litter and other fine fuels such as small twigs, grasses, and ferns [15]. It indicates the relative ease of ignition and flammability of fine fuels. The higher the FFMC, the greater the risk of a fire. The Duff Moisture Code is a rating of the average moisture content of moderately deep loosely compacted layers of decomposing organic matter. The Drought Code Index measures the average moisture content of deep, compact organic layers. The Initial Spread Index is calculated by taking the product of the Fine Fuel Moisture Code and Wind Speed. It is the expected rate of fire spread [16].

After cleaning the data, I used scikit-learn methods to split it into training and testing sets. For both the Algerian Forests and Montesinho region datasets, I split the data into 80% training and 20% testing. I used the RandomizedSearchCV scikit-learn method to randomly choose different combinations of hyperparameters. I then used the `best_estimator_` method to find the best combination of hyperparameters. This resulted in a high accuracy when the model was trained on the Algerian Forest Fires dataset. However, since the Monteinho Forest Fires

¹ Algerian Forest Fires Dataset:

<https://www.kaggle.com/datasets/sudhanshu432/algerian-forest-fires-cleaned-dataset>

² Montesinho Forest Fires Dataset: <https://www.kaggle.com/datasets/elikplim/forest-fires-data-set>

dataset still yielded a low accuracy, I further tuned the `n_estimators` (number of decision trees that are used in the random forest) hyperparameter (shown in Figure 1). After doing so, I found the optimal hyperparameters for the Montesinho forest fires dataset. When tuning the hyperparameters, I used accuracy as the evaluation metric. Table 1 shows the distributions I used to tune the hyperparameters.

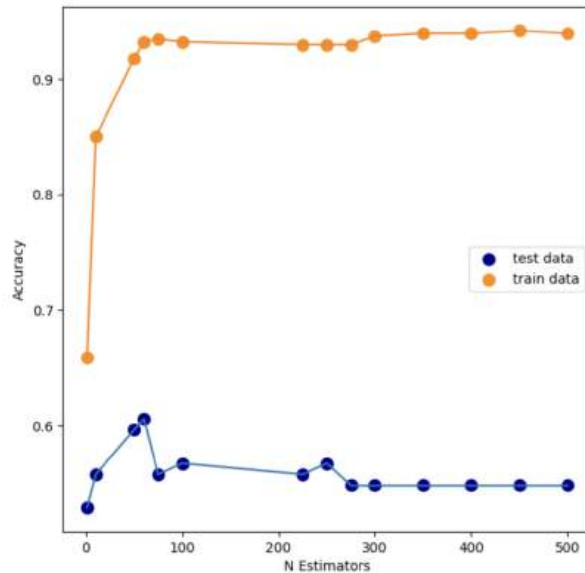


Figure 1: N Estimators vs Accuracy Plot for the model trained on the Montesinho Forest Fires Dataset.

As seen in Figure 1, when the number of estimators increases, the accuracy of the model on the training data also increases. The number of estimators refers to the number of decision trees in the random forest. There is a peak in the test data accuracy at around 50-60 estimators, but this is likely due to a fluke. The accuracy of the model levels off around 300 estimators which means including over 300 estimators will result in marginal benefits.

Hyperparameter Name	Description	Type and Range	Decided Value
<code>n_estimators</code>	The number of decision trees in the random forest.	Integer (50-3000)	350
<code>max_depth</code>	The maximum depth of the decision tree. If it is set to none, then the nodes will expand until all leaves are pure (where all data in that node belongs to a single class) or until all leaves contain less samples than <code>min_samples_split</code> .	Integer (1-20)	8
<code>max_features</code>	The number of features to consider when looking for the best fit. If “sqrt” then the features considered will be the square root of the total features. If “log2”, the	sqrt, log2, none	None

	number of features considered will be \log_2 (number of features). If “none” then all features will be considered.		
oob_score	Whether to test the model on random samples of data to estimate how well the model generalizes to new data.	True, False	True
min_samples_split	The minimum number of samples required to split an internal node	Integer (2-10)	7

Table 1: Distributions used to tune the hyperparameters for the model trained using the Montesinho Forest Fires dataset. The hyperparameters I tuned are “n_estimators”, “max_depth”, “max_features”, “oob_score”, and “min_samples_split”.

Results

The model works well on the Algerian Forest Fires dataset, with 100% accuracy. The confusion matrix in Figure 2 shows that each piece of data is correctly classified.

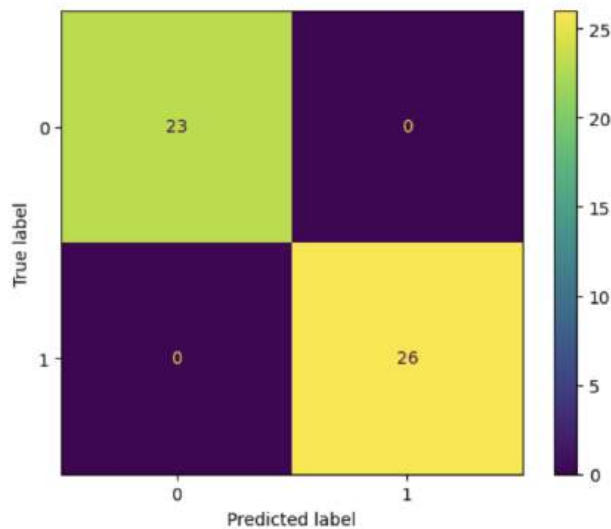


Figure 2: Confusion matrix of the model trained on the Algerian Forest Fires Dataset. The predicted label and true label matched for each piece of training data, meaning all the data was correctly classified. Here the label “0” represents that there was no forest fire and “1” indicates the presence of a forest fire.

The feature importance table in Table 2 indicates that the Initial Spread Index (ISI) and Fine Fuel Moisture Code (FFMC) have the greatest impact in making the decision for whether a fire will occur or not. On the other hand, features such as the month in which the data was recorded, wind speed, temperature, and relative humidity do not have a big impact on the final decision, as seen by their low feature importance values.

Feature	Importance
ISI	0.364363
FFMC	0.363650
DMC	0.094542
Rain	0.070288
DC	0.069959
Temperature	0.014830
RH	0.014468
Ws	0.005415
month	0.002485

Table 2: Feature importance table of model trained on the Algerian Forest Fires dataset. The Initial Spread Index (ISI) and Fine Fuel Moisture Code (FFMC) have the greatest importance. Features such as the month in which the data was recorded, wind speed, temperature, and relative humidity do not have a big impact on the final decision as shown by their feature importance values of less than .015 (or about 1.5%).

The model was trained with 500 estimators—which means that there were 500 decision trees in the random forest, from which the results were averaged together to determine whether a fire would emerge given the parameters. Figure 3 shows an example of one of the 500 decision trees. The tree is simple and does not have too many layers. Near the top of the tree, the model is making splits with decisive criteria, which split the data roughly in half. All the leaves in the decision tree are pure.

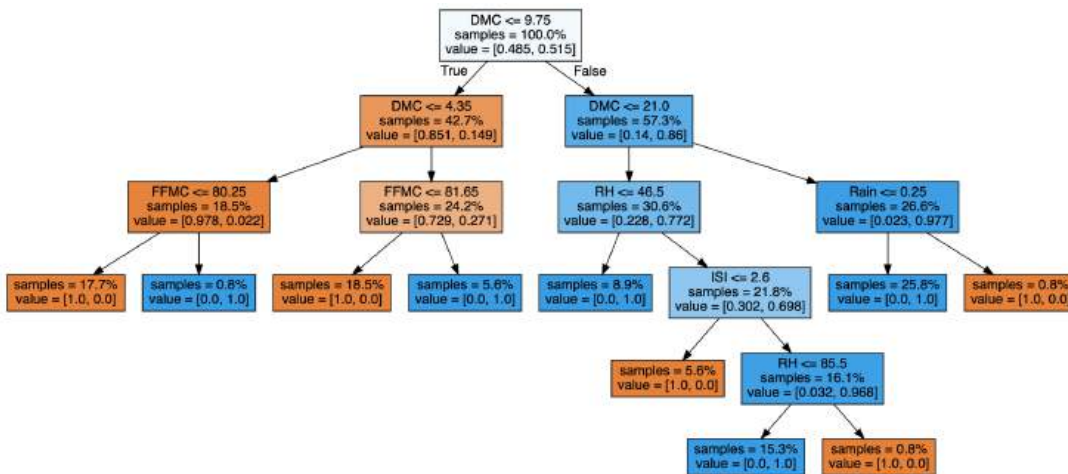


Figure 3: Example of one decision tree from the random forest model trained on the Algerian Forest Fires dataset. All the leaves in the decision tree are pure. The tree is simple and does not have too many layers. Near the top of the tree, the model is making splits with good criteria, which split the data roughly in half. All the leaves in the decision tree are pure.

In contrast, the model has a much lower accuracy for the Montesinho Forest Fires dataset. After hyperparameter tuning (shown in Figure 1), the model achieves an accuracy of 55.78%. The confusion matrix in Figure 4 shows that the model is not good at predicting when the true label for the data is “0”, meaning that the conditions would not cause a forest fire. As seen in the confusion matrix, when the true label is “0”, the model predicts correctly 20 out of 42 times, and predicts incorrectly 22 out of 42 times. This indicates that the model cannot find good patterns in the data when there are no forest fires, and therefore is unable to accurately make predictions.

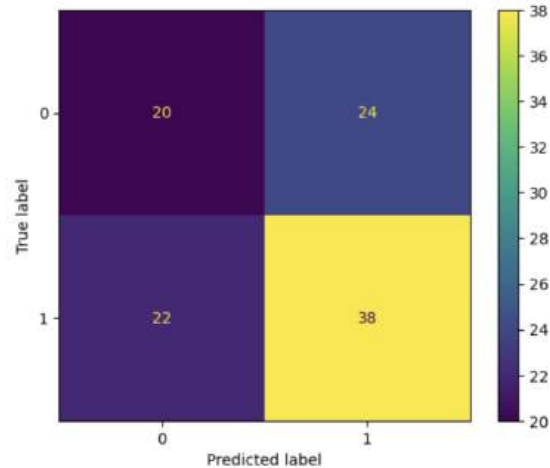


Figure 4: Confusion matrix of the model trained on the Montesinho Forest Fires Dataset. Here 0 represents no forest fire and 1 indicates the presence of a forest fire. When the true label is 0, the model cannot accurately predict whether there is a forest fire and essentially guesses—there is a 50% chance of the model predicting either label. This means that the model can not find good patterns in the data when there are no forest fires, and therefore is unable to accurately make predictions. When the true label is 1, the model is more accurate in its predictions but still not as accurate as in the Algerian Forest Fires dataset.

When looking at an example decision tree (Figure 5), it is evident that the model does not make good splits in the data early on. Instead of using conditions to split the data roughly in half, the model overfits to the data, isolating individual samples in the beginning rather than towards the middle/end. This means that the model is not able to accurately identify patterns in the data and instead is fitting very closely to the intricacies of the training data rather than producing generalizable results.

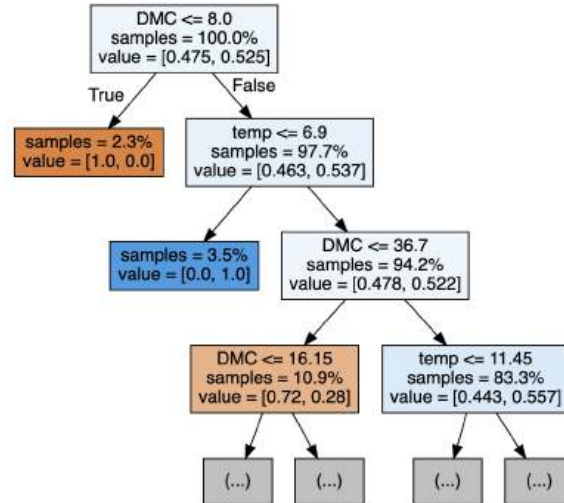


Figure 5: Example of the first three layers from one decision tree from the random forest model trained on the Montesinho Forest Fires dataset. The model overfits the data. Instead of using conditions to split the data roughly in half, the model isolates samples early on. For example, the initial split partitions the data into 97.7% and 2.3%, creating a pure leaf. If the model were learning the patterns in the data, the split would be closer to 50% and 50%.

The feature importance table in Table 3 shows that temperature, relative humidity, and the Duff Moisture Code are the most prominent factors in making the decision for whether a forest fire will occur or not. This means that a higher priority should be given to offset these factors. Unlike the feature importances for the Algerian Forest Fires dataset, where there is a clear indication of the features that are the most vital in making the final decision, the feature importances for each feature in the Portugal dataset are relatively similar to one another.

Feature	Importance
temp	0.217195
RH	0.176565
DMC	0.146804
DC	0.134249
FFMC	0.103174
wind	0.101554
ISI	0.094225
month	0.024026
rain	0.002207

Table 3: Feature importance table of model trained on the Montesinho Forest Fires dataset. Temperature, relative humidity, and the Duff Moisture Code are the most prominent factors in making the decision for whether a forest fire will occur or not. Unlike the feature importances for the Algerian Forest Fires dataset, where there is a clear indication of the features that are the most vital in making the final decision, the feature importances for each feature in the Portugal dataset are relatively similar.

Discussion

As seen in Figure 6, the decision trees for the model trained on the Algerian Forest Fires dataset were much simpler than those for the model trained on the Montesinho Forest Fires Dataset.

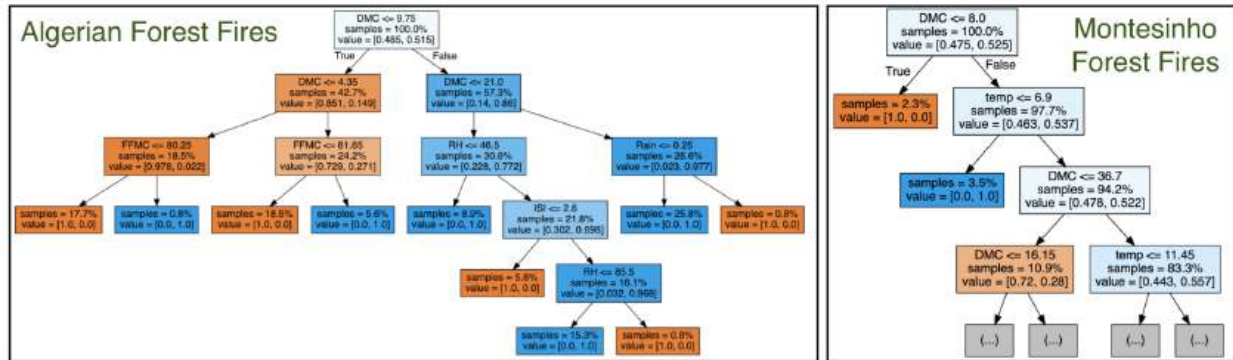


Figure 6: Side by side comparison of an example decision tree from the model trained on the Algerian Forest Fires dataset and the first three layers of an example decision tree from the model trained on the Montesinho Forest Fires dataset.

For the model trained on the Algerian Forest Fires dataset, the model made good splits early on in the data—splitting the data roughly in half in the beginning few splits. On the other hand, the decision trees for the Montesinho Forest Fires dataset isolated samples in the first few splits, which is an indication that the model was overfitting the data. Since the model worked well for the Algerian Forest Fires dataset, this can be an indication that there is no significant pattern in the features that correlates with forest fire presence in the Montesinho region (since the same features are used in both datasets). As stated in the literature review, forest fires are much more common in Algeria than Portugal—forest fires resulted in 74% [13] of tree cover loss in Algeria between 2001 and 2023 and only 34% [14] of tree cover loss in Portugal. This could potentially explain the low accuracy of the model trained on the Portugal dataset. Additionally, the data used for forest fires in the Montesinho Forest was originally a burn area dataset which I manipulated to serve as whether a forest fire occurred or not. I made an assumption that a 0 burn area indicated that a fire did not occur, but it is possible that a 0 burn area could have represented a very small fire.

Further, the feature importance tables for both datasets are different which could mean that the environmental factors that cause forest fires are specific to the region, meaning that results cannot necessarily be generalized to all areas. For the Algerian Forest Fires dataset, the important features were the initial spread index and Fine Fuel Moisture Code. This makes sense because high content of fine fuels leads to ease of fire ignition. For the Montesinho Forest Fires dataset, the temperature and relative humidity were the most dominant factors. The Montesinho Forest is located in northern Portugal, where areas experience higher humidity levels. Summers are humid and winters are not overly humid [17]. Due to the variation of humidity, it plays a

large role in forest fire occurrence. On days when it is more humid, fires are less likely to occur due to the dampening effect the moisture in the air creates.

Conclusion

I built a random forest classifier to predict whether a forest fire is likely to form given a set of environmental conditions. I trained the model on two datasets—Algerian Forest Fires and fires in the Montesinho region. The model achieved a 100% accuracy when trained on the Algerian Forest Fires dataset, and achieved 55.78% accuracy when trained on the Montesinho Forest Fires dataset. By looking at the example decision trees for both datasets, it is likely that the model overfitted the data from fires in the Montesinho forest. The decision trees were far more complex for the Montesinho Forest Fires dataset compared to the Algerian Forest Fires dataset. The feature importance for both datasets are significantly different. The key features for the Algerian Forest Fires dataset are Initial Spread Index and Fine Fuel Moisture Code. For the Monteinsho Forest Fires dataset, the dominant features are temperature, relative humidity, and Duff Moisture Code. This is an indication that the results of important features cannot necessarily be generalized to fires in all regions.

Forest fires are extremely destructive and detrimental to the environment, so it is crucial to prevent them. Understanding the environmental factors that cause forest fires is the first step in order to offset such conditions to reduce the likelihood of fires.

Future work includes analyzing datasets from forests around the world to develop a more cohesive understanding of the factors which greatly affect forest fire formation. It is also vital to understand the key environmental factors that cause forest fire formation and take preventative measures to prevent the conditions from reaching a point of harm.

Works Cited

- [1] “Wildfires Destroy 82m Hectares of Forest Worldwide in Decade.” *Anadolu Ajansı*, www.aa.com.tr/en/world/wildfires-destroy-82m-hectares-of-forest-worldwide-in-decade/2954918. Accessed 6 July 2024.
- [2] “5 Negative Effects of Wildfires.” *WFCA*, 4 Mar. 2024, wfca.com/wildfire-articles/negative-effects-of-wildfires/.
- [3] “Forest Fire.” Encyclopædia Britannica, Encyclopædia Britannica, inc., 30 July 2024, www.britannica.com/science/forest-fire.
- [4] “Wildfires and Climate Change.” Center for Climate and Energy Solutions, 14 July 2023, www.c2es.org/content/wildfires-and-climate-change/.
- [5] “Wildfires.” Education, education.nationalgeographic.org/resource/wildfires/. Accessed 30 July 2024.
- [6] “How Does Humidity Affect a Fire?” *WFCA*, 4 Mar. 2024, wfca.com/wildfire-articles/how-does-humidity-affect-wildfire.
- [7] Meghan Snow. “How Does Wildfire Impact Wildlife and Forests?: U.S. Fish & Wildlife Service.” *FWS.Gov*, 11 Oct. 2022, www.fws.gov/story/2022-10/how-does-wildfire-impact-wildlife-and-forests.
- [8] “Wildfires.” World Health Organization, World Health Organization, www.who.int/health-topics/wildfires#tab=tab_1. Accessed 30 July 2024.
- [9] “NCEI Monthly Fire Report.” National Centers for Environmental Information (NCEI), www.ncei.noaa.gov/access/monitoring/monthly-report/fire/. Accessed 11 July 2024.
- [10] “Wildfires Destroy 82m Hectares of Forest Worldwide in Decade.” *Anadolu Ajansı*, www.aa.com.tr/en/world/wildfires-destroy-82m-hectares-of-forest-worldwide-in-decade/2954918. Accessed 17 July 2024.
- [11] “Wildfire Climate Connection.” National Oceanic and Atmospheric Administration, www.noaa.gov/noaa-wildfire/wildfire-climate-connection. Accessed 17 July 2024.
- [12] MacCarthy, James, et al. “The Latest Data Confirms: Forest Fires Are Getting Worse.” World Resources Institute, 13 Aug. 2024, www.wri.org/insights/global-trends-forest-fires.
- [13] Vizzuality. “Algeria Deforestation Rates & Statistics: GFW.” Forest Monitoring, Land Use & Deforestation Trends, www.globalforestwatch.org/dashboards/country/DZA?category=fires. Accessed 5 Oct. 2024.
- [14] Vizzuality. “Portugal Deforestation Rates & Statistics: GFW.” Forest Monitoring, Land Use & Deforestation Trends, www.globalforestwatch.org/dashboards/country/PRT?category=fires. Accessed 5 Oct. 2024.
- [15] “LESSON 3 - EFFECTS OF WEATHER, TOPOGRAPHY AND FUELS ON FIRE BEHAVIOR.” Government of Nova Scotia, Canada, novascotia.ca/natr/forestprotection/wildfire/bffsc/lessons/lesson3/fuels.asp. Accessed 15 Sep. 2024.
- [16] Canada, Natural Resources. “Canadian Wildland Fire Information System: Canadian Forest Fire Weather Index (FWI) System.” Canadian Wildland Fire Information System | Canadian Forest Fire Weather Index (FWI) System, cwfis.cfs.nrcan.gc.ca/background/summary/fwi. Accessed 15 Aug. 2024.
- [17] “Climate: North Portugal.” *Worlddata.Info*, www.worlddata.info/europe/portugal/climate-north.php. Accessed 5 Oct. 2024.

Foreign Relations In Maoist and Post-Maoist China By Winston Vance

Abstract

Geopolitics in modern China are influenced by a complex mix of historical events, leadership priorities, and cultural narratives. Across its extensive history, China has often seen itself as the "Middle Kingdom," a term emphasizing its centrality and positioning it as a model for neighboring regions. Yet, its historical glory has been punctuated by periods of vulnerability and external threats. This history has led Chinese leaders, at times, to adopt isolationist policies to safeguard their sovereignty. Despite changes in leadership and geopolitical strategies, China's trajectory continues to be a balancing act between maintaining internal control and projecting global ambition. While China has engaged in the global economy and formed strategic alliances, it remains rooted in authoritarian governance, maintaining a cautious and assertive stance in international affairs. This approach is encapsulated in Deng Xiaoping's "24-Character Formula," which emphasizes patience and caution in foreign policy. China's evolution reveals an ongoing tension between its aspirations to be a global power and its need for domestic stability and control. China has struggled to maintain its sovereignty against foreign encroachment since the 17th century. This paradoxical awareness—of being a great civilization yet vulnerable when lagging behind global powers—has become a defining feature of modern Chinese identity. The tension between harnessing China's internal strengths and engaging with the world is evident in the contrasting policies of Mao Zedong and Deng Xiaoping.

China's Dual Identity

China has long viewed itself as the "Middle Kingdom," a translation of "Zhōngguó," symbolizing its perception as the central and superior civilization of the world. This term reflects China's historical role as the cultural and political core of East Asia, a position that reinforced its sense of superiority over neighboring regions (Kennedy). However, despite its self-assurance in its grandeur, China has endured significant periods of vulnerability to external threats. In response, its leaders often adopted isolationist policies to safeguard the nation. One of the earliest instances was the 1613 Edict of Haijin, issued by the Hongwu Emperor to curb piracy and foreign influence, which restricted maritime trade and limited foreign contact (Rowe). While these repressive policies protected China from foreign incursion, they also isolated the nation from modernization. This vulnerability became particularly apparent during the "Century of Humiliation" (1839–1949), starting with the First Opium War. China suffered a series of defeats at the hands of foreign powers, losing control of key territories and ports. In 1842, the Treaty of Nanjing forced China to cede Hong Kong to Britain and open "treaty ports" to Western nations. These events left lasting scars on China's national psyche and deeply influenced its modern identity and foreign policy. As recently as 2011, testimony before the U.S.-China Economic and Security Review Commission emphasized the enduring impact of the "Century of Humiliation" on China's national consciousness, likening its significance to America's fight for independence from Britain (Kaufman).

The contrast between the narratives of the "Middle Kingdom" and the "Century of Humiliation" reveals a profound duality in China's self-perception. While China takes immense pride in its rich history and vast territory, it remains acutely aware of the vulnerability it has experienced. The memory of territorial losses to foreign powers is deeply etched in China's collective memory, continuing to shape its domestic and international policies. This duality—between historical greatness and lessons from past subjugation—plays a central role in shaping China's contemporary cultural and political outlook.

Priorities Under Mao

Tensions between inwardly- and outwardly-focused policymaking were particularly marked during the Maoist period. Under the leadership of Chairman Mao Zedong, from 1949 to 1976, China adopted a notably inward-facing strategy. Drawing on its historical "Middle Kingdom" identity, Mao emphasized the mobilization of China's revolutionary peasant labor force and relied heavily on internal resources. A defining decision during Mao's leadership was his break from Soviet Communism to distance China from Soviet influence and assert domestic control in order to pursue a separate vision for peasant-led Communist revolution. When seeking allies, Mao rejected first-world nations and identified itself with third-world revolutionaries, rejecting the dominance of Western powers and further isolating itself from modernization.

Mao prioritized economic independence and reducing reliance on foreign nations, particularly Western powers, which he viewed as predatory. Even before his rise to power, Mao expressed his disdain for Western influences, famously criticizing foreign control during China's civil war. He stated, "The false show of friendship by foreigners (especially Englishmen and Americans) is merely a pretense of 'amity' in order that they may squeeze out more of the fat and blood of the Chinese people." His later remarks, such as in *We Must Do Economic Work*, reaffirmed his belief in self-reliance: "We stand for self-reliance. We hope for foreign aid but cannot be dependent on it; we depend on our own efforts, on the creative power of the whole army and the entire people." Mao firmly believes that China's greatest potential lay in the promise of mass mobilization and armed revolution with and external alliances limited to third-world nations.

Mao's ambitious vision aimed at transforming China into a powerful, self-sufficient nation. However, his Great Leap Forward, designed to rapidly industrialize China through mass mobilization, ended in disaster. The push for rapid modernization resulted in widespread famine, causing an estimated 45 million deaths. This human tragedy, along with Mao's rejection of foreign powers, left China economically isolated and technologically behind, with catastrophic consequences for China's geopolitical and economic standing.

The Sino-Soviet Split

The Sino-Soviet split took place in the late 1950s and early 1960s when Mao parted ways with Stalin due to an ideological divide and economic and infrastructural differences between the countries. The Soviets were more industrialized and saw the working class as being central to

communism. Mao, however, was facing a catastrophic economic situation following the failure of The Great Leap Forward. By splitting with the Soviets, Mao was more able to develop his own form of communism, departing from Soviet-style Marxism and Leninism. In the years following the establishment of the People's Republic of China in 1949, the Soviet Union was China's principal ally. By the late 1950s, the Soviets had constructed a network of modern industrial plants in China, supporting the production of warplanes, tanks, and warships. Despite their alliance, Mao harbored a deep distrust of Soviet premier Nikita Khrushchev due to his abandonment of Leninist and Stalinist doctrines. Relations were extremely tense by the late 1950s. As Mao solidified his power in Beijing, Khrushchev ridiculed the failures of the Great Leap Forward.

The split occurred in the late 1960s and was marked by minor battles along the Sino-Soviet borders. As Huanhee Kim notes, "The Sino-Soviet border clash of March 1969 signaled the militarization of the long-standing conflict between two former ideological allies," marking "the ultimate manifestation of the unremitting hostility that had characterized their relationship since the early 1960s" (Kim 113). Similarly, Yanan He observes that during the Cultural Revolution (1966–1976), Mao Zedong's followers deliberately orchestrated anti-Soviet propaganda to advance their political agenda. "Maoist polemicists discredited and attacked domestic foes by accusing them of disloyalty and collaboration with Soviet revisionists to undermine China's Communist system" (He 120).

The break from Soviet communism that followed the split allowed Mao to pursue his own ideologies, and he enacted his own policies and programs, including the Great Leap Forward and the Cultural Revolution. China's diplomatic relations with many countries were strained during this period, leading to its withdrawal from international organizations and increasing isolation on the global stage. A slogan appearing in Chinese propaganda between the years declared: "All people of the world unite, to overthrow American imperialism, to overthrow Soviet revisionism, to overthrow the reactionaries of all nations!" (Chinese Propaganda Dept.). Because Mao had divided and isolated China, he was more able to maintain firm control over domestic politics without external interference. However, by the end of the decade, as a result of China's isolation under Mao, the country had limited access to international trade and technological innovations and was losing ground economically.

Evolving Foreign Policy Strategy: Mao to Deng

At the end of the Cultural Revolution, Mao himself acknowledged that China lacked global influence. However, instead of seeking alliances with first-world nations, he aligned China with revolutionary movements in third-world countries in Asia, Africa, and Latin America. This strategic shift was foreshadowed by the 1955 Bandung Conference in Indonesia, where representatives from 29 governments discussed collaboration to resist first-world colonization and assert their own interests (Office of the U.S. Historian). By the 1970s, Mao positioned China as a leader among third-world nations, aiming to challenge the dominance of both first-world countries like the United States and second-world powers like the Soviet Union.

Mao's strategy to expand China's influence focused on supporting anti-colonial and anti-imperialist movements, providing assistance to liberation struggles as a way to counter Soviet influence in the developing world. As noted in *Global Asia*, “when Mao shifted his focus to the Third World, he viewed it as a ‘land of opportunity.’ His old alliances were disintegrating, and he was seeking new ones [by] positioning China as a member of the Third World, Mao aimed to establish Chinese political leadership and attain a measure of global influence” (Lee 5-6). Mao sought to unite developing nations in Asia and Latin America with China, building a coalition to resist the influence of first-world powers like the U.S. and the Soviet Union.

Mao’s strategy took a turn in the 1970s when he decided to pursue diplomatic relations with the United States. This move allowed China to balance against the Soviet threat while gaining access to Western technology and investment. This shift in diplomatic policy set the stage for an even greater transition in leadership policy after Mao’s death in 1976.

Mao's successor, Deng Xiaoping, would take this new direction further, abandoning Mao’s focus on revolution in favor of economic modernization and engagement with the global economy. Recognizing the failures of Maoist policies, Deng pivoted sharply. He understood the dangers posed by China’s technological and economic vulnerabilities and sought explicitly to strengthen China's global position through modernization, engagement with the global economy, and by creating external alliances. Deng argued that China’s survival as a socialist state required selective integration of market-facing capitalism. Unlike Mao, he believed that by modernizing and opening up to foreign investment and technology, China could close the gap with the West and compete on the global stage. Deng’s policies marked a dramatic departure from Mao’s anti-Western stance. After Mao's death, Deng prioritized economic reforms and global engagement, initiating China's rise as an economic powerhouse. This shift allowed China to re-enter the global order and pursue modernization, reversing the isolation and stagnation caused by Mao’s earlier policies. Through Deng’s pragmatic reforms, China began to position itself as a global economic player while maintaining its socialist foundations.

As Jian Chen noted for *The Wilson Center*, Deng was prepared to leave behind China’s revolutionary status, embracing the global capitalist market as a means to achieve modernization. “Deng was ready to abandon the ‘revolutionary country’ status that China had persistently claimed during the Maoist era . . . [China] would also gradually morph into an ‘insider’ of the existing international systems and institutions dominated by the United States and the capitalist West” (Chen 2). Deng himself emphasized the need for technological advancement and modernization, stating, “China must not be content to remain backward. It should participate in the development of high technology from the very beginning.”

However, Deng’s embrace of modernization did not mean he fully accepted Western ideologies. In his 1989 address, *Urgent Tasks of China’s Third Generation for Collective Leadership*, Deng reaffirmed the importance of maintaining a uniquely Chinese socialist path. “China would have no future if it did not follow the socialist road [W]e are trying to build socialism, a socialism suited to our own conditions. Otherwise, we should have to act in accordance with the will of the Americans, or of people in other developed countries or in the

Soviet Union. How much independence would we have then?”

While committed to preserving China’s socialist identity, Deng understood that modernization, technology, and foreign capital were essential to China’s success on the global stage. His pragmatic approach to foreign policy was outlined in his famous "24-Character Formula" in 1990, which advised Chinese leaders to: “Observe calmly; secure our position; cope with affairs calmly; hide our capacities and bide our time; be good at maintaining a low profile; and never claim leadership.” As James Holmes of the U.S. Naval Institute explains, “Deng wanted China to abstain from power politics until it was strong enough to assert itself in Asian and world politics without undue risk. Under his stewardship, China postponed—but did not cancel—its ambitions” (Holmes 1,143).

In essence, Deng’s vision for China was to pursue modernization and global integration while maintaining a distinctly independent socialist path. By adopting a strategy of cautious diplomacy and economic reform, Deng aimed to elevate China’s global stature without risking its sovereignty or succumbing to external influence.

Tiananmen Square Massacre

Until 1989, China underwent rapid economic growth and significant social transformation under Deng Xiaoping's leadership. The country recovered from the devastation of the Great Leap Forward and the Cultural Revolution, while adopting elements of a free-market economy. During this period of greater openness to the outside world, Chinese society experienced expanded cultural and personal freedoms, along with increased exposure to Western influences and democratic ideals. However, these changes also triggered anxieties about the country's future, both among the general population and within the leadership, leading to growing internal tensions (Liu).

Divisions reached a boiling point in April 1989 when students gathered in Tiananmen Square to protest the death of Hu Yaobang, a former General Secretary of the Chinese Communist Party (CCP) who was a proponent of political reform. Their protests called for greater democratic freedoms and continued to intensify throughout May, which prompted Deng to propose martial law. On June 3-4, China's leaders decided to suppress the demonstrations by force. The military was deployed and troops opened fire on the protesters (PBS). The exact death toll from the Tiananmen Square massacre remains uncertain, with estimates ranging from several hundred to several thousand deaths. These events shocked and outraged the world. The infamous Tiananmen Square Massacre resulted in China’s immediate ostracism on the global stage. Many countries imposed sanctions and withdrew from China diplomatically. China’s international reputation was severely damaged. Faced with choices to liberalize or return to authoritarianism, the CCP moved to radically suppress dissent and maintain strict control over internal affairs, ushering in a new era of political repression and isolation from Western powers, which is still playing out today.

Conclusion

Throughout the history of modern China, conflicting forces and narratives have played important roles in its evolution. On one hand, consistent with Mao Zedong's policies, China has shown a deep commitment to controlling its vast population through repressive measures and stringent internal regulations, often isolating itself from the West. On the other hand, China has pursued power within the framework of a global, modernized economy, echoing the policies of Deng Xiaoping. While both leaders sought external alliances, nationalism and authoritarian control have consistently asserted themselves whenever China's central authority has been challenged. Rather than indicating ideological shifts, Mao and Deng's alliances were strategically crafted to advance specific objectives, laying the foundation for China's emergence as a global power. Today, the tension between China's aspiration to be a dominant global player and its tradition of authoritarian rule continues. China actively engages in global affairs to bolster its power and influence. However, its core remains inward-focused and authoritarian, prioritizing stringent control over its populace. Despite policy decisions that may open or close doors, China's interests are fundamentally driven by two key narratives: the preservation of its self-image by reinforcing its superiority and the imperative to safeguard against external threats.

Works Cited

- "Bandung Conference (Asian-African Conference), 1955." Office of the Historian, U.S. Department of State, <https://history.state.gov/milestones/1953-1960/bandung-conf>. Accessed 16 May 2024.
- Chen, Jian. *From Mao to Deng: China's Changing Relations with the United States*. Wilson Center, Nov. 2019, www.wilsoncenter.org/sites/default/files/media/documents/publication/cwihp_wp_92_chen_jian_from_mao_to_deng_chinas_changing_relations_with_us_november_2019.pdf.
- Chinese Propaganda Department. "All People of the World Unite, to Overthrow American Imperialism, to Overthrow Soviet Revisionism, to Overthrow the Reactionaries of All Nations!" 1966-1976.
- Deng, Xiaoping. "Urgent Tasks of China's Third Generation for Collective Leadership." 1989, www.marxists.org/reference/archive/deng-xiaoping/1989/118.htm. Accessed 16 May 2024.
- He, Yinan. "China's Cultural Revolution and Mao's External Threat Inflation: Crushing Soviet-Style Capitalist Restoration, 1966–1969." *Journal of Cold War Studies*, vol. 26, no. 1, 2024, pp. 120–155. https://doi.org/10.1162/jcws_a_01194.
- Holmes, James. "Deng Xiaoping Was China's George Washington." US Naval Institute Proceedings, January 2020, www.usni.org/magazines/proceedings/2020/january/deng-xiaoping-was-chinas-george-washington.
- Jersild, Austin. *The Sino-Soviet Alliance: An International History*. University of North Carolina Press, 2014. JSTOR, doi:10.5149/9781469611600_jersild.
- Kaufman, Alison A. "The 'Century of Humiliation' and China's National Narratives. Testimony before the U.S.-China Economic and Security Review Commission: Hearing on 'China's Narratives Regarding National Security Policy,'" CNA, 10 Mar. 2011.
- Kennedy, Scott, editor. *Beyond the Middle Kingdom: Comparative Perspectives on China's Capitalist Transformation*. Stanford University Press, 2011, <https://www.sup.org/books/title/?id=17768>. Accessed May 15, 2024.
- Kim, Hwanhee. "The Making of a Cold War Turning Point: The Sino-Soviet Split and the Prelude to Detente with the United States, 1965-1968." *The Korean Journal of International Studies*, vol. 12, no. 1, 2014, pp. 113-136. <https://doi.org/10.14731/kjis.2014.06.12.1.113>.
- Larkin, Brian. "China and the Third World." *Current History*, vol. 69, no. 408, 1975, pp. 75–104. JSTOR, www.jstor.org/stable/45313350.
- Lee, J. T.-H. "China's Third World Policy from the Maoist Era to the Present." *Global Asia Journal*, Paper 3, 2008. http://digitalcommons.pace.edu/global_asia_journal/3.
- Liu, Melinda. "30 Years After Tiananmen: How the West Still Gets China Wrong." *Foreign Affairs*, 4 June 2019,

- Mao, Zedong. "The Cigarette Tax." Selected Works, no.38, 29 August 1923,
www.marxists.org/reference/archive/mao/selected-works/volume-6/mswv6_08.htm.
Accessed 16 May 2024.
- Mao, Zedong. "We Must Learn to Do Economic Work." Selected Works, vol. III, p. 241, 10
January 1945,
<https://www.marxists.org/reference/archive/mao/works/red-book/ch21.htm>. Accessed 16
May 2024.
- Rowe, William. China's Last Empire: The Great Qing. Harvard University Press, 2010, p. 123.
ISBN 978067405455.
- "Timeline: What Led to the Tiananmen Square Massacre." PBS, 5 June 2019,
www.pbs.org/wgbh/frontline/article/timeline-tiananmen-square/. Accessed 16 May 2024.

Model versus Defame: Chinese American Representation in American Cinema

By Yushu Wu

ABSTRACT

This study examines the evolution of Chinese American representation in cinema during the new media era, focusing on streaming platforms and algorithmic recommendation systems. Analyzing films from major streaming services and theatrical releases between 2020 and 2024, the research finds that while new media offers increased opportunities for diversity, Chinese American representation remains disproportionately low. Content analysis demonstrates persistent reliance on stereotypical narratives and character tropes despite some progress in expanding storytelling. The study identifies tensions and paradoxes between authentic cultural representation and the risk of reinforcing stereotypes. It concludes by calling for more nuanced approaches to Chinese American storytelling, balancing cultural specificity with individual diversity, and offers recommendations for promoting authentic representations in the digital age.

Keywords: Chinese American cinema, model minority, ethnic representation, new media, stereotypes

1. INTRODUCTION

Martin Scorsese has criticized contemporary films for being reduced to mere “content,” a term he argues has become ubiquitous in the media industry (Scorsese). His essay, “Il Maestro,” seems to target new media and the surge of over-the-top (OTT) providers. Media theorist Lev Manovich defines new media as “the cultural objects which use digital computer technology for distribution and circulation” (Manovich, “New Media from Borges to HTML”). What Scorsese’s essay presses on the most is the algorithm-centric curatorship of OTT streaming sites. While it can be said that content access is more democratic in the age of new media (Shapiro), Scorsese argues that the algorithm is purely for user retention, thus disregarding users’ taste variety.

New media is novel in terms of audio-visual language and narrative, stemming from the nature of new media (Manovich, *The Language of New Media*). Portable electronics allow for the instantaneous and often unconscious making of media (Friedman and Friedman). The long-standing linear “producer to distributor to viewer” chain is challenged by new media’s mass “circulation” model, which is participatory, community-based, and user-centric (Shapiro; Jenkins et al.; Obar and Wildman). This trend extends to contemporary traditional media, affecting narratives and aesthetics (Ng). As new media grants an outlet for each user, minorities can provide accounts of their lives, engaging in activism against stereotypes and naivety (Gheyntanchi and Moghadam; Kim). This is termed more effective as audiovisual context is used authoritatively in first-impression racial cognition (Nishime; Davé). An example of ethnic activism in cinema is that of Asian Americans challenging the predominantly Caucasian Hollywood, which often entails a bigoted portrayal of Asian culture, as seen in “Yellow Peril” and “model minority” stereotypes (Johnson; Kawai). The twenty-first century’s pop visual works

have produced more content on Asian American life, including *Fresh Off the Boat* (Nagaraj and Chien; Shah). On the qualitative side, however, the series perpetuates a “model minority” complex with the Huangs; likewise, other audio-visual media remain stagnant in Asian American narratives (Cheung; Besana et al.).

On the quantitative side, as of 2021, there were 24.0 million Asian Americans; the 5.2 million Chinese Americans constituted the largest Asian American group in the U.S. (United States Census Bureau). Meanwhile, Asian Americans accounted for only 5.9% of speaking roles in major films, despite making up 7.2% of the U.S. population (Yeun et al.). Another USC Annenberg Inclusion Initiative study in 2022 indicates an overall growth in U.S. screens’ inclusiveness compared to 2007’s statistics on speaking roles (Smith, Pieper, et al.). Among the foundations left by these previous studies, we find their reviews focused mainly on theatrical releases and isolated data from social foundations on the media industry’s efforts for new media and OTT. This paper aims to provide a more comprehensive picture of the contemporary fiction-narrative films scene, examining how the turn to OTT sources brings new energy to the minority image in American cinema.

This paper explores the statistical and content disparity between Chinese and Taiwanese Americans (abbreviated as Chinese American) life, taking them as a sample of the overall Asian American representation, and their film portrayal in the age of new media. It examines a relationship between Chinese American images and streaming sites: How has the representation of Chinese Americans in films changed in the new media era under the algorithmic recommender system of streaming services? Does it contribute to a more nuanced and multifaceted depiction or reinforce previous stereotypes? It aims to achieve this purpose through the following two:

1. Analyzing the quantity and nature of Chinese American content on major streaming platforms, specifically Netflix, Amazon Prime, and Disney+.
2. Identifying key changes in Chinese American portrayals between traditional and new media.

The aforementioned objectives are to be achieved via content analysis and case studies of streaming platforms, followed by algorithm analysis, interviews, and comparative examination of Chinese American portrayal and other Asian minorities, all studied within the period between 2020 and 2024, using films as a medium. In the ensuing discussion, the paper aims to conclude with a few pragmatic notes for contemporary filmmakers.

2. LITERATURE REVIEW

New media represents a revolution in media production, distribution, and consumption; it has brought new digital narratives, aesthetics, and networks, which developed as the era of Web 2.0 emerged, creating an internet collective that allows for more cooperation with fewer technological barriers (Wilson et al.; Sfetcu). New media has remodeled the classic “producer to distributor to viewers” dynamic into individual users, most commonly, user-user, user-content

interactivities, and content creators under a “participatory culture” (Jenkins et al.; Rice; Obar and Wildman). User interfaces are customized, and contents are “circulated.”

With the technical advancements, distinctive visual narratives and tropes of new media were born, which since have become naturalized in pop culture (Jennings; Baerg). Often cited as causal, the use of new media, the internet, and ICT account for accelerating the pace of modern life (Santarius and Bergener). Therefore, in the 2000s, Bordwell proposed his intensified continuity theory based on his observations of the progression of audio-visual language becoming more intense and continuous (immersive) for cinema in the burgeoning digital age (Bordwell, “Intensified Continuity Visual Style in Contemporary American Film”; Bordwell, “Intensified Continuity Revisited”).

Content has also been altered. New media encourages collective and spontaneous snapshots of life; sub-cultures and minorities thus find a means for profiling a realism of themselves, for instance, LGBTQ+ communities, with some researchers even drawing parallels to the “dans la rue” spirit of the French New Wave (Wargo; Miao). This is attributed, partly, to the instantaneity of the internet, and also to a “circulative” method of distribution of content on the web (Lobato; Crisp). Minor productions are now able to reach more targeted, individualized groups; the visibility of lesser-known subjects is raised; popular content is more easily shared, and also metabolized faster if not appealing anymore (Ono and Pham). One company that adopts the said strategy of niche, targeting indie films, is A24, building upon a business model Miramax had started years ago, but leveraging the directive power of the web (Tryon; Perren). The minority films they distributed, two of which we will discuss in this paper, found their audience precisely.

On the Asian American minority, the model minority myth has been targeted in mainstream society via media literacy campaigns, yet the media does not project on real-life perceptions and problem-solving of the stereotypes (Owens). Some mainstream media still iterate the dated racial concepts, such as Marvel (Ka Lok Shek); model minority stereotypes persist in American cinema (Byun).

3. METHODOLOGY

As of 2024, the 3 most-user OTT sites are Netflix, Amazon Prime, and Disney+ (Durrani). Taken as samples, this research paper examines the portrayal of Chinese Americans in new media films from January 2020 to July 2024 through 2 sections, each of its own method. The paper is generally qualitative while incorporating web data collection in some parts to reinforce its content and algorithm analysis.

Section I outlines a general view of Chinese American presence in original films. It compiles the number of Chinese-American-related titles, main characters, genre distribution, and critics’ scores, taken from the Internet Film Database (IMDb), The Film Database (TMDb), and Rotten Tomatoes. The compiling of data follows the model of the USC Annenberg Inclusion Initiative’s studies (Smith, Choueiti, et al.; Yeun et al.). The data are then analyzed to construct a picture of Chinese American portrayal.

Section II follows the basics of content analysis. Three films among the three sites' Chinese American films are chosen to compare with five from theatrical releases. In adherence to the core principles of sample manageability, relevance, and representativeness, this paper will specifically select films that display Chinese American culture principally or as a secondary element; films should also be mainstream in terms of critical and market reception (Fran and Seale). However, this method is not comprehensive for theatrical releases due to an unmanageable quantity of films. Historical relevance and contemporaneity are thus factored in, as they aim to demonstrate the evolution of Chinese Americans' presence on screen. The five films should reflect recent progress and the current status of Chinese Americans' image in American cinema. This paper will cross-analyze each film's opening (exposition) and closing (resolution) sequence, following Bordwell's *Film Art: An Introduction* and content analysis techniques (Bordwell and Thompson; Li et al.; Krippendorff). While the length of the opening scene may vary, this paper, for the sake of manageability, will sample only the first and last 5 minutes that contain narrative content (omitting pure titles). The discussion of the films' text will be examined separately in terms of structural (narrative) and iconographic (aesthetics) dimensions. The paper acknowledges an extension of Bordwell's intensified continuity theory into a continuation of pre-existing norms in narratives across films that have become cliché of characters, plot, themes, etc. (Bordwell, "Intensified Continuity Visual Style in Contemporary American Film"; Harbord; Schweinitz).

The research will conclude with a general comparative study of Chinese American portrayal of other Asian minorities, as well as the image of other ethnicities in American cinema. Data on other ethnicities shall be obtained from existing research, with an intent to maximize the overlap between the researched periods. The comparison shall focus on the prominence of the minority's characters in terms of the number of speaking roles and screen time, diversity and depth of roles, and representation of cultural elements; the categories of comparison stem from the previous work of Ono and Pham (2009).

4. ANALYSIS

4.1 SECTION I: GENERAL STATISTICS ON CHINESE AMERICAN REPRESENTATION

Of the 3 providers, Netflix is the most active in issuing original films, and also the oldest, having begun in 2012; the number of titles each year up to 2020 has seen a steady rise (Afilipoaie et al.). Disney+ and Amazon Prime started their original programming in 2019. The 2020 global lockdown fueled at-home digital media consumption in the U.S., which accelerated the growth of new media film (Gupta and Singharia).

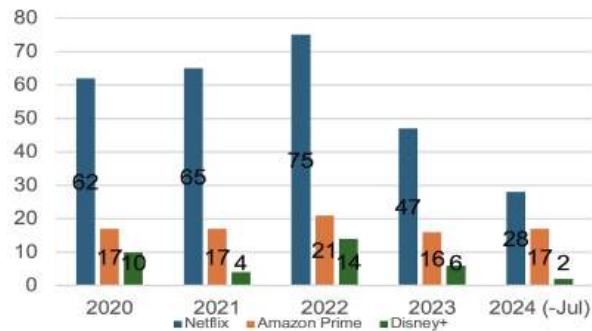


Fig. 1: Number of original film by each site, 2020-2024 July

Fig. 1 lists the number of original American English-language films by the 3 providers from Jan 2020 to Jul 2024. Out of a total of 401 films published, 277 are Netflix originals, demonstrating dominance in the original film ecosystem. Yet, in terms of percentage, Netflix’s Chinese American representation is less than Disney+’s contribution, the latter having 3 Chinese-American-related films among a total of 36, compared to Netflix’s 5 over a larger base number. Amazon is found to be absent in the making of Chinese American film content. A total of 8 Chinese-American-related films are found in this research.

Regarding the genre of Chinese American pictures, “comedy” and “family” films are the most prevalent. The genres of “drama” and “adventure” occupy a lesser portion. This aligns with the genre distribution of Netflix original films from 2015-2018 (Roig et al.). Chinese American films notably lack in the thriller category and tend to be family-related.

Upon critical reception, the mean scores of the 8 films’ public reception parameters are 6.34 (IMDb), 70 (TMDB), 82.1 (Tomatometer), and 68.7 (Popcornmeter). When compared to 2024’s Netflix’s mean Tomatometer index of 55.3 (Roeloffs), Chinese American films are more critically appealing. As of June 2021, the median IMDb score of Netflix original films was 6.40-7.05 (Luis), of which Chinese American films find themselves below.

This paper finds that for films that do not stress an engagement of Chinese American life or viewership, Chinese Americans are absent in major roles or speaking roles, while Chinese American culture can occur through symbols such as Chinatown, porcelain, and Chinese business. This paper then urges future researchers to focus on the fractional presence of Chinese Americans in major films that are not Chinese-American-centric and investigate their formulation of the ethnic picture, as they constitute a complementary part of Chinese American identity in American cinema.

On statistics, in comparison to other minorities, namely, other branches of Asian Americans, a study on the 100 most popular streaming titles in 2022 by USC Annenberg found the following regarding Asian American representation circumstances in the year:

1. The percentage of Asian American speaking characters in the top-grossing films (features) increased from 8.2% to 16.2% from 2019 to 2022.
2. “Only 24% of characters were explicitly identified as having heritage associated with a particular Asian country.
3. “68% of Asian characters never spoke to another Asian character, despite an average of 4 Asian characters per title.
4. “44% of Asian characters with a job were ‘knowledge workers’ who engaged in intellectual work rather than manual or service labor. More than half of these (54%) were in the fields of law or science, technology, engineering and mathematics (STEM)” (Giaccardi et al.).

The literature reflects a similar situation for Asian Americans – the quantity has risen, yet the quality does not match, as Asian American characters, in general, continue to see stereotypes reinforced in their identity, much like what Chinese American cinema shows as a part of the larger Asian American community. Other minorities in the United States, Black Americans, Hispanics, and Latino Americans, show mixed signs. In film casting, Black Americans occupy 11% of lead roles in comparison to their population share of 13.4%, yet most are engaged with race-related subjects, unable to jump out of a grand narrative about ethnic identity (Dunn et al.). While Black Americans see more success in quantitative results from activism and screen time, Hispanics see a poorer performance than the Asian American minorities, as their 62 million people in the U.S. comprise 19% of the country’s population, yet “Latinos hold less than 5 percent of leading on-screen, off-screen, and executive leadership roles in U.S. media” (Moslimani and Noe-Bustamante; Beccach et al.). Comparative research shows an equal or more serious underrepresentation of Latinos and African Americans in the on-screen population. For African Americans, their screen major role representation is about proportionate, yet the methods of portrayal are trite; the qualitative evaluation of Chinese American cinema, as expanded in the later Section II.

4.2 SECTION II: CASE STUDIES AND ANALYSIS

The onset of Covid-19 marked a new era for OTT films. This paper covers the period from 2020 to July 2024; 3 OTT films from this said period are chosen as representatives of new media films. 6 theatrical films made in the last 30 years are subjected to a comparative study to inspect the general trend in Chinese American portrayal in American cinema under the impact of the industry’s shift to new media.

Among the 8 films in Section I, 2 Netflix and 1 Disney+ original films are selected based on their reception and representativeness: *Turning Red* (2022), *Tigertail* (2020), and *The Half of It* (2020). The same principle applies to the selection of traditional theatrical releases, yet with more quantitative emphasis such as box office performance, critical reception, historical significance, etc.: *Joy Luck Club* (1993), *Better Luck Tomorrow* (2002), *Saving Face* (2004), *The Farewell* (2019), and *Everything Everywhere All at Once* (2022).

4.2.1 Quantitative Overview of Selections

With the aim of constructing a timeline of Chinese American images on screens over the last 30 years, I provide an overview of the theatrical titles selected and, separately, the OTT original films. This brief part aims to establish an evolutionary narrative regarding theatrical films that summarizes the transformation from *The Joy Luck Club* to the most recent A24 production, *Everything Everywhere All at Once*. This part also briefly assesses the distinctions OTT original films have made compared to their theatrical predecessors rooted in their nature of distribution.

A quantitative examination of the 5 theatrical releases shows no clear time-wise trend in their budget and gross, yet their distributors lean towards indie or arthouse: *The Joy Luck Club* by Hollywood Pictures; the last two of the selections both by A24, an independent publisher. Based merely upon this, it reveals that Chinese American content, over the years, has moved towards a more niche approach that either indicates shrinkage in Chinese American cinema or has become more mono-ethically targeted. Overall, the films, as their subject matter remains peripheral, have received increasingly complimentary public reception and shown growth in general, peaking at \$77,191,785 for *Everything Everywhere All at Once* with The Academy's Best Picture award ("Everything Everywhere All at Once (2022)").

On the OTT side, its Chinese American original films are plain and banal when compared to the epochal rise of their theatrical counterparts. OTT-exclusive releases are less statistically transparent in their financials and viewership; OTT films' profit is hard to chart as the sites' films are accessed through membership, not ticket sales. In measurable terms, streaming titles are more minority-oriented than theatrical releases (Ramón et al.). The BIPOC groups are proportionately (leads) or even overly represented (total actors) in OTT cinema. As mentioned in Section I, new media shows a more niche-oriented focus on film content, as the platform's nature grants atomization and independence, a ground for diverse ethnic representation.

4.2.2 General Content Analysis of Selections

Extending beyond quantitative views, the rest of this section presents an analysis of the theme, character, plot, opening sequence (exposition), and ending sequence (resolution) in their diversity of Chinese American portrayal. While data provides instructive insights into studying Chinese American presence, the fundamental form and content of the films are studied to offer a denser, anatomic view of the conditions of Chinese American imageries.

This section identifies films and comments that may not stand as honorable and pioneering without context. The sub-sections aim to construct a juxtaposing narrative of conforming Chinese American storytelling ideals and films that subvert conventional content and form.

4.2.2.1 Plot and Thematic Analysis

5 theatrical releases are cliché in repetitive themes on families and ancestry; their statements on rebelling, reconciling, or embracing the conservative Confucian Chinese traditions

show the Chinese American heritage's clash with progressive, liberal American values. In *Saving Face* and *Everything Everywhere All at Once*, it is lesbianism and unorthodox love. In *Joy Luck Club* and *The Farewell*, second-generation immigrants return to their ancestral soil and reconnect with their people. The films are mostly drama, comedy, and family genres. In tone, they are mellow, non-destructive, and mild. The films do not comment upon the advancement of the U.S. or the outdatedness of Chinese conservatism, even when viewed through a political-historical lens, which *Joy Luck Club* took for its depiction of the post-imperial China scenes; they propose a reconciliation of either, which fits the diversity ideal of America.

Better Luck Tomorrow is the more rebellious take on Asian Americans. The film pushes high school naivety and violence to an extreme to reveal the core theme of “there is no turning back” of adolescence. The film de-emphasizes a usual focus on Chinese American traits; its form and content serve for its coming-of-age theme rather than Chinese American matters via the narrated characters. For this reason, many argue it is not a Chinese American film but Caucasian and Hollywood at its core.

Later new media films do not present entirely new perspectives. *Tigertail* is a generational tale about crossing the Pacific and returning to the homeland, reminiscent of *The Joy Luck Club*. *Turning Red* is about a Toronto Chinese Canadian family and how the cultural lineage impacts the women of the household, especially in adolescence. *The Half of It* stands out in a way similar to *Better Luck Tomorrow*. Its portrayal of Ellie Chu, a Chinese American girl who is an outsider in a predominantly Caucasian town, no longer has “competent parents” or a major focus on the familial setting. Thematically, Ellie slowly accepts her lesbian identity, a theme Alice Wu has explored in her previous film, *Saving Face*. Yet, in the latter, the exploration of lesbianism is, to an extent, removed from a Chinese American contextualized conversation to an individual metric. No other notable exceptions worth detailing.

4.2.2.2 Character Analysis

Teen Chinese American characters are uniform in traits of “straight A’s” and studiousness, which bring about either esteem or abuse from peers. The elders can be portrayed as manipulative and expecting, yet also with profound wit. It is the middle-aged characters that show freedom in their position. In all films that in ways iterate Chinese stereotypes, they also subvert the cliché or build upon the cliché mindset, for example, the mother in *Saving Face*, who is rather reserved in parenting but entangles herself in the unorthodox plot of late-life pregnancy by a young man. Other characters who, through a time arc, redefine one’s identity include Ellie, Mei, Ben, Evelyn, etc., for whom film materials serve as revelation complexes that result in growth/character evolution. Yet, subversion or targeting of the cliché, as a color-conscious approach of filmmaking, can potentially affirm the stereotypes, for example, in *Saving Face*, the relatives frown at the mother and daughter for each’s relationship; they serve as caricatures to spotlight the protagonists, yet side characters themselves show a dated, conservative image. However, in the later films, the “model minority” ideal has diminished – characters are shown as less well-off in both digital and traditional cases, such as Evelyn, a laundromat businesswoman,

and Ellie's family, a lower-middle-class household led by an incompetent immigrant male, and Ellie has to find income in high school to pay off the bills overdue.

Regarding gender, 6 of the 8 selected films feature female protagonists, signifying the great regard filmmakers have levied on Chinese American women's experiences. It coincides in parallel with feminist cinema in that both peripheral narratives aim to challenge norms and conventions imposed, a correlation between Chinese cliché-breaking persona and modern females.

The films largely overlook the older Chinese settlers in the U.S. dating back to the 19th century, while they intend to produce a historical narrative with a heightened awareness of lineage tracing and re-establishing links with the homeland, which contradictorily left the early Chinese immigrants underrepresented. One film that signals the protagonists as earlier generations of immigrants is *Turning Red*. The Lees have lived in Toronto for several decades and run one of the oldest temples in the city. Or, it can be supposed the more Americanized and acculturated (that is to say, resemble less of a clear Chinese identity) Chinese American protagonists represent the earlier Chinese immigrants. Notable examples are Ellie from *The Half of It* and several in *Better Luck Tomorrow*, both focus more on teenage life, moving away from a fundamental discussion of Chinese Americans' familial and identity theme. *Chang Can Dunk*, a Disney+ original film also takes this attitude. Into the 21st century, films' content indicates a furthering of Chinese American character diversity, which can grow less relevant to Chinese ideals.

4.2.2.3 General Content

Upon politics, most do not explicitly comment on the deteriorating Sino-American relations and their impact upon individuals, except in *The Farewell*, which dedicated sections that subject Billi to inquiries about her view of the cross-cultural experience between the two nations. This, in its contemporary context. *The Farewell* captures the mindset of a modern Chinese immigrant to the United States that is less explored in other works – the U.S. continues to be a symbol for the Chinese middle class, as a land of the American dream, ideals, and opportunities that enable at the very least a betterment and potentially a transition in one's socioeconomic class. This is briefly depicted when Hongkongers Evelyn and Waymond make their way to the U.S. and purchase a laundromat. Yet, in both cases, the families are distant from the women in *The Joy Luck Club*. Reiterating here is a decline in the model minority portrayal of Chinese Americans, with film depictions of Chinese American lives returning to more basic and everyday experiences. What is clear is that these films fail, both independently and collectively, to paint a comprehensive picture of the Chinese Americans' state of life in the country. For the most part, they remain individual case studies, a representation dilemma.

Still, no film is not political. *The Joy Luck Club* takes a pro-American perspective on the Chinese American characters and their journey to the United States. This film is both an activist film for Chinese Americans and for women, with the latter taking an even larger role, as the weight placed on female characters is so prominent. The warmth of U.S. images contrasts with

the coldness and lean, almost sparse cinematography of scenes in China, and being an immigrant story, the American ideal of betterment forms the foundation of the film's core representation of the Chinese American experience.

Over the years, OTT films have not shown many evolving characters. While Confucianism versus Western ideals and lifestyle remain themes, along with nostalgia in these films, OTT demonstrates more alternative takes on the Chinese American identity. While *Better Luck Tomorrow* stands as more singular, we have *Chang Can Dunk* and *The Half of It*, two OTT pictures that collaborate to build a more fluid Chinese American image relative to the more culturally dense films that are didactic in handling Chinese American materials. Whether individually their character portrayal is superficial, Chinese American filmography altogether contributes to a complex account of ethnicity.

4.2.3 Content Analysis: Openings/Expositions

A film's opening usually sets the tone and attitude with which the work deals with Chinese American identity and life. Several traits can be summarized in the attitudes taken and the form of narrative used to present exposition and conflict-building.

Better Luck Tomorrow's opening 5 minutes establish the film's unconventional approach to Chinese American subject matter. The use of jump cuts, fast-paced editing, freeze frames, and intermissions in the exposition of various Asian American characters lays the groundwork for a highly dynamic story with a diverse array of Chinese American teens of varying mental states. The film opens with Virgil and Ben's self-contrasting conversation about college, as they are smoking. The Asian stereotypical studiousness is here contradicted by a rebellious cigarette, foreshadowing the unconventional life the protagonist teenagers will lead, which is later revealed through flashbacks. The serenity is broken by suspense, and the film transitions into a tale of their high school life. Few Chinese American films in this selection are intensified at the start with such a sense of suspense and high stakes.

Turning Red opens with a montage of photos accompanied by a monologue describing the family's expectations of their teenager, a creative method that establishes a glimpse of Mei's identity and obedience as part of the familial piety, seen in the first line: "Honor your parents." The respect for parents, being a theme, is presented without much obfuscation. The exposition portrays Mei in a balanced state between being herself and serving her parents, fulfilling their needs. She is a typical Asian student with a limited group of friends and is academically qualified (especially in math). The conflict of the film is also set – the coming of age of a girl through the menarche, which will disrupt her life balance, symbolized by a euphemist-signifier curse of the red panda, that remarks to a transition and an inciting incident. The opening ends with the Daisy Mart scene, in which the boy Mei is attracted, their first love is introduced, and the first disturbance to their ordinary and predictable life. While these symbols act common to all adolescent themes, the perspectives, and urgency they are captured within *Turning Red* embody them of Chinese characters that inform the viewers immediately of the film's intention.

Other films feature more conventional, familial openings. Both *The Joy Luck Club* and *Saving Face* open with familial dining scenes, initially focusing on the crowd and eventually narrowing in on the protagonist. In the case of the former, this leads to the main storyline of June and her return to China. The film opens with a monologue recounting a tale about a feather traveling from China to the U.S. “from afar and carrying with it all my good intentions.” The imagery is impressionist, with a Chinese freehand brushwork style, and the white background has a texture reminiscent of rice paper, all evoking an impression that the film is about China and Chinese immigrants. The presentation of the family dinner is American – featuring American football; the lines are inter-distorted and inaudible, serving as background noise. Yet, the camera movement directs the audience’s attention to the continuous presence of June, who takes the initiative, like a guide, in leading the camera around the household, providing an exposition of the scenario. The scene then focuses on the photo of the 8 women, the protagonists of this film, whom the film narrates parallelly in the rest.

In *Saving Face*, the opening is a montage of a day in the life of Wil, an aspiring young surgeon. In terms of content, it is similar to *The Joy Luck Club* in that it immediately establishes the character as a female who has an intellectual life and career, subverting the stereotype of a housewife and suppressed female identity. Yet, Wil is immediately bound by a more conventional side of her Chinese heritage – a set-up relationship with a man, with her mom not being aware of her lesbian identity. As we are introduced to the protagonist, Wil, as a guide as well, leads the camera to observe the Chinese club, and she, being a character with independence surviving in the city, is dragged back to her heritage-related obligations, which she is reluctant to be a part of. The hero is thus presented as being out-of-norm and, in some ways, rejecting and excluded by the social environment in which she is placed.

Ellie in *The Half of It* is also rejected by her social environment, but in this case, as a minority without Chinese community support, she is the lone kid who is bullied by racists and exploited for her academic capabilities. She is not popular in school, similar to Mei. In the aforementioned four films, the protagonists, who are all females, are subjected to some form of social pressure, either from the Chinese side of the family and/or from the minority factors posed by the setting. In almost all films, the opening establishes a heroic arc and the identity-fostered problems that will drive the narrative forward, including *Everything Everywhere All at Once*, which introduces Jobu Tupaki, or Joy, as a rebelling lesbian teenager, who is not accepted by the more conservative members of the household and in turn drives the familial side of the sci-fi story to fruition, as a fundamental motif of the plot.

Among all the films, *Tigertail*’s opening is the least focused on the common theme of individuals versus Chinese identity. We immediately recognize a flashback structure when we see a match cut of the protagonist, as a young man in Taiwan, to when he is old, sitting in a car, spoken to in English by his driver, which establishes the Chinese American context. Yet, unlike other films, the conflict is not immediately apparent in the opening. The pain of the protagonist is clear – the traumatizing absence of his parents during his childhood, yet fond memories of the

girl back in Taiwan – as a film that spans a greater time period, it is more subtle and gradual in revealing the protagonist’s core struggles and his mission when returning to his homeland.

The films in their opening can be categorized according to traits that primarily depict individuals versus environment, either as a minority in a Caucasian-dominant setting or a minority relative to the Chinese culture to which one belongs but has lost connection and attachment. Films in the latter category often present a rebellious, out-of-norm, well-adapted-to-American society character that stands out against familial expectations, while the former depicts a character more marginalized within American life. The relationship between character traits and Chinese culture may not be a specific point of focus in the film, and thus, may be absent in the opening, but stereotypical traits remain that are challenged as the film progresses. *Tigertail*’s opening is nostalgic, and it does not emphasize the duality of the Western and Eastern aspects of a character’s life that often lead to seemingly uncompromising conflicts. Other films more or less employ some stereotypes and then revolt against them, or they discuss the Chinese American identity theme explicitly, laying the groundwork for later development based on the clash of lifestyles across cultures in the opening sequence.

4.2.4 Content Analysis: Ending

Parallel to the conflict and stakes set in the opening, films often conclude with a state of self-acceptance, revolt, and expansion of perspective. To illustrate this point, *The Joy Luck Club* ends with June returning to China to find her sisters, not knowing their mother had died based on the letters. As she returns, she reconnects and serves as one of the most immediate messengers delivering the saddest news for a child – the passing of their mother. The distance between the two groups, separated by the ocean, is bridged as they find a certain parallel in fate and interest, connected by their mother as the roots of their family. Their family, although broken by war, experiences a physical and emotional reunion at this moment, and June, bearing the news, is able to find resonance with two cousins she had never met.

Similarly, as we discussed, *The Farewell* concludes with Billi embracing her homeland, although she was not initially repulsive towards its concept in the opening. Moreover, she is eager to return for her dying grandmother. The force opposing her travel is her family, who, we learn later, had leaned toward American ideals when they immigrated, yet are also conflicted by their familial obligation to let the elder have peaceful moments before passing. Her embracing is thus more akin to a discovery of the life she has grown estranged from, reaffirming her Chinese identity along the way. If we are to identify a pattern of this sort of Chinese American narrative, it is a typical return to the ancestral identity one is imbued with at birth. Or, we can argue that almost all Chinese American films adopt a positive theme in reaffirming the presence of one’s past and culture, despite it being unfamiliar and altered, for example, in *Tigertail*, where the old house ideal is shattered by the newer constructions. This paper shall refrain from passing judgment on these observations at present.

Regarding gender, the conservative notions in the films typically show characters opening up to new perspectives, as we see in *Everything Everywhere All at Once*, *Saving Face*,

The Half of It, and *Turning Red*, where they either confront the conservative and cliché aspects of their relatives and families or find acceptance of others' atypical acts and viewpoints through the journey of the narratives. Among the four named, three are to some extent comedic and lighthearted in the end; *The Half of It* takes a more serious and insightful approach to the transformation of Ellie. It's worth noting these four films all have female protagonists, indicating how females' confrontation with conservative mindsets is an underlying theme explored in this selection of films.

Better Luck Tomorrow remains a film that stands out, yet its uniqueness has been discussed extensively in this paper. Its re-establishment of a teenage-centric, not ethnicity-specific approach in treating the film materials is parallel to that seen in *The Half of It*, contributing to its dynamic spirit compared to the general field of Chinese American portrayals studied over the years.

5. CONCLUSION

Across the data and films examined, the researcher has drawn the following findings:

1. New media original films manifest a greater ability to represent ethnic diversity compared to theatrical films, achieving about proportional or even overrepresentation of BIPOC groups.
2. Chinese Americans do not receive a quantitatively proportionate level of representation on OTT sites. Of the three OTT sites studied, Disney+ shows the greatest initiative in incorporating Chinese American characters or life, followed by Netflix, and then Amazon Prime, which features no Chinese American lead characters or main plot/life elements.
3. In the selection, theatrical releases generally show more mundane and conformist approaches in stressing the family and conservative notions of Chinese American life and make efforts to discuss how individuals are affected under a grander narrative of Chinese versus American cultural dynamics. In some ways, these discussions reaffirm the model minority stereotype.
4. There exist exceptions to the aforementioned point. Films centered around Chinese American characters can sometimes avoid a discussion of family values and cultural heritage, allowing individuals to be free from the cultural bonds they are usually associated with, which can be stereotypical in a modern context. Yet, there are criticisms of this approach to filmmaking, arguing that it assimilates ethnicity into a non-cultural or White American state of mentality.
5. Stereotypes have evolved across the years in terms of characters and their situations, shifting from model minority to more marginal families of the ethnicity, which have become the focus of the screen. The genres are more diverse, and the parameters of each production are more nonconforming and inventive.

Overall, the ways that Chinese American plots are dealt with show an unsatisfactory stagnancy over the years since the earliest film in this research, *The Joy Luck Club*, a film from 30 years ago that had already started venturing into a conversation about the relation of 1st and 2nd generation immigrants to their cultural lineage. While some exceptions have emerged over the years, the mainstream exploration of Chinese American topics, or films that are praised for accurate or loyal portrayals of the Chinese American experience, appear to be confined to previous topics that had been established.

Drawing from these findings, the paper concludes that Chinese American films are often too conventional and conformist in handling ethnic-cultural materials. While it can be argued that a culture does not experience substantial variation but maintains a distinct narrative, the last three decades have seen dramatic shifts in the lived experiences of Chinese Americans, and continuing to discuss the same subjects as before feels repetitive when viewed from a historical point of view.

It is also found that films about the older generations of Chinese immigrants are less frequently portrayed in all pictures. Alternatively, it can be interpreted that the films that show no clear Chinese cultural characters are portrayals of these groups of the population. While they do provide alternatives to the conventional routes Chinese American cinema has taken, they are less related to Chinese culture and the Chinese diaspora in general, thus reducing the presence of Chinese lifestyle in their ways of living. Here, the two conditions of Chinese American life clash as they generate a possible point of contention regarding the purpose of such Chinese American cinema, a reiteration of the question asked, but now expanded upon: To whom and how are Chinese American films made?

From the artistic perspective, the paper suggests a more diverse portrayal of Chinese American characters that, sometimes, can mean a departure from a larger thematic, plot-wise thesis on Chinese culture thriving on American soil, yet providing a more dynamic viewing experience while still retaining realism or allowing artistic expression. The discussion section will expand upon this suggestion and further explore the ontology of Chinese American cinema, aiming to find a more complex response to these questions.

6. DISCUSSION

A natural question to ponder is, “Is there, to any extent, a general method for representing Chinese Americans or any minorities?” From this, we understand the awkward position minority narratives are in peripherally, as these groups, being less dominant, may demand a more natural and accurate account of their life to be revealed to the public, yet their films’ purpose should not fall into a fight solely to defend against the suppression over the dominant narratives, resulting in furthering a “good image” of themselves into a state of model minority. The terms “marginalized group” or “people of color” inherently suggest a natural counterforce to white dominance, which may lead to a disregard for the diversity within the minority as a whole.

There are two primary approaches minority narratives can take. One is to adopt a strong minority initiative, emphasizing the subordinate social status of these groups to the dominant,

and showing either a group or an individualistic revolt against the stratified class. Alternatively, a film can choose to neglect the minority status of the subjects and normalize their presence without praise or criticism of the attitudes towards the characters. Both methods are political in dealing with minorities; no method is apolitical in minority narratives. This dilemma persists unless minorities gain dominance, at which point the diversity and variability within the group can be freely and comfortably portrayed in films.

This discussion can be further explored in the special issue *Introduction: Peripheral Narratives, Minority Identities*, which examines the formation of minority identity through media and societal exposure, a process called minoritization (Moenandar and Godioli). This thus makes any contact between minorities and majorities a boundary-meeting matter embedded with power dynamics, as previously explained, which means any sort of cinema on Chinese Americans can be subsumed into an oppressive and stereotypical category.

Therefore, comes a constant tension of accusations. Characters' balance in their presentation of Chinese American traits can either be too Chinese-focused and thus cliché, as they inevitably, in trying to find an accurate Chinese life and mentality, fall into stereotypical territory (as these portrayals are not necessarily false, but over-discussed and repeated in past cinema); or, they are subjected to criticism for being assimilated to white culture, reflecting internalized racism (Trieu). The inevitable mentality of marginalized groups in a choice of acculturation or enculturation and the conflict between the individual and the social group that one identifies with are contributing to the confusion regarding one's Chinese American or minority status (David and Derthick). The book *Internalized Oppression* suggests a positive growth of psychological well-being for a person who enculturates oneself and stands firm with their cultural heritage, yet in the case of Chinese Americans, immobility in the factions of identity commonly discussed is the main concern, as it contributes to non-diverse ethnic statics.

The Joy Luck Club stands as a groundbreaking work in that it pioneers the activism of both women and Chinese Americans. It creates relatively new images of Chinese Americans and women in their confrontation with the past and with males while promoting a feminist ideal of independence. However, the theme has ceased to evolve, potentially exposing the masses to a risk similar to the dilemma of the "model minority," where Asian Americans are conscious of the stereotypes imposed but are more-or-less pushed to follow the expectations of being a model (Chou and Feagin 2015)(Chou and Feagin). In this manner, while Chinese Americans do embrace a sense of cultural identity, it is a stereotypical identity lost in context; the enculturation turns out to be more negative as it imposes a boundary for Chinese Americans to live within.

This criticism of the acculturation of Chinese Americans emerged at Sundance in 2002 regarding *Better Luck Tomorrow*, with critics arguing that its portrayal of Chinese Americans was extremely violent and rebellious. A question was raised asking Justin Lin whether he had considered his responsibility in representing the Asian American community and whether that manner of representation would bring about negative impacts. In response, Roger Ebert stated, "Asian American characters have the right to do whatever the hell they want." This comment liberates the film from the burden of representing the ethnicity itself – it is highly individualistic

and not concerned with realism in a larger scope or the necessity of presenting a righteous image of a group.

This discussion can extend to an examination of the social ontology of film. Bazin's realism theories still apply in some aspects here, but the scope of such realism is becoming less materialistic in modern commercial films. Instead, their themes, discussions, core messages, etc., are realistic, even if the content is eclectic. While the majority of content is real, isolated cases are realistic simultaneously, and there is no doubt that these isolated cases should not be abandoned in cinematic conventions. Their existence establishes a sense of majority and minority in storytelling or diversity within a concentration. This majority and minority dynamic, as previously discussed, can stem from American cinema in general. While considerations of ethnicity occupy a delicate position in the contemporary socio-political context, "minority" should not be a tag with a subtext of power; filmmakers intending to handle minority subjects, if possible, should always empower the minority by presenting unexpected and anti-stereotypical portrayals of their life and characters, which are free and non-dependent on stereotypes themselves. They should be internally driven, self-proving their right to exist.

The necessity for Chinese Americans or any minority cinema to produce films featuring less traditional narratives is undeniable. However, the extent to which these non-traditional cases should be mainstream or promoted depends on the films' purpose. A primary criticism of films depicting acculturated individuals is that they are made for white Americans. Yet, the author proposes an alternative explanation – films that iterate on the diversity and cultural clash between two cultures are primarily made for the white audience. As non-minorities, they do not experience the "individualistic versus collective" conflict stemming from the cultural shift experienced by 1st or 2nd generation immigrants (here, the author generalizes the majority as a cohesive body, while acknowledging they are also divided into diverse cultural phenomena; the assumption is that a majority adheres to the functions of a bounded society, conforming within itself to a standardized way of living, which for the U.S., is American). These cultural and, consequently, political films first respond to the majority's need for voyeurism of others' experiences, which is largely limited to their perspective, as well as confirming the notion of national diversity, a growing concept of togetherness and embodiment of a modern internationalist state.

Works Cited

- Afilipoaie, Adelaida, et al. "The 'Netflix Original' and What It Means for the Production of European Television Content." *Critical Studies in Television: The International Journal of Television Studies*, vol. 16, no. 3, Sept. 2021, pp. 304–25. *DOI.org (Crossref)*, <https://doi.org/10.1177/17496020211023318>.
- Baerg, Andrew. "Reviewed Work: Always Already New: Media, History, and the Data of Culture by Lisa Gitelman." *Symplokē*, vol. 15, no. 1/2, 2007, pp. 383–85.
- Beccach, Camilo, et al. "Latinos in Hollywood: Amplifying Voices, Expanding Horizons." *McKinsey & Company*, 7 Mar. 2024, <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/latinos-in-hollywood-amplifying-voices-expanding-horizons>.
- Benshoff, Harry M., and Sean Griffin. *America on Film: Representing Race, Class, Gender, and Sexuality at the Movies*. Third edition, Wiley Blackwell, 2021.
- Besana, Tiffany, et al. "Asian American Media Representation: A Film Analysis and Implications for Identity Development." *Research in Human Development*, vol. 16, no. 3–4, Oct. 2019, pp. 201–25. *DOI.org (Crossref)*, <https://doi.org/10.1080/15427609.2020.1711680>.
- Bordwell, David. "Intensified Continuity Revisited." *Observations on Film Art: Kristin Thompson and David Bordwell*, 27 May 2007, <https://www.davidbordwell.net/blog/2007/05/27/intensified-continuity-revisited/>.
- . "Intensified Continuity Visual Style in Contemporary American Film." *Film Quarterly*, vol. 55, no. 3, Mar. 2002, pp. 16–28. *DOI.org (Crossref)*, <https://doi.org/10.1525/fq.2002.55.3.16>.
- Bordwell, David, and Kristin Thompson. *Film Art: An Introduction*. 10. ed, McGraw-Hill, 2013.
- Byun, Brooke. *Asian American Representation in Media: An Interpretive Analysis of the Consumer*. 2021. University of Washington, Honors Thesis, <https://com.uw.edu/wp-content/uploads/2021/08/Asian-American-Representation-In-US-Media.pdf>.
- Cheung, Pricilla Holing. *Asian American Representation and Stereotypes on Mainstream American Television: A Case Study of Fresh Off the Boat*. 2019. School of Culture and Communication of Aarhus University, MA Thesis. *DOI.org (Datacite)*, <http://rgdoi.net/10.13140/RG.2.2.19351.83367>.
- Chou, Rosalind S., and Joe R. Feagin. *Myth of the Model Minority*. 0 ed., Routledge, 2015. *Crossref*, <https://doi.org/10.4324/9781315636313>.
- Crisp, Virginia. *Film Distribution in the Digital Age: Pirates and Professionals*. Palgrave Macmillan, 2015.
- Davé, Shilpa S. "Accents and Asian American Representation in Contemporary Culture." *Oxford Research Encyclopedia of Literature*, by Shilpa S. Davé, Oxford University Press, 2020. *DOI.org (Crossref)*, <https://doi.org/10.1093/acrefore/9780190201098.013.792>.

- David, E. J. R., and Annie O. Derthick. "What Is Internalized Oppression, and So What?" *Internalized Oppression*, edited by E. J. R. David, Springer Publishing Company, 2013, pp. 1–32, <https://doi.org/10.1891/9780826199263.0001>.
- Dunn, Jonathan, et al. "Black Representation in Film and TV: The Challenges and Impact of Increasing Diversity." *McKinsey & Company*, 11 Mar. 2021, <https://www.mckinsey.com/featured-insights/diversity-and-inclusion/black-representation-in-film-and-tv-the-challenges-and-impact-of-increasing-diversity>.
- Durrani, Ana. "Top Streaming Statistics In 2024." *Forbes*, 13 June 2024, <https://www.forbes.com/home-improvement/internet/streaming-stats/#:~:text=Netflix%20still%20has%20the%20most,impressive%20260.28%20million%20subscribers%20globally>.
- "Everything Everywhere All at Once (2022)." *Box Office Mojo*, <https://www.boxofficemojo.com/title/tt6710474/>. Accessed 9 Sept. 2024.
- Fran, Tonkiss, and Clive Seale. "Content and Comparative Keyword Analysis." *Researching Society and Culture*, Fourth edition, Sage, 2012. *BnF ISBN*, <http://eprints.lse.ac.uk/id/eprint/59103>.
- Friedman, Linda Weiser, and Hershey H. Friedman. "The New Media Technologies: Overview and Research Framework." *SSRN Electronic Journal*, 2008. *DOI.org (Crossref)*, <https://doi.org/10.2139/ssrn.1116771>.
- Gheytauchi, Elham, and Valentine N. Moghadam. "Women, Social Protests, and the New Media Activism in the Middle East and North Africa." *International Review of Modern Sociology*, vol. 40, no. 1, Spring 2014, pp. 1–26.
- Giaccardi, Soraya, et al. *A Balancing Act for Asian Representation in Streaming Visibility Doesn't Always Mean Cultural Specificity*. USC Annenberg Norman Lear Center, Nov. 2023, https://learcenter.s3.us-west-1.amazonaws.com/NLC_GH+Asian+Representation+in+Streaming.pdf.
- Gupta, Garima, and Komal Singharia. "Consumption of OTT Media Streaming in COVID-19 Lockdown: Insights from PLS Analysis." *Vision: The Journal of Business Perspective*, vol. 25, no. 1, Mar. 2021, pp. 36–46. *DOI.org (Crossref)*, <https://doi.org/10.1177/0972262921989118>.
- Harbord, Janet. "The Potency of Film Editing: Rose Hobart Stop Return." *Contemporary Theatre Review*, vol. 25, no. 1, Jan. 2015, pp. 68–72. *DOI.org (Crossref)*, <https://doi.org/10.1080/10486801.2015.992245>.
- Jenkins, Henry, et al. *Spreadable Media: Creating Value and Meaning in a Networked Culture*. New York University Press, 2013.
- Jennings, Pamela. "Narrative Structures for New Media: Towards a New Definition." *Leonardo*, vol. 29, no. 5, 1996, p. 345. *DOI.org (Crossref)*, <https://doi.org/10.2307/1576398>.

- Johnson, Laura (Laura Renee). *Asian and Asian American Representations in American Film*. 2004. Western Washington University, WWU Honors Program Senior Projects. 210., https://cedar.wwu.edu/cgi/viewcontent.cgi?article=1210&context=wwu_honors.
- Ka Lok Shek, David. *Mockery, Deviants, Aliens, and Asians in Marvel Films: A Media Analysis of Asian Representation*. 2022. Aalborg University, https://projekter.aau.dk/projekter/files/473810243/10_Semester_THESIS.pdf.
- Kawai, Yuko. "Stereotyping Asian Americans: The Dialectic of the Model Minority and the Yellow Peril." *Howard Journal of Communications*, vol. 16, no. 2, Apr. 2005, pp. 109–30. *DOI.org (Crossref)*, <https://doi.org/10.1080/10646170590948974>.
- Kim, Dasol. "The Growing Up Asian American Tag: An Asian American Networked Counterpublic on YouTube." *International Journal of Communication*, vol. 15, 2021, pp. 123–42.
- Kim, Minjeong, and Rachele J. Brunn-Bevel. "Hollywood's Global Expansion and Racialized Film Industry." *Humanity & Society*, vol. 44, no. 1, Feb. 2020, pp. 37–66. *DOI.org (Crossref)*, <https://doi.org/10.1177/0160597619832045>.
- Krippendorff, Klaus. *Content Analysis: An Introduction to Its Methodology*. SAGE Publications, Inc., 2019. *DOI.org (Crossref)*, <https://doi.org/10.4135/9781071878781>.
- Li, Ying, et al. "Movie Content Analysis, Indexing and Skimming Via Multimodal Information." *Video Mining*, edited by Azriel Rosenfeld et al., Springer US, 2003, pp. 123–54. *DOI.org (Crossref)*, https://doi.org/10.1007/978-1-4757-6928-9_5.
- Lobato, Ramon. *Netflix Nations: The Geography of Digital Distribution*. New York University Press, 2019. *Open WorldCat*, <https://www.jstor.org/stable/j.ctv12pnnk3>.
- Luis. *Netflix Original Films & IMDB Scores*. Kaggle, 2021, <https://www.kaggle.com/datasets/luiscorter/netflix-original-films-imdb-scores/data>.
- Manovich, Lev. "New Media from Borges to HTML." *The New Media Reader*, edited by Noah Wardrip-Fruin et al., The MIT Press, 2003. *K10plus ISBN*, http://manovich.net/content/04-projects/033-new-media-from-borges-to-html/30_article_2001.pdf.
- . *The Language of New Media*. 1st MIT Press pbk. ed, MIT Press, 2002. *Library of Congress ISBN*, https://dss-edit.com/plu/Manovich-Lev_The_Language_of_the_New_Media.pdf.
- Miao, Ruomu. "'Cinema Was Becoming More Real': From the French New Wave to the Vertical Screen Aesthetics of Mobile Short Video." *Revista FAMECOS*, vol. 30, no. 1, Oct. 2023, p. e44854. *DOI.org (Crossref)*, <https://doi.org/10.15448/1980-3729.2023.1.44854>.
- Moenandar, Sjoerd-Jeroen, and Alberto Godioli. "Introduction: Peripheral Narratives, Minority Identities." *Frontiers of Narrative Studies*, vol. 8, no. 2, Jan. 2023, pp. 133–38. *DOI.org (Crossref)*, <https://doi.org/10.1515/fns-2022-2017>.
- Moslimani, Muhamad, and Luis Noe-Bustamante. "Facts on Latinos in the U.S." *Pew Research Center*, 16 Aug. 2023, <https://www.pewresearch.org/race-and-ethnicity/fact-sheet/latinos-in-the-us-fact-sheet/>.

- Nagaraj, Sriganeshvarun, and Puu Wen Chien. "Asian Stereotypes: Asian Representation in Hollywood Films." *INTI Journal*, vol. 2020, no. 63, Dec. 2020, <http://eprints.intimal.edu.my/id/eprint/1485>.
- Ng, Waiping. *People's Tendency of Watching Films Has Already Been Influenced the Development of Information Age: Advances in Social Science, Education and Humanities Research*, 2021, pp. 2830–36. *DOI.org (Crossref)*, <https://doi.org/10.2991/assehr.k.211220.491>.
- Nishime, LeiLani. "Mixed-Race Asian Americans and Contemporary Media and Culture." *Oxford Research Encyclopedia of Literature*, Oxford University Press, 2020. *DOI.org (Crossref)*, <https://doi.org/10.1093/acrefore/9780190201098.013.854>.
- Obar, Jonathan A., and Steve Wildman. "Social Media Definition and the Governance Challenge: An Introduction to the Special Issue." *Telecommunications Policy*, vol. 39, no. 9, Oct. 2015, pp. 745–50. *DOI.org (Crossref)*, <https://doi.org/10.1016/j.telpol.2015.07.014>.
- Ono, Kent A., and Vincent N. Pham. *Asian Americans and the Media*. Polity, 2009.
- Owens, Lydia Susan. *Perceptions of Asian American Representation in Entertainment Media*. 2020. West Virginia University, Graduate Theses, Dissertations, and Problem Reports, <https://researchrepository.wvu.edu/cgi/viewcontent.cgi?article=8778&context=etd>.
- Perren, Alisa. *Indie, Inc: Miramax and the Transformation of Hollywood in the 1990s*. 1st pbk. ed, University of Texas Press, 2013.
- Ramón, Ana-Christina, et al. *Hollywood Diversity Report 2024: Featuring Film; Part 2: Streaming*. UCLA College, Institute for Research on Labor & Employment, 2024, <https://socialsciences.ucla.edu/wp-content/uploads/2024/06/UCLA-Hollywood-Diversity-Report-2024-Film-Streaming-5-23-2024.pdf>.
- Rice, Ronald E., editor. *The New Media: Communication, Research, and Technology*. Sage Publications, 1984.
- Roeloffs, Mary Whitfill. "Critically Panned Movies 'Mother Of The Bride,' 'Unfrosted' Top Netflix Charts." *Forbes*, 15 May 2024, <https://www.forbes.com/sites/maryroeloffs/2024/05/15/critically-panned-movies-mother-of-the-bride-unfrosted-top-netflix-charts/>.
- Roig, Antoni, et al. "Netflix Fictional Feature Film Originals: An Analysis of Release Strategies." *Communication & Society*, Apr. 2021, pp. 125–39. *DOI.org (Crossref)*, <https://doi.org/10.15581/003.34.2.125-139>.
- Santarius, Tilman, and Jens Bergener. "Does the Use of ICT Speed up the Pace of Life?" *Proceedings of the 7th International Conference on ICT for Sustainability*, ACM, 2020, pp. 107–15. *DOI.org (Crossref)*, <https://doi.org/10.1145/3401335.3401642>.
- Schweinitz, Jörg. *Film and Stereotype: A Challenge for Cinema and Theory*. Columbia University Press, 2011.
- Scorsese, Martin. "Il Maestro: Federico Fellini and the Lost Magic of Cinema." *Harper's Magazine*, no. March 2021, Mar. 2021, pp. 24–32.

- Sfetcu, Nicolae. *Small Business Management for Online Business - Web Development, Internet Marketing, Social Networks*. MultiMedia Publishing, 2014. *DOI.org (Crossref)*, <https://doi.org/10.58679/tw86436>.
- Shah, Hemant. “‘Asian Culture’ and Asian American Identities in the Television and Film Industries of the United States.” *SIMILE: Studies In Media & Information Literacy Education*, vol. 3, no. 3, Aug. 2003, pp. 1–10. *DOI.org (Crossref)*, <https://doi.org/10.3138/sim.3.3.002>.
- Shapiro, Andrew L. *The Control Revolution: How the Internet Is Putting Individuals in Charge and Changing the World We Know*. PublicAffairs, 1999.
- Smith, Stacy L., Marc Choueiti, et al. *Inequality in 1,200 Popular Films: Examining Portrayals of Gender, Race/Ethnicity, LGBTQ & Disability from 2007 to 2018*. USC Annenberg Inclusion Initiative, Sept. 2019, <https://assets.uscannenberg.org/docs/aai-inequality-report-2019-09-03.pdf>.
- Smith, Stacy L., Katherine Pieper, et al. *Inequality in 1,600 Popular Films: Examining Portrayals of Gender, Race/Ethnicity, LGBTQ+ & Disability from 2007 to 2022*. USC Annenberg Inclusion Initiative, Aug. 2023, <https://assets.uscannenberg.org/docs/aai-inequality-in-1600-popular-films-20230811.pdf>.
- Trieu, Monica M. “Understanding the Use of ‘Twinkie,’ ‘Banana,’ and ‘FOB’: Identifying the Origin, Role, and Consequences of Internalized Racism within Asian America.” *Sociology Compass*, vol. 13, no. 5, Mar. 2019, <https://doi.org/10.1111/soc4.12679>.
- Tryon, Chuck. *On-Demand Culture: Digital Delivery and the Future of Movies*. Rutgers University Press, 2013.
- United States Census Bureau. *Asian American, Native Hawaiian and Pacific Islander Heritage Month: May 2023*. 3 Mar. 2023, <https://www.census.gov/newsroom/facts-for-features/2023/asian-american-pacific-islander.html>.
- Wargo, Jon M. “‘Every Selfie Tells a Story ...’: LGBTQ Youth Lifestreams and New Media Narratives as Connective Identity Texts.” *New Media & Society*, vol. 19, no. 4, Apr. 2017, pp. 560–78. *DOI.org (Crossref)*, <https://doi.org/10.1177/1461444815612447>.
- Wilson, David W., et al. “Web 2.0: A Definition, Literature Review, and Directions for Future Research.” *AMCIS 2011 Proceedings - All Submissions*, no. Paper 368, Aug. 2011, https://aisel.aisnet.org/amcis2011_submissions/368. AIS Electronic Library (AISeL).
- Yeun, Nancy Wang, et al. *The Prevalence and Portrayal of Asian and Pacific Islanders across 1,300 Popular Films*. USC Annenberg Inclusion Initiative, May 2021, https://assets.uscannenberg.org/docs/aai_aapi-representation-across-films-2021-05-18.pdf.

Challenges and Recent Advancements in Deep Space Craft Design Regarding Galactic Cosmic Radiation Protection By João Pedro O. Toledo Piza

Abstract

Humans are inherently curious, and it is clear that our next step is to venture and make a new home for humanity outside of Earth. Given that no planet on the solar system is suitable for long-term housing of humans (without severe terraformation and/or human adaptation), if we are to become an extraterrestrial species we must look outwards, outside of our solar system for a new home. Interstellar space, however, is an incredibly hazardous environment. Specifically, the galactic cosmic radiation that hits spacecraft is ionizing, which can cause severe health problems. This paper looked at the issues and advances in radiation protection systems for deep space spacecraft design and recommends the use of cryo-hydrogen shielding for large vessels and electrostatic membranes for small vessels, in addition to keeping the cryoprotective drug amifostine as a possible treatment for any imparted effects of radiation.

Keywords

Physics and Astronomy; Nuclear and Particle Physics; Radiation; Deep Space Exploration; Manned Spacecraft; Deep Space Radiation

Introduction

On Earth, humans are protected from solar and cosmic radiation by Earth's^{6,12}. Furthermore, 2 of the 3 cardinal principles of radiation protection (time, distance, and shielding) cannot be applied in space²¹. Time cannot be reduced since the voyage is bound to take excessive amounts of time, and distance cannot be reduced since Galactic Cosmic Rays (GCR) are isotropic, this leaves only shielding as a way to protect the crew of an interstellar voyage²⁰.

The effects of ionizing radiation on the human body can be severe. According to the World Health Organization, radiation can "impair the function of tissues and organs", as well as produce many adverse health conditions, including death⁸. According to NASA, exposure to GCR can lead to cancer and degenerative diseases¹². All of this leads to the conclusion that any manned deep-space mission must have radiation shielding.

Earth's magnetosphere and atmosphere provide shielding from solar and GCR, but the same cannot be said for interstellar travel, as discussed above. Furthermore, shielding is the only way to protect a spacecraft in deep space from GCR. Below, 4 possible solutions to this issue are discussed: cryo-hydrogen, plasma shield, electrostatic membranes, dietary supplements, and passive material shielding. These 4 solutions were picked from modern-day papers following up on "Shielding from cosmic radiation for interplanetary missions: Active and passive methods"⁹.

The methods were compared via a quantitative and qualitative analysis of the minimum areal density required for their radiation protection to fall above the baseline. This paper has drawn from the technical brief NASA-STD-3001 as a baseline, which sets the maximum yearly

exposure of 600 millisieverts (mSv)⁷. Solutions were measured in their areal density needed to protect a human from a dose equivalent of 600 mSv.

Methods

Below are all the methods analyzed in this paper, as described in the introduction.

Passive Material Shielding

Material shielding is the use of certain materials that block ionizing radiation around the living and radiation-sensitive areas of a ship¹. The most common type of material shielding in space or weight-critical applications is hydrogen or hydrogen-rich materials¹⁷. Hydrogen is specially used because it fragments heavy ions found in GCR, and slows down neutrons formed when GCR interacts with matter¹¹. For this reason, polyethylene (a common plastic) and kevlar are relatively effective given their weight and simplicity, which is why kevlar is used as radiation shielding in the International Space Station (ISS)¹⁶. Polyethylene requires an 86.2 (kg/m²) areal density to maintain safe levels of radiation exposure, while aluminum requires 130 (kg/m²)⁴.

Cryo-Hydrogen

Another possible material shielding is cryo-hydrogen⁴. This method involves the freezing of hydrogen gas for increased density in liquid or solid states³. The benefit of this method is increased protection but at a much higher cost. Another benefit is that hydrogen can be used for final burns in case needed¹⁸.

The thermal protection needed to maintain the system at temperatures below 20 degrees Kelvin (the point at which hydrogen becomes liquid), is another factor to be considered³. The areal density of the cryogenic thermal shielding is 30 (kg/m²)⁴. A cryocooler is suggested as a backup, which requires 1.4 kWh to be operated, with the added mass being less than 1 (kg/m²). This system can work indefinitely as long as it has a stable energy source, and even without an energy source, the thermal shielding can maintain the system at stable temperatures for up to 3 years. Altogether, the areal density of the system (cryocooler included) must be 38.2 (kg/m²) to “reduce the radiation exposure of astronauts to allowable limits”¹⁸.

One important consideration, however, is that there has only been one major study to discuss the possibility of cryo-hydrogen as a radiation protection device. Given that the data collected here was not done so experimentally, only through simulations, there might be some discrepancies with reality.

Electrostatic Membranes

Electrostatic shielding is the use of artificial electromagnetic fields, similar to Earth's, to protect a spacecraft from GCR⁵. The method discussed here involves the deployment of inflatable charged gossamer structures¹⁴. To put it simply, these are deployable structures that form a large magnetic field, protecting the ship in its center^{2,14}. This is advantageous given that the safe zone created between the fields is incredibly safe from radiation¹³. One complication is

that some instruments that depend on a stable electromagnetic field, or lack thereof, might suffer with this instrument, and large ships need tremendously high voltages, which become unrealistic^{13,14}.

Biological Shielding

Biological shielding is the practice of shielding organic life forms from radiation via medicine²². One 2013 paper discusses the idea of dietary supplements that can help with exposure to protons and HZE particles, which can prevent " radiation-induced detrimental effects even when applied several days after radiation exposure" ²². Of the 15 agents suggested in the article, only 1 (amifostine) is FDA-approved^{10,20}. Amifostine, however, should only be treated as a countermeasure, and not as the main line of defense against cosmic radiation¹⁹.

This approach is incredibly light and cheap, especially compared to the previous methods mentioned above²². The downside, however, is that it only works well as a treatment, as its preemptive use yields very limited results¹⁹. Another downside is that radiation-critical structures cannot be protected, since they are not made of organic matter²⁰. To conclude, the low cost and weight of this approach, paired with its effectiveness as medication to be applied after radiation exposure, helps make its case as a secondary protection mechanism, to be used in spacewalks or if other protection devices fail¹⁹.

It is also important to mention that one of the core papers used here ("Deep space missions and the issue of overcoming the problem of space radiation") which analyzes amifostine as a possible radiation protection device, while peer-reviewed, did not publish their data or explain their methodology.²² Furthermore, another paper ("Pharmacological Agents for the Prevention and Treatment of Toxic Radiation Exposure in Spaceflight"), even though published 5 years before, presents much more solid evidence, including their methodologies and results. Another important factor there considered was how safe (how many negative side effects) these agents were to the human body. For this reason, this paper has considered "Pharmacological Agents for the Prevention and Treatment of Toxic Radiation Exposure in Spaceflight" as a better indicator of the state-of-the-art in biological radiation protection, and so has drawn data from that paper more extensively.^{19,22}

Methodology

In an effort to better understand how the methods mentioned above can be compared, this section explores the methodology here used to determine each method's characteristics.

To ensure realistic data, all data from this paper was collected from peer-reviewed journals. The collected data in this paper is divided into a quantitative and a qualitative analysis. While all of the possible solutions were evaluated holistically in qualitative analysis, only the cryo-hydrogen, aluminum, and Polyethylene were analyzed quantitatively. The reason for dietary supplements not being analyzed quantitatively is due to their lack of physical properties. The reason for electrostatic shielding not being analyzed quantitatively is due to its unorthodox scaling properties. In an effort to lower these possible discrepancies in the data, specially when

comparing radiation protection methods quantitatively, passive material shielding data and cryo-hydrogen data were both collected from the same paper.

Results

Table 1: Radiation Protection methods* ranked by areal density, and sourced		
Methods	Areal Density (kg/m ²)	Sources
Cryo-hydrogen	38.2	(18)
Polyethylene	86.2	
Aluminum	130.0	
* Only shows radiation protection methods mentioned above		

Table 2: Radiation Protection methods and respective advantages and disadvantages, sourced			
Method	Advantages	Disadvantages	Sources
Cryo-Hydrogen	<ul style="list-style-type: none"> - Efficient - Light - Requires very little electricity - Occupies a large space - Possible use as propeller fuel - Great loadbearing capabilities 	<ul style="list-style-type: none"> - Requires electricity - Little research 	(18)
Aluminum	<ul style="list-style-type: none"> - Commonly used: A lot of research, practice - Not as expensive as other methods - Great loadbearing capabilities 	<ul style="list-style-type: none"> - Very heavy - Occupies a large space 	(16) (18) (11)
Polyethylene	<ul style="list-style-type: none"> - Very cheap - Commonly used: A lot of research, practice 	<ul style="list-style-type: none"> - Very heavy - Occupies a large space - Not safe for bearing load 	
Dietary Supplements (Amifostine)	<ul style="list-style-type: none"> - Light - Cheap - Physically small - Possibly used to treat other medical issues 	<ul style="list-style-type: none"> - Only usable as a treatment - Not very efficient 	(19) (22)
Electrostatic Membrane	<ul style="list-style-type: none"> - Physically small - Expensive - Requires large amount of electricity 	<ul style="list-style-type: none"> - Complex: large chance of error - Very expensive to scale 	(14)

	<ul style="list-style-type: none"> - Protects astronauts in extravehicular activities - Due to the inverse square law and its unorthodox scaling, gets cost per m³ covered can decrease as m³ increases, depending on the case. 		
--	---	--	--

Table 3: Qualitative analysis of electrostatic membranes compared with cryo-hydrogen as radiation shields, sourced		
Advantages of electrostatic membranes	Advantages of cryo-hydrogen	Sources
Attenuates more particles	Less energy is needed than electrostatic membranes	(18) (14) (13)
Lighter	Easier to scale (More effective in larger vessels)	(2) (3)
Cheaper		
Smaller size		
More effective in smaller vessels		
Better loadbearing capabilities		

From Table 1, it is clear to see that cryo-hydrogen is the method with the lowest areal density by far, being around 3.4x less dense than aluminum and 1.5x less dense than Polyethylene.

From Table 2, it is possible to understand that aluminum and polyethylene are not viable materials, leaving amifostine, electrostatic membranes, and cryo-hydrogen. Given that amifostine cannot be used as a primary radiation shield.

From Table 3, it is possible to understand the advantages and disadvantages in the two methods that showed the biggest potential from Table 2.

Discussion

Discussion of sources has been presented in their respective sections and in "Methodology".

Conclusion

As seen above, it becomes clear that active electrostatic shielding and cryo-hydrogen are the best options for deep space radiation protection. When compared to biological shielding, it is

much more expensive and dense, but, unlike biological shielding, works as a pre-emptive measure and works on non-organics. Amifostine, however, is the only one that can help somewhat reverse the effects of radiation exposition.

Due to the inverse-square law, the cryo-hydrogen shielding is much more effective the larger the ship, depending on how the shape of the ship's outer shell, and the fact that the "power requirement [of the electrostatic membranes] becomes increasingly more challenging as the spacecraft voltage is increased", which is needed in larger vessels. On the other hand, more cryocoolers might be needed depending on the surface area. Thus, for larger vessels, cryo-hydrogen becomes the better option.

However, when considering smaller vessels, electrostatic membranes are much more efficient than cryo-hydrogen. This is because, at smaller voltages, electrostatic membranes block much more radiation, even if they require more energy. One important challenge that must be covered for their use, nevertheless, is that they require either large electric storage solutions or an active source of electricity.

In conclusion, this paper suggests the use of cryo-hydrogen shielding for large vessels and electrostatic membranes for small vessels, in addition to keeping amifostine as a possible treatment for any imparted effects of radiation. This space, however, still demands empirical evidence of these observations, with each of these methods being tested in deep space-like conditions.

Acknowledgments

I would like to thank Dr. Steve Bullock and Dominique Paladini for their help revising this paper and helping me with the idealization, and Dr. Kieran Tait for his help with source acquisition. Additionally, I would like to thank Indigo Research for their help.

Works Cited

1. Kanagaraj, B.; Anand, N.; Raj, S.; Lubloy, E. Advancements and Environmental Considerations in Portland Cement-Based Radiation Shielding Concrete: Materials, Properties, and Applications in Nuclear Power Plants– Review. *Cleaner Engineering and Technology* **2024**, *19*, 100733. <https://doi.org/10.1016/j.clet.2024.100733>.
2. Moore, C. L. Gossamer Space Structures Panel. In *2003 IEEE Aerospace Conference Proceedings (Cat. No.03TH8652)*; 2003; Vol. 8, p 8_3993-8_3994. <https://doi.org/10.1109/AERO.2003.1235583>.
3. Moradi, R.; Groth, K. M. Hydrogen Storage and Delivery: Review of the State of the Art Technologies and Risk and Reliability Analysis. *International Journal of Hydrogen Energy* **2019**, *44* (23), 12254–12269. <https://doi.org/10.1016/j.ijhydene.2019.03.041>.
4. Li, X.; Mustafi, S.; Boutte, A. Preliminary Thermal Design of Cryogenic Radiation Shielding; 2015.
5. Tripathi, R. K.; Wilson, J. W.; Youngquist, R. C. Electrostatic Space Radiation Shielding. *Advances in Space Research* **2008**, *42* (6), 1043–1049. <https://doi.org/10.1016/j.asr.2007.09.015>.
6. Su, J.; Wei, Z.; Liu, G.; Xu, J.; Fei, T.; Zhou, B.; Li, Y.; Pan, Y.; Gao, D.; Han, H. Simulation of Radiation Environment and Design of Multilayer Radiation Shield for Orbital Exploration of Jupiter. *Advances in Space Research* **2024**, *73* (5), 2652–2662. <https://doi.org/10.1016/j.asr.2023.12.006>.
7. NASA; OCHMO. *Design for Ionizing Protection*; Technical Brief NASA-STD-3001; NASA, 2022.
8. World Health Organization. *Radiation and health*. <https://www.who.int/news-room/questions-and-answers/item/radiation-and-health> (accessed 2024-08-04).
9. Spillantini, P.; Casolino, M.; Durante, M.; Mueller-Mellin, R.; Reitz, G.; Rossi, L.; Shurshakov, V.; Sorbi, M. Shielding from Cosmic Radiation for Interplanetary Missions: Active and Passive Methods. *Radiation Measurements* **2007**, *42* (1), 14–23. <https://doi.org/10.1016/j.radmeas.2006.04.028>.
10. Citrin, D.; Cotrim, A. P.; Hyodo, F.; Baum, B. J.; Krishna, M. C.; Mitchell, J. B. Radioprotectors and Mitigators of Radiation-Induced Normal Tissue Injury. *Oncologist* **2010**, *15* (4), 360–371. <https://doi.org/10.1634/theoncologist.2009-S104>.
11. Thibeault, S.; Fay, C.; Lowther, S.; Earle, K.; Sauti, G.; Kang, J. H.; Park, C.; McMullen, A. Radiation Shielding Materials Containing Hydrogen, Boron, and Nitrogen: Systematic Computational and Experimental Study - Phase I; 2012.
12. NASA. *Why Space Radiation Matters*. <https://www.nasa.gov/missions/analog-field-testing/why-space-radiation-matters/> (accessed 2024-08-04).

13. Tripathi, R. K.; Wilson, J. W.; Youngquist, R. C. Electrostatic Active Radiation Shielding - Revisited. In *2006 IEEE Aerospace Conference*; IEEE: Big Sky, MT, USA, 2006; pp 1–9. <https://doi.org/10.1109/AERO.2006.1655760>.
14. Tripathi, R. Meeting the Grand Challenge of Protecting Astronaut ' s Health : Electrostatic Active Space Radiation Shielding for Deep Space Missions Final Report; 2012.
15. Palaszewski, B. Solid Hydrogen Experiments for Atomic Propellants. In *36th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*; American Institute of Aeronautics and Astronautics: Las Vegas,NV,U.S.A., 2000. <https://doi.org/10.2514/6.2000-3855>.
16. Narici, L.; Casolino, M.; Di Fino, L.; Larosa, M.; Picozza, P.; Rizzo, A.; Zacont, V. Performances of Kevlar and Polyethylene as Radiation Shielding On-Board the International Space Station in High Latitude Radiation Environment. *Sci Rep* **2017**, 7 (1), 1644. <https://doi.org/10.1038/s41598-017-01707-2>.
17. Iguchi, D.; Ohashi, S.; Abarro, G. J. E.; Yin, X.; Winroth, S.; Scott, C.; Gleydura, M.; Jin, L.; Kanagasegar, N.; Lo, C.; Arza, C. R.; Froimowicz, P.; Ishida, H. Development of Hydrogen-Rich Benzoxazine Resins with Low Polymerization Temperature for Space Radiation Shielding. *ACS Omega* **2018**, 3 (9), 11569–11581. <https://doi.org/10.1021/acsomega.8b01297>.
18. Li, X.; Mustafi, S.; Boutte, A. J. Cryogenic Hydrogen Radiation Shield for Human Spaceflight. US10144535B2, December 4, 2018. <https://patents.google.com/patent/US10144535B2/en> (accessed 2024-08-04).
19. Langell, J.; Jennings, R.; Clark, J.; Ward, J. B. Pharmacological Agents for the Prevention and Treatment of Toxic Radiation Exposure in Spaceflight. *aviat space environ med* **2008**, 79 (7), 651–660. <https://doi.org/10.3357/ASEM.2113.2008>.
20. Sihver, L. Physics and Biophysics Experiments Needed for Improved Risk Assessment in Space. *Acta Astronautica* **2008**, 63 (7–10), 886–898. <https://doi.org/10.1016/j.actaastro.2008.04.013>.
21. Kim, J. H. Three Principles for Radiation Safety: Time, Distance, and Shielding. *Korean J Pain* **2018**, 31 (3), 145–146. <https://doi.org/10.3344/kjp.2018.31.3.145>.
22. Mortazavi, S. M. J.; Mozdarani, H. Deep Space Missions and the Issue of Overcoming the Problem of Space Radiation. *International Journal of Radiation Research* **2013**, 11 (3), 199–202.

GenExpressivity: The impact of rights to freedom of expression on adolescents' mental health outcomes in the United States

By Anya Khera

Abstract

The right to freedom of expression is a cornerstone of adolescent development. Yet, its acceptance in the context of school, digital, and home settings presents significant challenges in the United States. This paper explores the association between adolescents' rights to freedom of expression and the effects on their mental health outcomes, emphasizing the critical role that supportive environments play in enhancing well-being. Through a systematic review of existing literature and case studies, this paper examines the complex dynamics of expression rights in education, digital platforms, and family settings. The findings suggest that adolescents who are empowered to express their identities, values, and opinions freely experience better mental health, including lower levels of depression and anxiety. However, these benefits are contingent on the presence of environments that balance freedom with safety. Schools, online platforms, and families that fail to protect or encourage expression may contribute to adverse mental and emotional well-being. The paper concludes by offering evidence-based recommendations to enhance the rights and opportunities for self-expression among youth in the United States, aiming to foster favorable mental health outcomes.

GenExpressivity The impact of rights to freedom of expression on adolescents' mental health outcomes in the United States

Introduction

"Congress shall make no law...abridging the freedom of speech..." This powerful assertion from the First Amendment to the United States Constitution enshrines the importance of free expression as a fundamental human right. However, in the context of American schools, the right to freedom of expression for adolescents is a contentious issue, between protection and regulation. The United States Court of Appeals for the Third Circuit in *Sypniewski v. Warren Hills Regional Board of Education* (2002) asserted, "There is no constitutional right to be a bully...Students cannot hide behind the First Amendment to protect their 'right' to abuse and intimidate other students at school." This highlights the delicate balance between safeguarding free speech and maintaining a safe educational environment where adolescents feel protected and seen.

The right to freedom of expression is crucial for adolescents at a time when their mental health is at risk. Over the past decade, adolescents in the United States have experienced escalating rates of mental and emotional problems, with statistics showing an alarming rise in depression, anxiety, and suicidality. In 2021, 20% of adolescents aged 12–17 had at least one major depressive episode during the prior year. Emergency department visits for deliberate self-harm by children increased by 329% between 2007 and 2014, and youth suicides rose by over 30% between 2014 and 2017. Among adolescents, females are particularly vulnerable, with the prevalence of major depressive episodes being more than twice as high (29%) compared to

males (12%).

Adolescents' rights to freedom of expression are recognized as critical to their development. The First Amendment protects students' rights to express their opinions on matters of public concern without fear of retribution. The landmark case *Tinker v. Des Moines Independent Community School District* (1969) established that students do not "shed their constitutional rights to freedom of speech or expression at the schoolhouse gate." This decision underscores the importance of protecting students' rights to express themselves as long as it does not disrupt the educational process. However, subsequent cases, such as *Bethel School District v. Fraser* (1986) and *Morse v. Frederick* (2007), have introduced exceptions to *Tinker*, limiting students' free speech rights under certain circumstances.

Understanding the impact of adolescents' rights to freedom of expression on their mental health outcomes is vital. Restrictions on self-expression can hinder adolescents' ability to form their identities and voice their opinions, contributing to increased stress, anxiety, and depression. Conversely, environments that support and encourage free expression can enhance adolescents' mental health and well-being by fostering a sense of belonging, esteem, and self-actualization.

This paper explores the complex relationship between adolescents' freedom of expression and mental health outcomes in the U.S., focusing on three key environments: schools, social media, and the home. By analyzing case studies, existing policies, and psychological research, this study aims to shed light on how the right to self-expression can either protect or harm adolescents' mental well-being. Through this examination, we will argue that fostering environments where adolescents can freely express their identities, values, and opinions — without the fear of retribution — can significantly improve their mental and emotional well-being.

Methodology

Research Design

In this literature review, a systematic approach was employed to synthesize relevant publications on youth mental health and associated freedom of expression, in schools, social media, and home settings. The literature research was conducted across multiple databases, including PubMed, ScienceDirect, Google Scholar, and SpringerLink. The search terms and keywords used included 'youth mental health,' 'freedom of expression,' 'student rights,' 'social media,' 'digital age,' 'home,' and 'adolescent mental health.' Boolean operators were applied to combine keywords and refine search results.

The following inclusion and exclusion criteria were used to select the studies for review:

- Inclusion criteria: Studies published between 2000 and 2024, peer-reviewed articles, and literature focused on youth mental health, freedom of expression, student rights, and the impact of social media and family.
- Exclusion criteria: Non-peer-reviewed sources, such as editorials, blogs, and opinion pieces, studies not available in English, and studies that did not focus directly on youth

mental health or freedom of expression in the digital age.

A multi-step process was followed to ensure a comprehensive and unbiased selection of literature. First, the titles and abstracts of all retrieved articles were screened. After an initial screening, full-text articles were assessed for relevance to the research question. Duplicates were removed, and a final list of 50 studies was selected for review.

The selected studies were analyzed using a thematic synthesis approach. Studies were grouped based on the topic of the current paper, and key patterns, trends, and themes were identified. Discrepancies between studies were highlighted and discussed where relevant.

This review is limited by the exclusion of non-English language papers and studies older than 2000, and by the fact that only peer-reviewed publications were considered, potentially excluding valuable insights from gray literature.

Synthesis

The extant literature has shown that adolescents' rights to freedom of expression have significant and positive effects on their mental health and well-being in the United States. The reviewed studies indicate that environments where adolescents feel safe to express themselves — whether in schools, online, or at home — are linked to lower levels of anxiety and depression and higher levels of self-esteem and psychological well-being. Schools that promote free expression contribute to better mental health outcomes, while restrictive policies are associated with psychological distress. In online settings, social media provides vital platforms for identity formation and self-expression, but the risks of cyberbullying and harassment underscore the need for balanced protections. At home, open communication with parents correlates with reduced depression and anxiety, though this dynamic varies based on cultural, socioeconomic, and family factors. Overall, the studies emphasize that freedom of expression is crucial for adolescents' development, yet its effectiveness depends on supportive environments and the mitigation of associated risks.

Theme 1 School Context

The right to freedom of expression significantly impacts adolescent students' mental health outcomes in schools. A 2023 study highlights that adolescents who feel their voices are heard in school are less likely to experience anxiety and depression. The researchers surveyed over 1,200 students across various U.S. high schools, measuring levels of anxiety, depression, and self-esteem and students' perceived freedom to express themselves in the classroom. The study findings showed that schools fostering environments where students can freely express their thoughts and opinions contribute to better mental health outcomes, with students reporting lower levels of stress and higher levels of self-esteem and overall psychological well-being (Novis, 2023).

Schools can provide students with avenues for self-expression through classroom discussions, extracurricular activities, and creative outlets such as art, music, and drama

programs. Schools that encourage participation in these activities not only allow students to express their unique identities but also help them develop critical thinking skills, confidence, and resilience. For example, debate clubs allow students to articulate their views on various topics, while art programs give them the freedom to explore their creativity. When students feel they can share their ideas without judgment or fear of punishment, they are more likely to experience a sense of belonging and personal fulfillment, which can lead to better academic and emotional outcomes.

This finding is supported by a 2024 briefing paper, which argues that affirming First Amendment rights in schools is crucial for youth development and mental well-being (Huddleston, 2024). The paper emphasizes the role of schools as spaces where students not only learn academically but also develop their sense of identity through self-expression. Encouraging students to communicate their viewpoints through written assignments, presentations, or even school-wide constructs like student councils empowers them to take ownership of their learning experiences. Restricting students' expression within school settings can lead to feelings of isolation, helplessness, and anxiety, ultimately impacting their mental health negatively. An article by Bouchrika also stresses the importance of upholding student rights to foster a positive educational environment (Bouchrika, 2024). Schools that limit student expression may inadvertently contribute to adverse mental health outcomes by stifling creativity, critical thinking, and personal development. The article discusses the legal policies that govern student speech in schools, noting that overly restrictive measures can create a climate of fear and repression, discouraging students from sharing their ideas. It advocates for educational practices that encourage student participation, linking these practices to improved academic performance and mental health.

A synthesis of these studies indicates that when rights to free expression are respected in schools, students exhibit improved mental health and overall well-being. Moreover, schools that provide active platforms for expression such as student newspapers, open forums, or cultural venues allow students to share their perspectives on social, political, and personal topics. Conversely, educational environments that suppress student expression tend to correlate with higher levels of psychological distress. A 2023 study focuses on student psychological outcomes, noting the direct correlation between expression and mental health (McDermott, 2023). It provides evidence that supportive school climates, where teachers encourage open dialogue and creative expression, are crucial for mitigating anxiety and depression among adolescents. Additionally, another article emphasizes the educational and developmental benefits of freedom of expression, arguing that school policies should prioritize protecting these rights (Huddleston, 2024). Bouchrika suggests practical approaches for schools to nurture student expression, such as incorporating student feedback into decision-making processes, offering opportunities for peer collaboration, and creating forums for unstructured discussion. The article underscores the importance of balancing discipline with freedom, ensuring students feel safe expressing themselves without fear of punishment.

Critically, while the studies converge on the positive effects of freedom of expression,

they also highlight the complexity of implementation within schools. The 2023 study cautions that the mere provision of expression rights is insufficient without supportive and inclusive school cultures. Schools must also train teachers and staff to create safe spaces where students feel comfortable voicing their opinions, whether in classroom or extracurricular settings. Research.com highlights successful case studies where schools have implemented supportive practices, such as arts programs and student-led initiatives, resulting in better mental health and academic outcomes.

Overall, these findings underscore the necessity of balancing freedom of expression with supportive environments to enhance adolescents' mental health in educational settings. Schools that succeed in this equilibrium create environments where students feel valued and heard, leading to improved mental well-being, self-expression, and academic success.

Theme 2 Online Platforms

Freedom of expression is a critical contributor to the development and mental health of adolescents in the United States. This freedom encompasses the use of online platforms (hereafter social media), which in turn, can lead to public expression of opinions and personal expression through forms such as fashion. These mechanisms provide adolescents with platforms for communication and identity formation, which are crucial for their mental well-being. However, the complexities inherent in these freedoms also pose significant challenges, influencing adolescents' mental health in multifaceted ways.

Social media offers opportunities to express opinions publicly, and this is another crucial aspect of adolescents' rights. A report by Edison Lanza underscores the importance of protecting young people's rights to voice their opinions (Lanza, 2019). This freedom is essential for their personal development and active participation in society, for example through political discussion and voting. However, the report by Lanza also acknowledges the potential negative repercussions of public expression. Adolescents who express their opinions online can become targets of online harassment and abuse, which can severely affect their mental health. Balancing the protection of free speech with safeguards against harassment is, therefore, essential to support adolescents' mental well-being.

Social media has also amplified forms of personal self-expression, such as fashion. Fashion allows teenagers to explore and express their identities, as shown by research from a 2010 study, which highlights that adolescents use clothing as a tool to communicate individuality and autonomy (Miles et al., 2010). By choosing what to wear, teens assert their personal tastes, cultural affiliations, and social positions, fostering a sense of self-confidence and belonging. Approximately 25% of adolescent online posts are related to fashion and personal style, reflecting its importance in shaping their identities. By showcasing their clothing and cosmetic choices online, teens receive validation and feedback from peers, which can boost their self-esteem.

Comparing these studies reveals a nuanced picture of how freedom of expression impacts adolescents' mental health. Lanza's report addresses the legal and societal aspects of public

expression, highlighting the need for a supportive legal framework to protect young voices. Meanwhile, the 2010 study takes a more personal approach, considering the everyday choices adolescents make, such as their fashion, and how these can impact their sense of self and mental well-being.

Critically analyzing these studies reveals a complex interplay between freedom of expression and mental health. While expression can significantly enhance self-esteem, autonomy, and social connectedness, it also comes with risks. The unregulated nature of social media can exacerbate mental health issues through exposure to negative content and cyberbullying for content as personal as clothing and appearance. Similarly, while public expression of opinions is fundamental for personal growth and societal participation, it must be balanced with protections against harassment to safeguard adolescents' mental well-being.

Theme 3 Home Context

The ability of adolescents to freely express themselves within their home environments is critical to their mental and emotional health and overall well-being. This theme explores how the right to freedom of expression within the family unit impacts adolescents' mental health outcomes, focusing on the perceived safety of expressing their thoughts and feelings to their parents. By examining this dynamic, we can better understand the role of familial support in fostering healthy emotional development and identify potential barriers that might hinder open communication.

A study on adolescent emotional health and communication with parents, published on Springer Link, delves into the relationship between adolescents' emotional health and their ability to communicate with their parents. The study finds that open and supportive communication correlates with lower levels of depression and anxiety among adolescents (Magalhães, Calheiros, & Antunes, 2017). Adolescents who feel safe to express themselves are more likely to experience positive mental health outcomes, such as higher self-esteem and lower stress levels.

Prior research on youth risk behavior and mental health has confirmed the importance of a supportive home environment in mitigating potentially dangerous behaviors among adolescents (Perou et al., 2013). Adolescents reporting strong parental support and open communication are less likely to engage in risky behaviors and more likely to report higher life satisfaction and mental well-being. This aligns with the findings from the 2017 study, further underscoring the critical role of open communication in adolescent mental health (Magalhães et al., 2017).

An article by Rosenthal discusses the legal frameworks supporting adolescents' rights to freedom of expression and their implications for mental health, examining the balance between parental authority and adolescents' self-expression (Rosenthal, 2019). Legal protections for freedom of expression play a crucial role in empowering adolescents, which positively influences their mental health outcomes by fostering a sense of autonomy and self-worth. This perspective adds a legislative dimension to the discussion, highlighting how rights and laws can

support mental health outcomes.

A 2009 study investigates the impact of parent-adolescent relationships on mental health outcomes, focusing on communication patterns. It finds that positive communication is a significant predictor of mental health (Chandra et al., 2009). Adolescents who perceive their communication with parents as open and non-judgmental report lower levels of psychological distress and higher levels of emotional well-being. This study complements the findings of the other studies, emphasizing the importance of supportive communication.

The studies reviewed collectively illustrate the significant role that open communication and perceived safety in expressing oneself in the family context have on adolescents' mental health. They emphasize that a supportive home environment can significantly mitigate mental health issues such as anxiety and depression, and enhance overall emotional well-being. However, these studies also reveal several complexities and limitations that warrant a more in-depth critical analysis.

Firstly, while the studies consistently highlight the benefits of open communication, they often simplify the dynamic by assuming uniformity in what constitutes "open" or "supportive" communication. The quality of communication between parents and adolescents can be highly variable and influenced by numerous factors, including cultural norms, socioeconomic status, and individual family dynamics. For instance, what is considered supportive in one cultural context may be perceived as intrusive or inadequate in another (Magalhães et al., 2017). This variation complicates the ability to generalize findings across diverse populations and underscores the need for culturally sensitive approaches to assessing and fostering effective communication.

The studies also tend to overlook the impact of socioeconomic factors on communication and mental health outcomes. Economic stressors, such as financial instability or job loss, can strain familial relationships and limit the capacity for supportive interactions (Perou et al., 2013). Adolescents from lower socioeconomic backgrounds might experience different communication patterns compared to their more affluent peers, which can affect their mental health differently. This disparity suggests that while the right to freedom of expression is crucial, its effectiveness and the level of perceived safety may be compromised by socioeconomic barriers that impede open dialogue and emotional support.

The 2017 article by Rosenthal provides an important perspective by highlighting the role of legal frameworks in supporting adolescents' rights to freedom of expression. While legal protections are vital, they are often insufficient on their own to ensure meaningful self-expression within the home (Rosenthal, 2017). Legal rights must be complemented by practical support systems and family-based interventions to be effective. The gap between legal protections and their real-world application can be significant, and there is a need for policies that not only protect rights on paper but also facilitate their practical realization in everyday family interactions.

Furthermore, the studies reveal that while open communication is associated with positive mental health outcomes, the relationship is not always straightforward. Adolescents who

feel safe to express themselves at home are indeed more likely to report lower levels of psychological distress, but there may be underlying factors that influence this relationship. For example, adolescents who are more resilient or have strong social support networks outside the family may fare better even in less supportive home environments (Chandra et al., 2009). Conversely, those who are more vulnerable or lack external support may not benefit as much from open communication at home. This variability suggests that the impact of parental communication on mental health is part of a broader, multifaceted interplay of individual, familial, and social factors.

Given these complexities, there is a pressing need for targeted interventions that address specific barriers to effective communication and support within families. Programs designed to enhance family communication skills, offer parenting support, and address socioeconomic challenges can help bridge the gap between the theoretical benefits of open expression and its practical implementation. Furthermore, research should focus on developing tailored approaches that consider the diverse needs of different families and communities to ensure that all adolescents have the opportunity to experience supportive and open communication.

In summary, while the importance of freedom of expression within the home for adolescents' mental health is well-supported by the studies, the practical realities of achieving effective communication are more nuanced. Addressing the gaps identified through these studies — such as the influence of cultural and socioeconomic factors, the application of legal protections, and the variability in mental health outcomes — requires a comprehensive approach that combines legal rights with practical, culturally sensitive, and targeted support systems. Overall, the right to freedom of expression within the home is vital for adolescents' mental health, with supportive parental relationships being a key factor in fostering a safe environment for self-expression.

Conclusion

This paper aimed to explore the impact of the right to freedom of expression on adolescents' mental health outcomes across different settings in the United States, including schools, digital environments, and homes. Review of the literature revealed that freedom of expression plays a critical role in adolescent development, contributing significantly to their psychological well-being. In school settings, environments that promote open expression were linked to lower levels of stress, higher self-esteem, and overall better mental health outcomes. Social media platforms offer opportunities for connection, self-expression, and identity-building. At home, open and supportive communication between parents and adolescents was generally associated with lower levels of anxiety and depression, highlighting the importance of a safe environment for self-expression.

The findings also underscore the importance of balancing the right to freedom of expression with supportive environments. In schools, for example, while it is crucial to uphold students' rights to express themselves, it is equally important to cultivate inclusive and supportive cultures that nurture this expression. Similarly, in digital settings, protecting

adolescents' rights in social media is essential, but there must also be safeguards against the potential harms associated with online expression. At home, fostering open and non-judgmental communication between parents and adolescents is vital for ensuring that the right to self-expression leads to positive mental health outcomes.

The implications of these findings are far-reaching. They highlight the need for advocacy and policy changes that prioritize the mental health of adolescents by protecting their rights to freedom of expression. Schools and policymakers must work together to create educational and digital environments that not only allow but encourage open expression while also providing the necessary support systems to mitigate potential risks. This includes implementing programs that teach students about safe and responsible online behavior and creating school policies that protect students without stifling their voices. Advocacy efforts should also focus on increasing awareness among parents about the importance of open communication with their children and providing resources to help families navigate the challenges of adolescent self-expression.

However, this review also revealed gaps and limitations in the literature. Many studies focused primarily on the benefits of freedom of expression without adequately addressing the complexities and challenges of implementing this right in diverse settings. There is a need for more research on how cultural, socioeconomic, and familial factors influence adolescents' ability to express themselves freely and the subsequent impact on their mental health. Additionally, while legal frameworks supporting freedom of expression are important, there is a lack of research on how these laws are applied in everyday family and school interactions.

Future research should focus on exploring these gaps, particularly the role of diversity, equity, and inclusion in shaping the dynamics of adolescent self-expression and mental health. Research should also examine the effectiveness of various interventions designed to promote healthy expression in schools, social media, and homes. Additionally, studies should investigate how to better integrate legal protections with practical support systems to ensure that all adolescents, regardless of their background, can safely and effectively express themselves. By addressing these areas, future research can contribute to more comprehensive and effective strategies for supporting adolescent mental health through the protection of their rights to freedom of expression.

Works Cited

- Casey Foundation. (2023, August 28). Social media's concerning effect on teen mental health. The Annie E. Casey Foundation.
<https://www.aecf.org/blog/social-medias-concerning-effect-on-teen-mental-health>
- Centers for Disease Control and Prevention. (2013). Mental health surveillance among children—United States, 2005–2011. *Morbidity and Mortality Weekly Report*, 62(Suppl 2), 1-35. <https://www.cdc.gov/mmwr/preview/mmwrhtml/su6202a1.htm>
- Frison, E., & Eggermont, S. (2017). Browsing, posting, and liking on Instagram: The reciprocal relationships between different types of Instagram use and adolescents' depressed mood. *Child Indicators Research*, 10(4), 1053-1070. <https://doi.org/10.1007/s12187-017-9511-6>
- Gattis, M. N., Sacco, P., & Cunningham-Williams, R. M. (2019). Substance use and mental health outcomes among homeless youth: A latent class analysis. *Child Indicators Research*, 12(1), 271-289. <https://doi.org/10.1007/s12187-017-9511-6>
- Gould, M. S., & Lake, A. M. (2009). The contagion of suicidal behavior. *Journal of Adolescent Health*, 45(4), 317-319. <https://doi.org/10.1016/j.jadohealth.2009.07.003>
- Hurd, N. M., & Zimmerman, M. A. (2010). Natural mentors, mental health, and risk behaviors: A longitudinal analysis of African American adolescents transitioning into adulthood. *Journal of Adolescence*, 33(1), 171-181.
<https://doi.org/10.1016/j.adolescence.2009.05.004>
- Inter-American Commission on Human Rights. (2023). Freedom of expression and the regulation of digital platforms. Organization of American States.
https://www.oas.org/en/iachr/reports/pdfs/lexmedios_eng.pdf
- National Center for Biotechnology Information. (2023). Understanding the impact of social media on adolescent mental health. National Institutes of Health.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10476631/>
- National Center for Injury Prevention and Control, Centers for Disease Control and Prevention. (2013). Surveillance summaries: Youth risk behavior surveillance—United States, 2013. *Morbidity and Mortality Weekly Report*, 62(Suppl 2), 1-35.
https://www.cdc.gov/mmwr/preview/mmwrhtml/su6202a1.htm?utm_campaign=enews20220303&utm_medium=email&utm_source=govdelivery
- Sweeney, C. T. (2019). The juvenile justice system's role in children's mental health. *Buffalo Human Rights Law Review*, 25, 101-124.
<https://heinonline.org/HOL/LandingPage?handle=hein.journals/bufhr25&div=7&id=&page=>
- Tabor, N. M. (2019). The moral and legal foundation of youth's right to vote. *Buffalo Human Rights Law Review*, 25(1), 1-46.
<https://heinonline.org/HOL/LandingPage?handle=hein.journals/bufhr25&div=7&id=&page=>
- Youth.gov. (n.d.). Mental health promotion and prevention. U.S. Department of Health & Human Services.
<https://youth.gov/youth-topics/youth-mental-health/mental-health-promotion-prevention>

The Impact of Artificial Intelligence and Data Centers on the Stock Prices of Utility Companies By Lucas Camacho

Abstract

In the stock market, the growth of artificial intelligence has accelerated the growth of companies in many sectors. One sector that has been strongly connected to the rise of AI is utilities, as many of these companies are necessary for building large data centers for AI use. Dozens of utility companies were selected to have their past four quarterly reports downloaded and then data scraped for mentions of terms relating to artificial intelligence. Then, a linear regression of these mentions of artificial intelligence and the stock prices of these companies over the same time period was graphed for each individual stock. This resulted in twenty-nine linear regression models and corresponding graphs, which were found to have an average coefficient of determination value of 0.295. Additionally, 20 of the 29 graphs stocks displayed a positive correlation between artificial intelligence mentions and price. Due to the lack of clarity from the determination values, these graphs provide not enough evidence to strongly support the theory that companies that are intertwined with the use and growth of artificial intelligence, specifically ones in the utilities sector, have seen an increase in their stock prices because of that connection.

Introduction

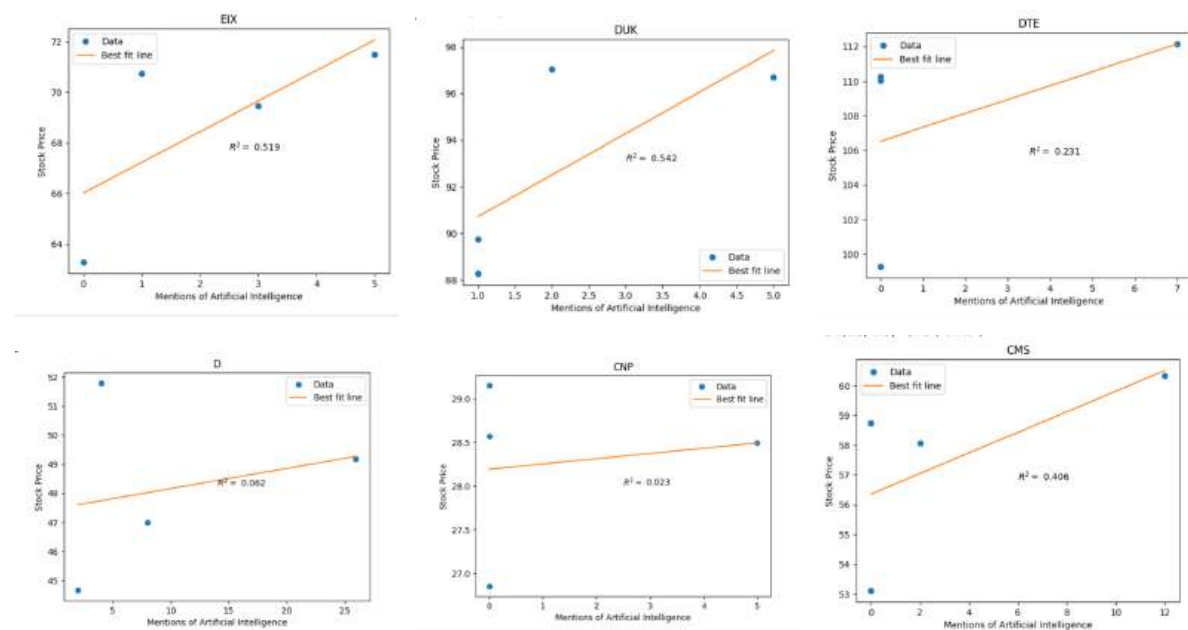
Encompassing only 2.9% of the stock market and being seen as a relatively unchanging sector, utilities are not often discussed in developments of the stock market. However, recent research is causing economists to suspect there may be a new trend in the area; utilities companies in artificial intelligence (AI). According to the Wall Street Journal, Vertiv, and Vistra, companies that supply artificial intelligence data centers with the necessary utilities, as well as others have seen considerable increases in their stock prices. This has resulted in a rush to invest in these companies, and a need to determine whether or not AI is truly the cause behind the changes in these stock prices. This experiment uses all of the quarterly reports of S&P 500 utility companies, scrapes them for AI mentions, and then compares them to their stock prices to evaluate the strength of their correlation. Using a Python data scraper and a linear regression, we seek to understand whether the mentions of AI-related terms and stock prices of utility companies in the S&P 500 are positively correlated.

Methods

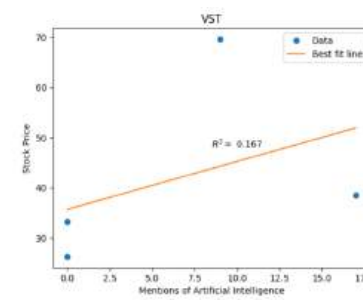
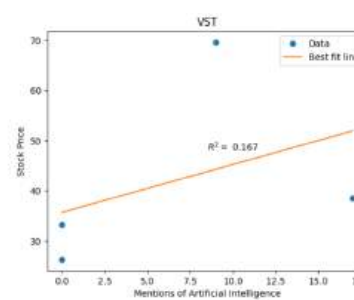
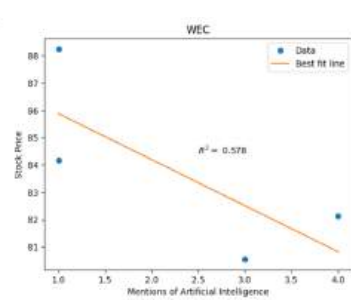
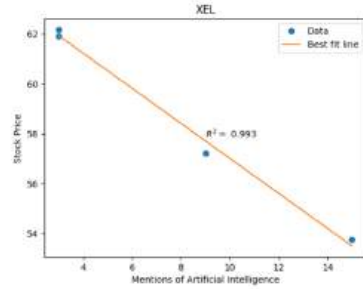
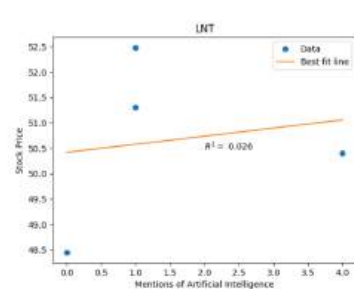
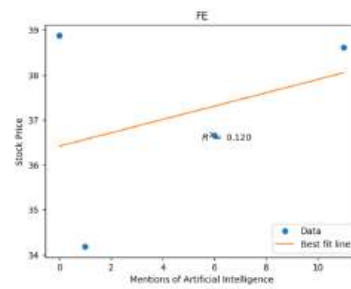
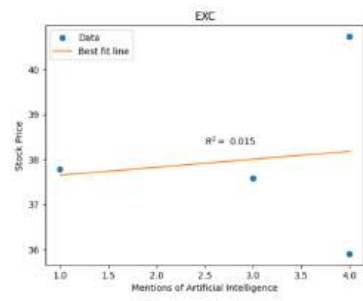
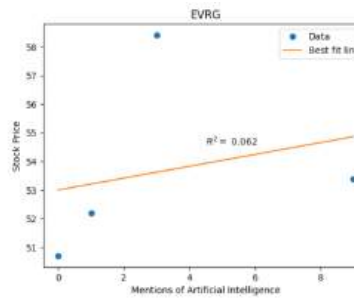
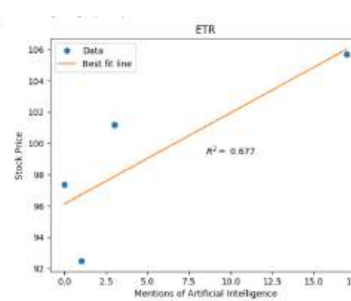
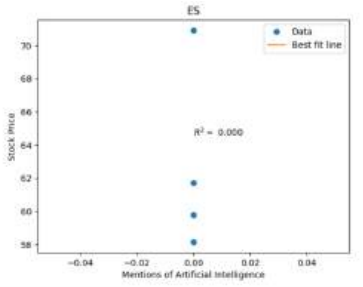
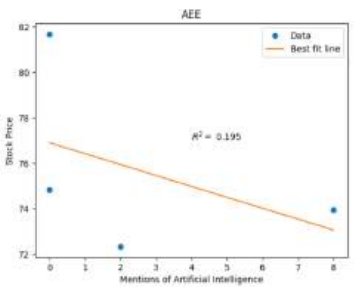
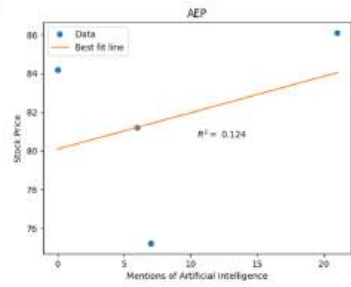
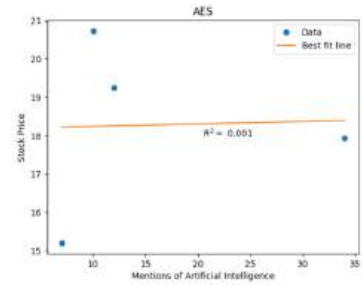
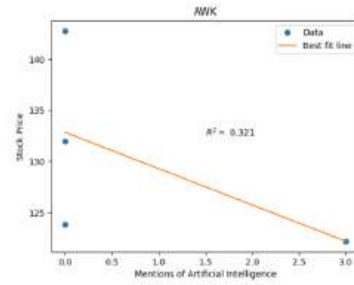
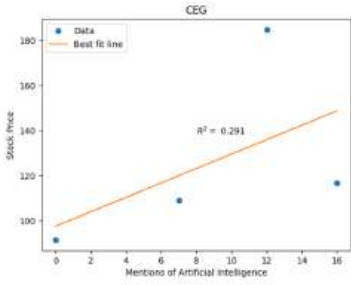
To begin, a decent number of utility companies and their quarterly reports were needed that were relatively similar in order to reduce selection bias. Twenty-nine stocks from the S&P 500 suited this need as they had easy-to-find quarterly reports and held the similarity of being all power-based utility companies that have some of the most investments across the globe, meaning that the findings of this experiment could have a larger impact. Using the website Seeking

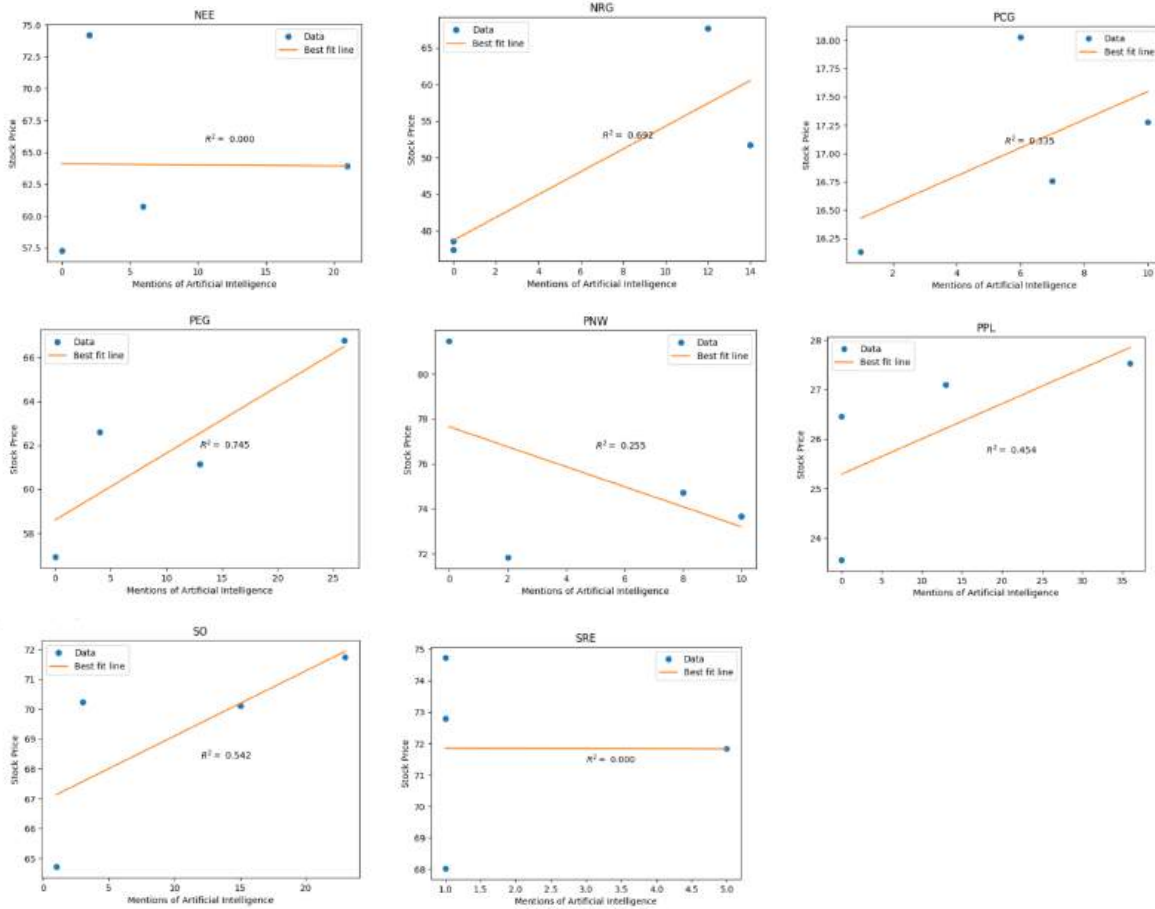
Alpha³, the four most recent quarterly reports of each of these companies (from Q2 2023 to Q1 2024) were downloaded and moved into a Google Drive folder. Last year saw a huge spike in investment in artificial intelligence and the utility sector, so this was determined to be the best time period to analyze the correlation between the two. After downloading the dataset, the data scraper was coded in Python on Google Colab using the library *re* to scan documents for artificial intelligence-related terms such as “AI”, “data-center”, “ML, and “machine learning”. Furthermore, the library *Google.colab* allowed for the Google folder holding the quarterly reports to be opened and searched for specific documents. The results of the data scraper were moved into a Google sheet to be referenced later. Once each of the reports was data scraped, the linear regressions were performed on each stock individually, as when performed on the entire dataset, there were too many outliers to determine any tangible trend and there needed to be normalization as one stock could average \$10 while another could average \$1,000. Since there may be different relationships between the AI mentions and stock price for each company, performing 29 individual linear regressions would be a better option. To perform this, the mentions of artificial intelligence and data centers were inputted as the x-coordinates, and the corresponding stock prices were inputted as the y-coordinates; a testing set was not used due to lack of data points. *Statsmodels.api* and *numpy* were used to create a line of best fit for each dataset, and then *matplotlib.pyplot* graphed the data points, and line of best fit, and displayed the coefficient of determination value.

Results



³ <https://seekingalpha.com/>





These 29 graphs display the relationship between the artificial intelligence mentions of S&P 500 utility companies, representing the x-coordinates, and the corresponding stock prices, representing the y-coordinates in USD. These four blue dots on each graph represent data points of stock prices and artificial intelligence mentions at the end of each quarter beginning with Q2 of 2023 through Q1 of 2024. The orange line displays the line of best fit that was found from performing a linear regression. The company ticker for each graph can be found at the top, and the coefficient of determination value (R^2) can be found in the center of the graph. A positive correlation between the USD value of the stock and the AI mentions would support the idea that investment in AI and data centers by utility companies leads to higher stock value. While the majority of the stocks held a positive correlation, very few had high coefficient of determination values (over 0.5 for the purposes of this experiment), meaning that the certainty on this relationship is quite low.

Discussion

Over the 1 year period, it was found that 6 of the 29 stocks held a positive correlation with R^2 values of above 0.5, while another fourteen stocks had a positive correlation with R^2 values that are less than 0.5. Three of the stocks had non-monotonic relationships with small R^2

values, and there were five stocks with negative relationships, only three of which had R^2 values of over 0.5. With few data points over the past year, many of the stocks that had low R^2 values were affected by a single outlier point. Also, there was a small decrease in the overall value of S&P 500 utilities stocks from \$344.01 to \$333.49, or roughly a 3.1% decrease. Although this is a small decrease, this may have caused some of the stock prices to trend downward despite having increasing mentions of artificial intelligence.

Conclusion

By using a Python data scraper, stock prices, and linear regression, this experiment was expected to find a strong correlation between the artificial intelligence mentions and the corresponding stock prices of utility companies. However, there is not a strong enough correlation between the two in this experiment. Perhaps future extensions of this work could use a larger time frame, more companies, or companies from other related sectors. This experiment can help show the usefulness of data scrapers in finding trends in the stock market, as well as potentially justifying whether or not utility companies should invest in artificial intelligence.

A Retrospective of SPACs

By Alexander Cao, Andrew Cun, Conrad Krol, Ryan Lung

Abstract

This paper analyzes the correlation between passive ETF investments and SPAC pricing along with the roles and impacts of each party involved within the de-SPAC process. We find that ETFs serving as passive investment vehicles are primary holders of SPACs as seen through the -15.56% median decline in SPACs following the delisting of SPAK, a leading SPAC based ETF. We further analyze the opportunities within the de-SPAC process for insiders to receive personal benefits through the means of payouts, founder shares, and more.

Keywords: SPACs, Shell Companies, ETFs, Passive Investing, De-SPACs

Consider a scenario in which one's investment has plummeted by 90% overnight. This is not a hypothetical scenario, but a recurring reality in the world of de-SPACs, where companies post-IPO face extreme volatility and dramatic selloffs. The recent increase in usage of SPACs as an alternative to traditional IPOs emphasizes the importance of understanding the risks that are weighed against the investor.

Throughout our research, we analyze a dataset containing over 550 of the most recent de-SPACs from 2016 to 2024. Our data includes the extent of the sell-off that occurred after the de-SPACs; the drawdown experienced when entering short positions; and the monetary effects of ETFs. We find that of these de-SPACs, a staggering 99% fall below their initial share price. This suggests that these stocks are overvalued, where the fall in price reflects the actual underlying value. According to NYU Law School Professor Michael Ohlrogge, "These companies were aware the valuation the SPAC was giving them was exceptionally generous. It's a no-brainer to take advantage of that." (McGinty et al.) Despite this drop in price, no SPACs dropped below \$0.01 prior to the end of the lock-up period, which is when founders are allowed to sell their shares. Since founder shares are typically bought for nominal prices, around \$0.0001 per share, a net profit is ensured for these particular shares if the de-SPAC process finalizes successfully. Individual investors are compensated if the SPAC merger doesn't go through. (Jenkinson & Sousa 3-5) This creates a possibility for a conflict of interest between the two parties involved. We also believe that passive ETFs have been a considerable source of funding for these SPACs.

We believe this paper meaningfully contributes to the realm of SPAC-based research by establishing a connection between passive ETFs and SPACs, as well as examining the incorporated risk factors within these ETFs. These risk factors were highlighted through insider involvement, the Fama French Three Factor Model, and the analysis of past SPAC to ETF price correlation data.

I. Background

Imagine waking up to find that the value of your investment has plummeted by 90% overnight. This is not a hypothetical scenario, but a recurring reality in the A SPAC is a Special Purpose Acquisition Company, a shell company created specifically for the purpose of raising capital through a public offering to acquire or merge with an existing private company. SPACs offer an alternative, faster route for a private company to go public compared to the traditional IPO (Initial Public Offering) route.

SPACs have existed for decades; however, their popularity has only recently surged in global markets. Originating in the 1980s, the concept of a SPAC was initially coined as a “blind pool”, where investors would invest in a shell company without any prior knowledge of how their money would be used. However, these companies faced immense criticism because they rarely benefited the investors, enriching the promoters who took fees at the cost of investors. The concept of SPACs began with the investment bank GKN Securities in the early 1990s, run by David Nussbaum. GKN Securities redefined the concept of blind pools, introducing several changes that would help protect the investor. GKN Securities introduced the concept of holding raised capital in interest-bearing escrow accounts. The escrow accounts ensured that capital was protected until a viable acquisition target was approved by shareholders. Crucially, a defined investment timeline was required for all SPACs, typically between 18 to 24 months. If no acquisition is made within this period, all capital is returned to the respective investors. Additionally, GKN recognized the importance of investor participation. GKN created new provisions which allow shareholders to vote on and approve any proposed acquisition. Going into the early 2000s, early pioneers such as EarlyBirdCapital, led by aforementioned Nussbaum, continued redefining the concept of a SPAC, fighting to gain legitimacy in the financial sphere. However, SPACs only began gaining significant traction in the late 2010s and early 2020s. (Shachmurove & Vulcanovic 14-23)

When the COVID-19 pandemic broke out in late 2019, markets became increasingly more volatile globally, making traditional IPOs challenging. SPACs quickly became a popular alternative to traditional IPOs, offering a more flexible and faster route for private companies to enter public markets. (Karnjanonun 1-17)

Today, a SPAC is formed by a group of sponsors or founders that contribute initial capital to initiate the SPAC, who typically receive a stake in the SPAC as compensation through founder shares. The SPAC goes public through an IPO, which includes selling units to investors. After establishing the SPAC, the management team identifies a suitable private company to merge with, after which the SPAC and the target company sign a letter of intent that includes the proposed valuation and other key deal terms. Once both parties agree on the terms, they enter a merger agreement which legally binds them to complete the transaction. The SPAC files a proxy statement with the SEC (Securities and Exchange Commission) providing a detailed report about the target company and the proposed merger. SPAC shareholders must then vote to approve the merger. If the majority votes in favor, the deal can proceed. Once shareholder approval is obtained, the SPAC and the target company close the merger. The SPAC changes its name to

reflect the target company's identity and begins trading under a new symbol. This ending step is often referred to as a "de-SPAC". Often, additional capital is raised through a Private Investment in Public Equity (PIPE) transaction which involves selling shares to institutional investors, providing extra funds for the merger or future growth. (Blankespoor et al. 6-7)

II. Analysis of Previous Papers

Heyman's paper "From Blank Check to SPAC: The Regulator's Response to the Market, and the Market's Response to the Regulation", Heyman suggests that SPACs have been used to bypass regulations passed by the "Penny Stock Reform Act", retaining many of the previous benefits of the blank check company while still retaining investor protections. He details some of the benefits of SPACs for investors, including the ability for them to gain "...exposure to private equity style deals without having millions of dollars to invest" (Heyman 548-549). Seeing as these shares are exchanged for small sums on the market, there is a possibility of significant returns for the investor. On the other hand, there are also many potential dangers. Continuing to hold a position pertains to the same risk as a long-term investment in any other small-cap growth company. Private equity investors are also incentivized, as SPACs provide a lucrative, low-risk and cheap avenue to make additional profits. This causes confusion as to why investors continue to take their chances with SPACs.

The analysis in Blankespoor et al.'s paper "A Hard Look at SPAC Projections", discovers that the projections provided by SPACs are overly optimistic in their valuation forecasts. They report: "For the sample of SPAC mergers with observable post-merger revenue, we find that only 35% of firms meet or beat their projections. This proportion declines for forecasts that are longer horizon, and non-serial SPAC sponsors miss forecasts by greater percentages." (Blankespoor et al. 1) Although it is understandable that these firms would want to remain confident in the longevity of the business, it is worrying that investors are misled by these inaccurate statistics. It suggests that instead of choosing to remain transparent and honest with the investors, these firms choose to misdirect them and provide false hope and unrealistic expectations.

According to the analysis in Spamann and Guo's paper "The SPAC Trap: How SPACs Disable Indirect Investor Protection", they believe the investor protections inherent in most public securities keep most investors safe without fully comprehending the business and the terms. However, SPACs remove this defense because of the two possible compensations offered from the SPAC share, either the redemption value or a post-de-SPAC share. While the first option is typically selected by knowledgeable investors, uninformed investors choose the second, either intentionally or automatically, by default. Seeing as the SPAC share value pre-de-SPAC is at a higher payoff, these novice investors pay more than necessary. These margins are exploited by SPAC sponsors and other IPO investors, allowing them to make profits from the SPAC even if the value collapses. We further investigate this point by examining the relationship of passive-ETFs to these SPACs. We find significant correlation and evidence suggesting that clueless investors are contributing significantly to these profit margins.

III. SPAC v. IPO Process

In the SPAC process, companies are exempt from a series of hindrances that are generally involved in IPOs. This lack of oversight consists of the lack of disclosure within investors, essentially leaving the future of the merger and company on the management team. (Anconetani et al.) This is further emphasized within the valuation of the equity being based purely off of the management team as well as the potential post-merger value which will not be subject to financial audit. (Agarwal 3) IPOs on the other hand are subject to a series of hindrances that overall slow down the public offering process in exchange for the safety of investors. First, the target company must file a detailed prospectus to ensure investors can evaluate a specific business and its prospects. Next, the IPO equity will be valued based on the current market conditions and stock price, along with a mandatory financial audit.

Additionally, there is an overall increase in mandatory financial transparency with regulators and banks. The final difference is the lockup period which all sponsors and insiders are subject to, ensuring the immediate liquidation of mass funds will not occur. These differences further emphasize the most probable current and future use of de-SPAC events to publicize an SPAC company. The most probable use overall favors the company over the individual investors. (Spamann & Guo 2-5) This potential favoring of the company can be analyzed through a series of key aspects that are solely embedded within SPACs. First, the emphasis on fast-tracking the public offering process along with the valuation completely relying on the management team's valuation. The most probable usage of these SPACs would then be for sponsors and company executives to gather funding through PIPE investors with these overinflated valuations and benefit off the consequent financial packages and founder's shares. These overinflated valuations and aspects of the de-SPAC public offering aren't seen throughout traditional IPOs, further indicating the usage of SPACs as a means to benefit insiders within the company. (Lin)

A. Drawbacks of de-SPACs

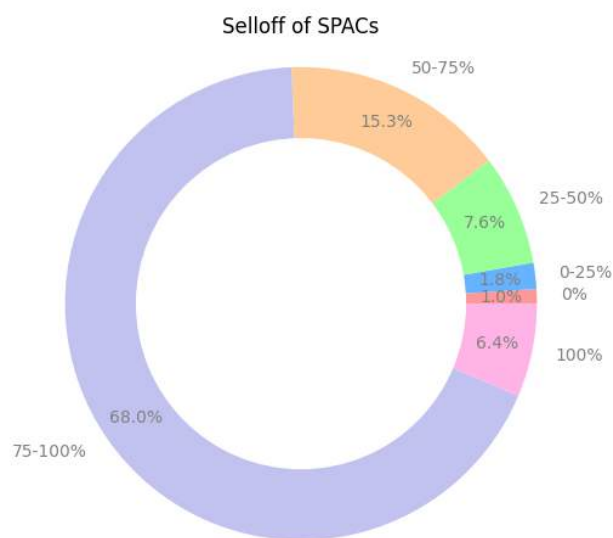
With the supposed drawbacks of de-SPACs, the continued use of de-SPACs is questionable. Since the de-SPAC process is a less regulated public transition route, there have been multiple instances of corrupt internal practices and exploitation of ignorant investors in this emerging process. (Henriksson & Hansson) For example, in the SEC v. Shvartsman case in \$DJT and \$DWAC, insiders utilized confidential information to capitalize on the merger before it was announced. Some parties involved in these fraudulent de-SPACs were repeat offenders, like Shvartsman, revealing a lack of regulation within the market. More than 50% of de-SPACs involve insider trading in 2023 according to the WSJ. (McGinty et al.) This lack of oversight heightened the risk of future instances of internal corruption. Since then, the SEC has implemented further regulation to minimize such behavior and attempt to secure SPAC's longevity as a means of public offering. The SEC has adopted new rules to protect investors in SPAC IPOs and de-SPAC transactions by increasing investor disclosure, sponsor compensation, and dilution. (SEC)

B. Corrupt Internal Practices

In recent years, a common sentiment has grown that the industry of SPACs are corrupt as a whole. This has been difficult to prove due to lack of substantive evidence. However, our findings paint industry practices as a scheme in which high-level executives attempt to take advantage of their own company for personal profits, putting the underlying stock and company in jeopardy. We arrive at this conclusion through three main suspicions. Firstly, more than 50% of de-SPAC'd companies involved insider selling, (McGinty et al.) SPAC management teams overestimate post-merger value, and operating company executives clearly understanding the consequences of de-SPAC mergers.

It is undeniable that de-SPAC'd companies may be manipulated by insider traders, often in the form of a substantial sell-off in price directly following the de-SPAC. This is confirmed through a recent research study in which “[out of] 460 companies that did SPAC deals, 232 [consisted of] insider sales based on a review of SEC filings.” (McGinty et al.) This research further solidifies the claim of internal corruption within de-SPAC mergers along with providing potential participants and specific parameters regarding percentage increase in the underlying equity that may be implemented in the form of regulation. These parameters also provide potential solutions to insider trading such as limiting insider ownership as a majority of sell-offs are incited by shareholders with more than 10% ownership. Specific internal roles are also mentioned such as corporate officers or the SPAC management team, which highlights further insight and proof of the corrupt process that occurs internally. (U.S. Securities and Exchange Commission)

Another key suspicious indicator is the fact that SPACs are known to fail with no



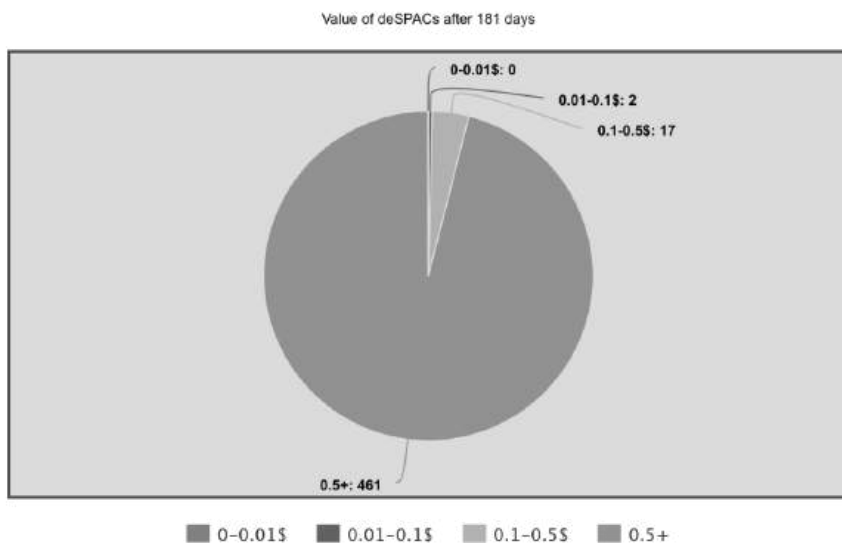
supposed benefit for the target company, yet these companies continue to participate in this IPO process. Our data shows that the de-SPAC process results in immense sell-offs in price post-de-SPAC. From our data, which we were given by SPAC Research, we find that there is a significant increase in the volatility of the underlying stock after de-SPACs events. This typically causes the equity's price to fall significantly, often to a point where recovery is near unattainable. These sell-offs occurred in over 95% of all the cases we analyzed. 89.7% of the time, the underlying stock falls over 50%. Only

occasionally, in exceedingly rare circumstances, will the price of the stock increase above the opening price, in 1% of cases. A potential rational reason would be insider benefits to specific company executives and SPAC sponsors. SPAC sponsors generally benefit due to the fact that they receive 20-25% of founders shares and associated warrants following the de-SPAC,

meanwhile target company CEOs receive lump sum payouts, stock warrants, and founder shares. (Spamann & Guo 2-5) This has previously been assumed to be further motivation for insider collaboration and trading by several research institutions, some even concluding that "executives from private companies targeted by (or themselves courting partnership with) SPACs have often extracted significant financial packages from SPAC transactions [additionally] insiders have won big despite post-combination share prices plummeting." (Mignardi) This provides clear motive and benefits to each individual insider party adding to the argument of internal corruption within the collaborating parties. Additionally, it allows the opportunity for a potential bypass of recent regulations that mandate publicizing internal disputes. This is because all those involved in the merger understand the result of the de-SPAC is the devaluation of their company in exchange for the monetary benefits to insiders. This in turn may result in no conflict in interests, allowing for all insider trading and operations to remain private throughout the process.

IV. Incentives for SPAC Executives

Due to various factors, we believe that it is largely in the interests of company executives to go public through a SPAC. Such factors include the speed of the entire process, the fact that



sponsors are compensated with these founder shares, and a greater incentive for the process to complete successfully. (Jenkinson & Sousa 3-5)

One of the key benefits of choosing SPAC when deciding to IPO is the swift registration of the company. A typical SPAC merger can take as few as three to five months to complete, while a

traditional IPO can take around nine to twelve. (Shachmurove & Vulcanovic 14-23) This discrepancy can be attributed to the shell company being publicly listed, which avoids many of the regulations and steps in a standard IPO. This shortened process allows the marketers of the SPAC to effectively capitalize on their marketing.

The ability to purchase a huge percentage of founder shares at a nominal price is largely weighted in the founder's favor. These founder shares are typically priced around \$0.0001, as is the case for \$ATP for example. According to our data, this indicates that in every case, even years after a merger or extreme outliers, the founders will always redeem profits from these shares. (Fortney 6) The only possibility for these founders to not turn a profit through these shares is if the deal fails. Regardless, in a failing de-SPAC the founders don't lose their shares,

the shares remain illiquid until a successful merger. It cannot be understated how advantageous this deal is for the founders, largely at the expense of investors. Our data shows that 0 out of the 480 post-de-SPAC publicly listed companies fall below \$0.01 after 181 days after their closing date. Since the typical lockup period for founder shares is between 180-365 days (Moon 7-8), this data shows how lucrative de-SPACs are for the founders, even if the companies sells off after going public. (Anconetani et al.)

V. Data Collection Processes

A. SPAC Data

To begin, we utilize data given to us from spacresearch.com, containing 583 tickers of de-SPAC events, spanning from July of 2016 to July of 2024. The dataset encompasses the entirety of the recent rise and fall of SPAC deals, providing us a comprehensive representation of de-SPAC events. In addition, the dataset notes the date of close for every SPAC deal, which will prove vital for later data analysis. Upon acquiring the dataset, we gather additional data from each SPAC ticker enabling us to perform further analysis. With the use of Python, Yahoo Finance, and other relevant libraries, we are then able to measure prices of the market open after the date of close, as standard investors are unable to purchase shares during aftermarket hours. We then calculate the maximum amount of drawdown an investor would have with a short position, along with the profit gained on the short position. We also record the number of trading days that pass from the time of going public until the minimum value is reached.

Consider the following entry from the dataset:

Ticker	Percent Min	Percent Max	Date Open	Date Min	Date Max	Missing Vals
ADGM	-50.0	16.75	2024-08-01	6	1	Y

B. SPAK ETF Data

Alongside data for de-SPAC events, we also must gather data on Defiance Next Gen SPAC Derived ETF (SPAK), the ETF used in this case study. This process is relatively straightforward, with all mutual funds being required to publicly disclose their portfolios on a quarterly basis (ETF.com). Thus, we are able to access the latest publicly available filing on the SEC's EDGAR database for SPAK, before its delisting (ETF Series Solutions). Finally, after cleaning the text file obtained from the SEC, we are left with a list of every stock in SPAK's holdings prior to delisting, as well as the amount of capital invested in each holding.

VI. Passive Investing, ETFs, and Their Effects on SPACs

Given the significant and well-documented decline in SPAC activity over the past few years—illustrated by the stark drop in capital raised, from \$160.75 billion in 2021 to only \$13.42 billion in 2022 (O'Connor & Nazir)—the question arises: why do SPACs still exist in today's market? Our paper has posited that due to high levels of information asymmetry and conflicting interests among key stakeholders during de-SPAC transactions, SPACs have evolved into a corrupt vehicle that disproportionately benefits insiders at the expense of retail investors. However, all this hinges on a certain level of external investment. With SPACs becoming a relatively niche form of investment, and with the continual failure of SPACs in recent years, it is puzzling why anyone would actively choose to invest in them today.

Thus, this section will propose a theory on the origins of external investment in SPACs.

A. Active v. Passive Investing

Traditionally, the financial world has been dominated by active investing. Even today, when one thinks of investments, they would imagine the busy scenes of wall street or an active day trader. Active investors engage in the proactive purchasing and selling of assets, aiming to outperform the market with their own information and beliefs. Passive investing on the other hand is defined as “an investing strategy that aims to maximize returns over the long run by keeping the amount of buying and selling to a minimum. It involves building a portfolio to mirror a market index and then holding onto those investments without reacting to market fluctuations.” (Inspired Economist)

The influence of passive investing has grown significantly over the past few decades. In 1993, passive funds invested in U.S. stocks managed \$23 billion of assets, representing only 3.7% of the combined assets of active and passive funds invested in U.S. stocks and just 0.44% of the U.S. stock market. By 2021, however, passive assets had surged to \$8.4 trillion, overtaking active funds to account for 53% of the combined active and passive fund market and 16% of the entire U.S. stock market. (Jiang et al. 3)

In Gârleanu and Pedersen's paper “Active and Passive Investing: Understanding Samuelson's Dictum”, they postulate a significant link between ‘uninformed’ investors and passive investors given that they both seek to maximize their performance subject to a minimal cost. Unlike active investors, who dedicate resources to research and analysis in an attempt to outperform the market, uninformed and passive investors accept market returns with minimal intervention.

Their findings reveal a critical insight regarding the behavior of passive traders. They argue that passive investors tend to be less informed compared to their active counterparts, often operating with limited knowledge about the specific assets they are purchasing, or in other words, this places them on the unfavorable side of the information asymmetry. This observation becomes particularly relevant in the context of SPACs, who are characterized by their high degree of information asymmetry. As revealed earlier in this paper, insiders and sponsors possess

far greater amounts of knowledge and insights about the business compared to the average investor. Thus, external investors for SPACs could potentially originate from passive sources.

B. Impact on SPACs

SPACs have attracted significant investment from passive funds in the past. Popular ETFs such as SPAK and SPCX were created solely to capitalize on the upcoming trend of SPACs during the late 2010s and early 2020s. (Dierking) (Likos) Many broader-based ETFs automatically include SPACs as part of their diversified holdings, usually in sectors relating to SPACs. (Chamberlain)

Our findings suggest that passive ETFs may play a significant role in artificially inflating SPAC prices, contributing to market inefficiencies that can be exploited by insiders.

Passive acquisition occurs automatically as part of the ETF's rebalancing or investment strategy, regardless of the actual value or performance potential of the SPAC. Consequently, increased demand generated by these ETFs can inflate SPAC prices, whose value, therefore, does not align with the underlying fundamentals of the business. This creates an environment where insiders, who are more informed about the true value of the SPAC, can capitalize on inflated prices by exiting their positions, often to the detriment of less-informed retail investors.

In other words, ETFs act as liquidity providers in markets that are not otherwise highly liquid.

C. Case Study: Defiance Next Gen SPAC Derived ETF

Defiance Next Gen SPAC Derived ETF (SPAK) was launched in September 2020, during the SPAC boom. It was an exchange-traded fund that focused primarily on SPACs and companies derived from de-SPAC events. SPAK aimed to track the performance of the Indxx SPAC & NextGen IPO index. The Indxx SPAC & NextGenIPO index, SPAK's underlying index, included both companies before and after de-SPAC merger events, reflecting a broad representation of the SPAC environment. As the first ETF specifically created for SPACs and SPAC-derived companies, SPAK sought to capitalize on the recent surge in interest in these blank-check companies.

Having launched during a time that coincided with the SPAC boom of the early 2020's, SPAK attracted significant attention from investors. The ETF experienced early success in both performance and capital inflows (Yahoo Finance). However, due to problems within the underlying fundamentals regarding countless SPAC companies, as outlined previously in this paper, several high-profile SPAC deals underperformed after their de-SPAC event.

DraftKing's all-time high stock price of \$71.72 in March '21 has fallen ~84% to \$11.67 per share as of June 2022, as the company continues to post operating expenses well above revenues, resulting in materially lower EBITDA of -\$1.562bn. (Yahoo Finance)

Virgin Galactic's ATH of \$59.41 per share has plummeted to \$6.02 as the company expected FYE '21 revenues of \$210mm and EBITDA of \$12mm, whereas their actual FYE '21

results showed a much different story of \$3mm in revenue and -\$310mm in EBITDA. (Yahoo Finance)

Similarly, Nikola's ATH stock price of \$79.73 has decreased 91% amidst fraud accusations, as well as material misses in performance expectations. (Yahoo Finance)

Nikola projected an FYE '21 EBITDA of -\$211mm, but ended up over 3x lower at -\$685mm." (Paraboschi & Hill, 2022)

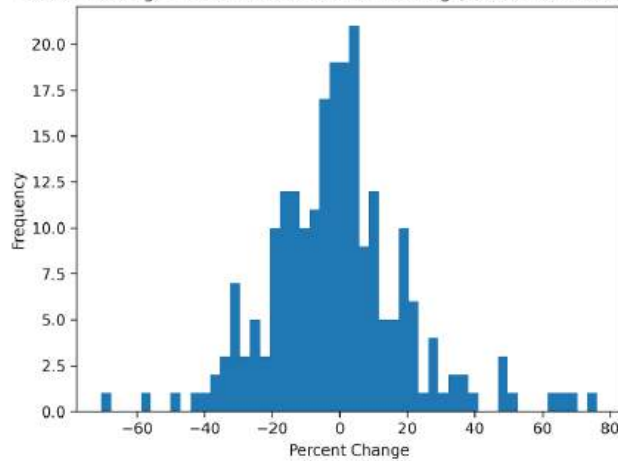
By mid-2021, SPAK's performance had declined sharply, and the ETF saw a drop in assets under management" (AUM). (ETFDB)



C.1. Statistics

The statistics surrounding the price changes of SPAK holdings before and after its delisting offer valuable insights into the broader impact of the ETF's collapse on the SPAC market. These figures, derived using Python to analyze SPAK holdings and SPAC performance data, highlight the downturn in value, both for SPAK-associated holdings and SPACs in general.

Percent Change 1 Month Before SPAK Delisting (7/30/2022 - 8/30/2022)

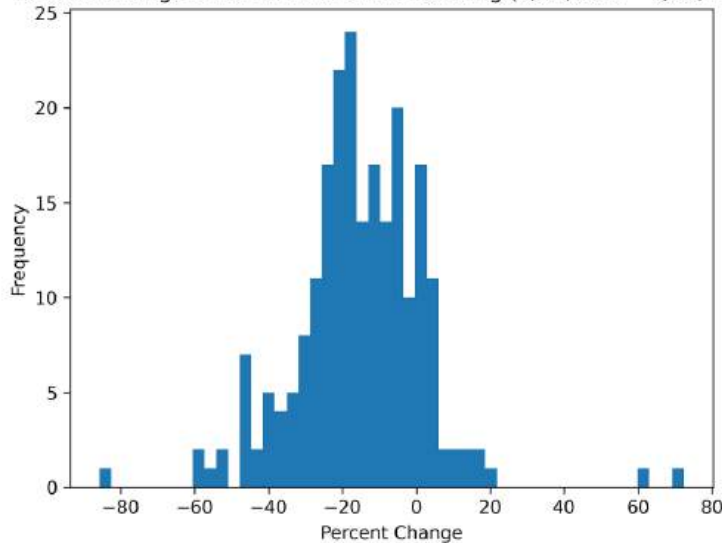


Average percent change: -0.91%

Median percent change: -1.32%

Before delisting, SPAK's holdings were already experiencing negative price performance, although the decline was relatively modest. An average decrease of just -0.91% suggests that investor sentiment toward SPACs was becoming increasingly negative, but this hadn't yet turned into a full-blown sell-off. The median price change of -1.32% reflects similar trends, indicating that the majority of holdings were seeing small, yet notable, declines. This modest drop could be attributed to market uncertainty about the future of SPACs, which were starting to face increasing scrutiny and regulatory pressure at the time.

Percent Change 1 Month After SPAK Delisting (8/30/2022 - 9/30/2022)

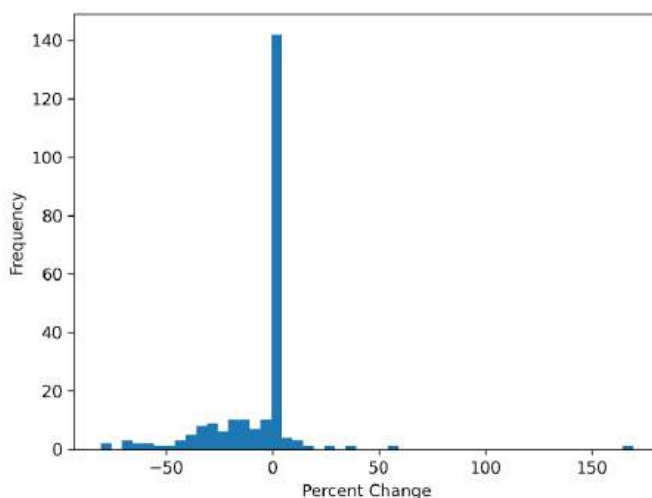


Average percent change: -15.49%

Median percent change: -15.56%

The data paints a much bleaker picture following SPAK's delisting. The average percent change plummeted to -15.49%, with the median at -15.56%. This significant drop illustrates how closely tied the fortunes of these SPACs are to SPAK's existence. With the ETF no longer offering support and liquidity, many SPAC holdings experienced sharp declines, reflecting the growing lack of confidence in SPACE-related investments during this period. Price changes of all SPACs excluding SPAK holdings 1 month after SPAK's delisting 8/29/2022:

Price Changes of all SPACs Excluding SPAK Holdings 1 Month After SPAK's Delisting on 8/29/2022:



Average percent change: -7.172030278080107

Median percent change: 0.09920862061037806

When analyzing a broader perspective including SPACs, not just SPAK's specific holdings, a similar downturn is evident, although the decline isn't as drastic. The average percent change of -7.17% indicates a broader market trend of declining SPAC valuations, but the median change of 0.09% suggests that not all SPACs affected as heavily as SPAK's specific holdings. The wider spread between the average and median points to significant underperformance among a few key SPACs that dragged down the overall average.

C.2. Fama-French Three-Factor Model

In order to gain an in-depth understanding on the significance of SPAK, SPAK's performance using the Fama-French Three-Factor Model must be analyzed.

The Fama-French Three-Factor Model functions to explain the returns of an equity by incorporating additional risk factors. The model provides the option to break down SPAK's returns into three key components:

- **Market Risk Premium (Mkt-RF):** This represents SPAK's sensitivity to overall market risk.
- **Size Factor (SMB - Small Minus Big):** This measures SPAK's exposure to small-cap stocks, which tend to have an increased volatility.
- **Value Factor (HML - High Minus Low):** This captures the favor toward value stocks (high book-to-market ratio) or growth stocks (low book-to-market ratio), helping us understand whether SPAK's returns are due to speculative, growth-oriented investments often seen through SPACs.

Given these variables, the Fama-French Three-Factor Model can be represented by the following equation.

$$R_{SPAK} - RF = \alpha + \beta_1(Mkt - RF) + \beta_2SMB + \beta_3HML + \epsilon$$

Where:

- R_{SPAK} is the return of SPAK.
- RF the risk-free rate.
- α is the model's intercept, or the portion of SPAK's returns that is not explained by the three risk factors.
- $\beta_1, \beta_2, \beta_3$ are the sensitivities (factor loadings) to the market risk premium ($Mkt - RF$), size factor (SMB), and value factor (HML), respectively.
- ϵ is the error term representing any unexplained return.

This model helps quantify the relationship between traditional market risks and SPAK's returns and whether its exposure to generally abstract small-cap and growth stocks contributes to the inflated prices seen in the SPAC market.

C.2.1. Key Findings

Using the Fama-French model, the following understandings can be observed:

Market Risk Premium ($Mkt - RF$ coefficient = 0.0108):

SPAK's low market beta indicates that there is a relatively positive correlation to the market, yet an overall lack of sensitivity to broader market movements. This suggests that SPAK's returns aren't reliant on overall market trends and returns. Instead, other factors, such as further speculative small-cap companies and the SPAC market, likely play a more significant role. This lack of correlation to the broader market supports the claim that SPAK's returns are primarily reflective of the nature of SPAC investments, rather than being affected or driven by overall market fundamentals.

Size Factor (SMB coefficient = 0.0112):

The positive and statistically significant size factor indicates that SPAK is more involved with small-cap stocks, which generally indicates higher risk and volatility. Because SPACs are generally small-cap or micro-cap companies in the early stages of their growth

cycles that are valued based on post-de-SPAC potential, they rely heavily on investor optimism and future potential rather than current performance. This exposure to these small-cap stocks further supports the argument that SPAK's holdings contain inherent risk. This heavy implementation of small-caps can also explain why SPAK was a primary reason for inflated SPAC prices, as passive ETFs tend to buy these SPACs regardless of their performance and fundamentals..

Value Factor (*HML* coefficient = -0.0047):

The value factor involving a negative coefficient indicates that SPAK favors growth stocks rather than traditional value stocks. Growth stocks generally consist of higher valuations based on future earnings potential, which is a direct correlation to the nature of SPACs—many of which are micro-cap stocks with overestimated growth expectations based on the SPAC sponsors' valuation. This inherent favor toward growth stocks solidifies the argument that SPAK is a lead contributor to the inflation of SPAC prices, as passive investment ETFs such as SPAK concentrate on high-growth, speculative companies while disregarding past financials and performance.

C.3. Analysis and Implications

These statistics emphasize the differences between the performance of SPACs prior to and following the delisting of SPAK. Prior represented data reveals a correlation between the delisting of a SPAC-focused ETF and the decline in SPAC pricing. These drastic drops in price indicate that SPAK, as one of the largest and first ETFs focused on SPACs, acted as a pillar of support for the performance of SPACs by funneling passive investment into each liquidity pool. Following delisting, SPACs faced steeper sell-offs, further intensifying the already weakening market.

This sell-off could also be attributed to the drain of liquidity from SPACs as a whole. The delisting of ETFs such as SPAK results in SPAC shares demand falling, leading to a rapid devaluation process. Additionally, the high price volatility among SPACs, along with the increase in regulation from the U.S. Securities and Exchange Commission (SEC), likely aggravated these sell-offs and lack of investor confidence in the aftermath of SPAK's delisting.

The results of the Fama-French model further strengthen the argument that SPAK functioning as a passive investment vehicle plays a significant role in the inflation of SPAC prices. The significant involvement within small-cap and growth stocks, combined with the lack of correlation to broad market trends, indicates that SPAK's returns are driven by these speculative small-cap equities including SPACs.

This exposure to speculative small-cap stocks and SPACs can result in a feedback loop in which passive inflows leading into SPAK further inflate the pricing of SPACs, unrepresentative of their true value. As further shares of SPAK are purchased, the ETF is forced to purchase more of the SPACs, overall driving SPAC prices higher. This relationship between SPAK and SPACs can explain why many SPACs underwent significant price run-ups in 2020 and 2021, followed by significant periods of downturn as investor sentiment shifted.

As previously mentioned, the positive size factor (*SMB*) and negative value factor (*HML*) indicate that SPAK's holdings are heavily invested within small-cap and shell companies. These companies come with an inherent risk and generally tend to be more speculative, which supports the claim that SPAK contributed to the unrepresentative price runs in the SPAC market.

The automated allocation of capital to speculative SPACs causes passive ETFs like SPAK to inflate SPAC prices without considering the success or fundamentals of the underlying company. The results of the Fama-French model show that SPAK's performance is driven more by its involvement within the speculative small-cap, growth stocks and SPACs than by broad market performance, further solidifying the argument that passive investment inflows play a key role in the pricing increase of SPACs.

The moderate R-squared value (0.512) shows that broad market activity, size, and value explain part of SPAK's returns, yet the primary portion of returns is unaccounted for, likely being driven by SPAC-specific risks. This leaves investors at risk to generally randomized volatility and probable losses when SPAC valuations inevitably correct. The lack of significant alpha in the model implies that SPAK doesn't provide investors with excess returns, meaning there aren't compensations for the additional risk they are taking by investing in these speculative companies.

The Fama-French model indicates that SPAK's performance is strongly influenced by its exposure to SPACs and micro-cap, growth stocks. By gradually allocating capital to these riskier investments, SPAK contributes to the artificial price inflation in the SPAC market, ultimately adding to the several opportunities for insiders to profit while leaving individual investors exposed to significant risk. These findings support the broader argument that passive investment in SPACs through ETFs like SPAK further aggravates market inefficiencies and inflates prices beyond fair value levels.

In sum, the statistical data emphasizes the overall fragility of the SPAC market and the potential for volatile and unpredictable price swings in the absence of institutional support from passive investment vehicles such as SPAK. This supports the broader statement that passive ETFs play a key role in providing liquidity and supporting prices in generally volatile sectors of the market, such as SPACs.

VII. Live Portfolio Application

Although purchasing SPACs as long-term investment vehicles has been proven ineffective, the de-SPAC process still may present a notable opportunity to engage in the price sell-off that frequently follows these events. Previous analysis within this paper and historical data from 2016 to 2024 shows that 99% of de-SPAC scenarios result in price valuing under immediate post-de-SPAC value, demonstrating a consistent phenomenon that is likely to continue. Despite the high probability of these sell-offs, implementing a successful investment strategy comes with several challenges.

A significant challenge that may be faced is managing drawdown-related risk when attempting to short to fair value. After a de-SPAC, prices can exhibit unpredictable price

fluctuations before the significant sell-off, increasing the chances for substantial drawdown. The issue arises from the difficulty in accurately timing the sell-off and minimizing drawdown prior to the sell-off. Another issue faced by investors is managing risk in profitable scenarios. The sell-off is often rapid although it may take over 100 trading days to reach its bottom. This variation in timing makes it challenging to predict when the underlying stock will bottom out causing complications in position sizing decisions. This unpredictability complicates risk management for investors looking to maximize gains from profitable positions.

To effectively capitalize on the de-SPAC sell-off opportunity, a well-defined strategy that addresses key issues such as risk management and portfolio allocation is essential. In order to create an effective risk management strategy, we will use a method proven through backtesting historical data to determine the optimal risk-to-reward ratios and their corresponding win rates. By examining past de-SPAC events, investors can identify signatures in price and set strategic limits to balance potential losses and gains. For example, implementing a 13% drawdown limit while aiming for 71.16% profit provides a 5.47 risk-to-reward ratio. This strategy requires an approximate 15.47%-win rate to break-even but has a win rate of 29.79% in 2024 as of 7/28/2024. Year to date as of 7/28/2024, this risk management strategy has a win rate of 35.05% and in 2023 it had a win rate of over 27%. These risk management values simply function as an example that has been tested and can be adjusted based on historical analysis, but they should be refined continually with forward testing on new de-SPAC events to stay responsive to changing market conditions. To confront the issue of portfolio allocations, we have found that a 2% portfolio allocation per position will be most effective as a high-risk option, as many de-SPAC events can go for as long as 200 trading days before reaching 71.16% take profit.

$$RRR = \lambda\phi \div \zeta$$

Where:

- RRR is the risk-to-reward ratio.
- $\lambda\phi$ is the take profit percentage.
- ζ is the drawdown limit percentage.

$$BEW = \left(\frac{1}{1+RRR} \right) \times 100$$

Where:

- BEW is the break-even win rate.
- RRR is the risk-to-reward ratio.

Although this strategy historically is back-tested to work, real-life application is difficult due to liquidity issues as well as the fact that brokerages will not allow individual investors to short the company at its post-de-SPAC opening price right after its closing date. Individuals will only be allowed to short on the third day after opening day, complicating the backtesting process.

VIII. Conclusion

The volatility and danger of SPACs raise skepticism about why investors remain vested in the space. Instead of choosing to invest in a safe equity with consistent profits, SPACs have a notorious reputation for insider trading and dodgy deals, and one that is well-earned. We seek to identify the group of uninformed investors that continue to fuel the success of these SPACs.

Analyzing passive investment ETFs and their role in the SPAC market highlights the noteworthy correlation between ETFs such as SPAK and the inflation of SPAC prices. By passively acquiring SPAC stock shares, ETFs inadvertently contribute inefficiencies in market pricing, driving up the prices of SPACs which are disregarding underlying business fundamentals. This creates opportunities for informed insiders to profit from inflated valuations while leaving passive and retail investors exposed to the potential for significant losses when the price reevaluates.

Works Cited

- McGinty, Tom, et al. "Company Insiders Made Billions Before SPAC Bust." *The Wall Street Journal*, 30 May 2023, <https://www.wsj.com/articles/company-insiders-made-billions-before-spac-bust-4607a869>.
- Jenkinson, Tim, and Miguel Sousa. "Why SPAC Investors Should Listen to the Market." *Journal of Applied Finance*, vol. 21, 2015, pp. 3–5, <https://doi.org/10.2139/ssrn.1341771>.
- Shachmurove, Yochanan, and Milos Vulcanovic. "SPAC IPOs." *Oxford Handbook of IPOs*, 2017, pp. 14–23.
- Karnjanonun, Kanin. *SPAC Sponsors and Their Effects on SPAC Prices*. Chula Digital Collections, 2021, <https://doi.org/10.58837/chula.is.2021.68>.
- Blankespoor, Elizabeth, et al. "A Hard Look at SPAC Projections." *Management Science*, vol. 68, no. 6, 2022, pp. 6–7, <https://doi.org/10.1287/mnsc.2022.4385>.
- U.S. Securities and Exchange Commission. *Special Purpose Acquisition Companies, Shell Companies, and Projections* (Release Nos. 33-11265; 34-99418; IC-35096; File No. S7-13-22). U.S. SEC, 2022, <https://www.sec.gov/rules/final/2022/33-11265.pdf>.
- Mignardi, Lilliana W. "SEC Enforcement Continues SPAC Crackdown as Founder Trading Profits Generate Scrutiny." *Holland & Knight Insights*, 15 June 2023, <https://www.hklaw.com/en/insights/publications/2023/06/sec-enforcement-continues-spac-crackdown-as-founder-trading>.
- Bazerman, Max, and Punit Patel. "SPACs: What You Need to Know." *Harvard Business School Faculty & Research*, July 2021, <https://www.hbs.edu/faculty/Pages/item.aspx?num=60545>.
- Anconetani, Riccardo, et al. "Are SPACs a Good Investment Deal for Investors? A Performance Comparison Between SPACs vs IPOs." *Journal of Small Business and Enterprise Development*, 2024, <https://doi.org/10.1108/jsbed-11-2023-0555>.
- Agarwal, Rohan. "An Insight into SPACs and Their Valuation Conundrum." *SSRN Electronic Journal*, 2021, <https://doi.org/10.2139/ssrn.3882261>.
- Spamann, Holger, and Hao Guo. "The SPAC Trap: How SPACs Disable Indirect Investor Protection." *SSRN Electronic Journal*, 2022, pp. 2-5, <https://doi.org/10.2139/ssrn.4135558>.
- Lin, Zhong Hao. *Abnormal Returns from Insider Trades on Special Purpose Acquisition Company's (SPAC's) Warrants*. Saint Mary's University, 2022.
- Henriksson, Magnus, and Tobias Hansson. *How Do IPOs and SPACs Differ in Terms of Post-Listing Performance, and How Do Insider and Institutional Ownership Affect These Differences?* University of Gothenburg School of Business, Economics and Law, 2023.
- Fortney, Hannah. "SPAC Attack: An Examination of SPAC Director Compensation and Its Legal Implications." *SSRN Electronic Journal*, 2021, pp. 6, <https://doi.org/10.2139/ssrn.3911337>.

Moon, Lawrence. "An Overview of SPACs and the 2020-2021 SPAC Surge." SSRN Electronic Journal, 2021, pp. 7–8, <https://doi.org/10.2139/ssrn.3979359>.

O'Connor, Michael D. N. "SPAC IPOs, Deals Fell in 2022." S&P Global Market Intelligence, 30 Jan. 2023, <https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/spac-ipos-deals-fell-in-2022-73994241>.

"Inspired Economist." "Passive Investing: An In-Depth Explanation of Low Cost, Long Term Investing." Inspired Economist, 23 Oct. 2023, <https://inspiredeconomist.com/articles/passive-investing/>.

Jiang, Hao, et al. "Passive Investing and the Rise of Mega-Firms." SSRN Electronic Journal, 2020, pp. 3, <https://doi.org/10.3386/w28253>.

Gârleanu, Nicolae, and Lasse Heje Pedersen. "Active and Passive Investing: Understanding Samuelson's Dictum." *The Review of Asset Pricing Studies*, vol. 12, no. 2, 2021, pp. 1–58, <https://doi.org/10.1093/rapstu/raab020>.

Dierking, David. "Top 4 SPAC ETFs for 2022." *TheStreet*, 29 Nov. 2021, <https://www.thestreet.com/etffocus/trade-ideas/top-4-spac-etfs-for-2022>.

Likos, Peter. "Do You Want to Invest in SPACs? Here Are Three New SPAC ETFs on the Market." *U.S. News & World Report*, 23 Apr. 2021, <https://money.usnews.com/investing/funds/articles/spac-etfs-to-consider>.

Chamberlain, Linsley. "Dispelling the Hype: An Examination of SPAC Common Equity Performance." SSRN Electronic Journal, 2021.

Zong, Yichen. "Why Passive ETFs Are Really Active, and What That Means for Investors." Knowledge at Wharton, July 2022, <https://knowledge.wharton.upenn.edu/article/why-passive-etfs-are-really-active-and-what-that-means-for-investors/>.

Paraboschi, Charles, and Max Hill. "SPAC Market Update: Who Turned on the Lights?" Valuation Research Corporation, 14 Sept. 2022, www.valuationresearch.com/insights/spac-market-update-who-turned-on-the-lights/.

Hamid, Zubair, et al. "Fama and French Three Factor Model: Empirical Evidence from Financial Market of Pakistan." *African Journal of Business Management*, vol. 6, no. 8, 2012, pp. 1–6, <https://doi.org/10.5897/ajbm11.1765>.

Heyman, David K. "From Blank Check to SPAC: The Regulator's Response to the Market, and the Market's Response to the Regulation." *Entrepreneurial Business Law Journal*, vol. 2, no. 1, 2008.

"ETF Education: How Transparent Are ETFs?" *ETF.com*, Jan. 2018, <https://www.etf.com/sections/news/etf-education-how-transparent-are-etfs>.

ETF Series Solutions. SEC Filing: NPORT-P. U.S. SEC, 2022, <https://www.sec.gov/Archives/edgar/data/1540305/000114554922052011/0001145549-22-052011.txt>.

Growth Dynamics and Doubling Time of *Mycobacterium Smegmatis* Under Controlled Laboratory Conditions By Sophia Qin

Abstract

This study works towards examining the growth dynamics and doubling time of *Mycobacterium smegmatis* in a controlled environment. The experiment monitored the optical density (OD) readings at regular intervals to analyze the bacterial growth phases and calculate the doubling time. My results demonstrate the adaptability of *M. smegmatis* across different concentrations and have implications for optimizing industrial processes involving similar bacterial cultures.

Introduction

Bacteria make up a major domain of prokaryotic microorganisms, which are the most varied and adaptable organisms. These unicellular organisms live in various environments, from the human gastrointestinal tract to terrestrial and aquatic habitats. Bacteria are metabolically versatile, having adaptive mechanisms built into them that allow them to survive in different conditions.

The bacterial genome evolves through rapid mutative processes, causing changes in metabolic pathways that help these organisms use other sources of energy or become antibiotic-resistant to adapt to new environments or habitat harsh environmental conditions. With time, beneficial mutations are naturally selected, helping bacterial populations evolve and keep survival rates high over many territories. This ability ensures that bacteria will never be lost from any niche on earth.

The growth of bacteria in a controlled liquid medium follows a characteristic pattern comprising four distinct phases: the lag phase, the log (or exponential) phase, the stationary phase, and the death phase (Figure 1). The initial lag phase is a period of metabolic adjustment where bacteria synthesize new enzymes and adapt to the new environment. During this phase, growth is slow as the bacteria prepare for active replication. The lag phase is succeeded by the log phase, where bacteria proliferate at their maximum rate, doubling in number during each replication cycle.

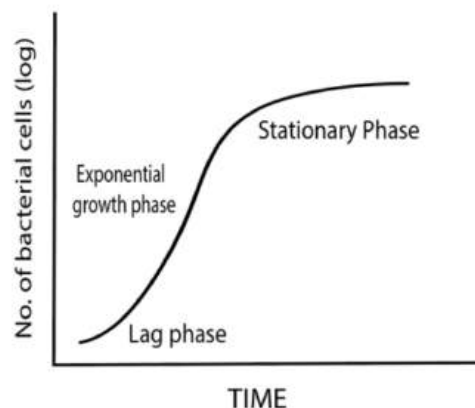


Fig 1. Generalized Growth Curve of Bacteria Demonstrating the Lag Phase, Exponential Growth, and Stationary Phase

As the bacterial population increases, resources required by microorganisms such as food and space become limiting factors for their continued multiplication. This leads to a point where the number of dividing cells stays constant over time and equals the number of dying cells; hence a steady state arises known as the stationary phase. Eventually, the culture enters the death phase, characterized by a decline in the bacterial population due to the depletion of nutrients and accumulation of waste products. The rate of cell death surpasses the rate of replication, leading to a reduction in the viable bacterial population.

The growth of bacterial populations during the log phase can be mathematically modeled using the exponential growth equation: $N(t) = N(0) \times 2^{(t/td)}$ where $N(t)$ is the population size at time t , $N(0)$ is the initial population size, t is the time, and td is the generation time or doubling time.

While the exponential growth model assumes constant growth under optimal conditions, real-world scenarios often involve environmental constraints. To account for these limitations, the logistic growth model is employed: $N(t) = \frac{A}{1 + (\frac{A-N(0)}{N(0)})e^{-rt}}$

where A represents the carrying capacity or maximum population size, and r is the intrinsic growth rate, defined as the difference between the birth and death rates per generation. Other mathematical models, such as the Gompertz, Monod, Baranyi-Roberts, Richards, and Linear Growth models, provide alternative approaches to describe bacterial growth under varying conditions (Table 1). Each model offers specific advantages and limitations, and the choice of model depends on the experimental context and the particular growth characteristics of the bacterial species under study.

Table 1. Comparison of Bacterial Growth Models Under Varying Conditions

Growth Model	Key Characteristics	Common Applications	Advantages	Limitations
Exponential	Constant growth rate during exponential phase; assumes unlimited resources.	General bacterial growth studies, where resources are not a limiting factor.	Simple, widely used, easy to apply	Only valid during exponential phase
Logistic	Accounts for carrying capacity; growth rate slows as resources become limited.	Ecological studies and situations where environmental limits impact growth.	Accounts for environmental carrying capacity	Assumes constant environment and carrying capacity

Gompertz	Handles complex growth patterns; useful under stress or varying environmental conditions.	Food microbiology and other fields where bacteria face varying conditions.	Good fit for complex growth curves, particularly in food microbiology	Can be complex to fit, parameters are not always intuitive
Monod	Growth rate limited by substrate concentration; used when specific nutrients limit growth.	Environments where nutrient concentration is a critical factor in growth.	Describes relationship between growth rate and substrate concentration	Assumes only substrate limitation impacts growth rate
Baranyi-Roberts	Includes lag phase and deceleration; widely used in predictive microbiology.	Predictive microbiology for assessing food safety and spoilage.	Detailed model with representation of all growth phases	Requires more parameters and detailed initial conditions
Richards	Flexible model that can describe various growth shapes; accommodates asymmetry in growth patterns.	Describing growth under variable conditions; often used in ecology.	Flexible, can model various shapes of growth curves	Complex with many parameters, fitting may require detailed data
Linear	Assumes constant growth rate; simple and straightforward, often used for short growth periods.	Basic growth modeling, particularly in controlled lab environments.	Simplest model, easy to understand and apply	Doesn't represent lag or stationary phases

Doubling Time

Doubling time, the time it takes for a bacterial population to double in size, is a critical determinant of its growth and reproductive success. It is central to understanding bacterial growth dynamics. This parameter is influenced by both intrinsic factors, such as genetic and physiological traits, and extrinsic factors, including pH, temperature, nutrient availability, oxygen levels, and the presence of toxins. Accurate determination of bacterial doubling time is essential for optimizing laboratory experiments and industrial processes. Doubling time varies widely among different species of bacteria. For instance, *Escherichia coli* can double its

population in as little as 20 minutes under optimal conditions while *Mycobacterium tuberculosis* requires approximately 10 hours to do so.

Doubling time can be calculated using various mathematical models, each tailored to specific growth phases and environmental conditions of bacterial populations. The Exponential Growth Model is most effective during the exponential phase, where bacteria grow at a constant rate, and doubling time can be calculated using the equation $td = \ln(2)/\mu$, with μ representing the specific growth rate constant. This model assumes abundant resources and no growth limitations, making it straightforward and accurate for this phase. In contrast, the Logistic Growth Model considers the carrying capacity of the environment, accounting for the slowing of growth as resources become scarce. Although more complex, this model is useful for ecological studies where environmental limitations impact bacterial growth. For more intricate growth patterns, the Gompertz and Richards Models are employed, especially in scenarios like food microbiology, where bacteria grow under stress or varying environmental conditions. These models often require numerical methods to determine doubling time due to their complexity. Lastly, the Monod Model describes bacterial growth in environments where the growth rate is limited by substrate concentration. In this model, the doubling time is influenced by the availability of specific nutrients, with calculations based on the relationship between the specific growth rate and substrate concentration.

Dataset and Data Analysis

The growth of *M. smegmatis* was monitored over an 8-hour period under three different conditions. The Optical Density (OD600) was measured hourly to assess bacterial growth. The data is summarized in Table 2 and graphically represented in Figure 2.

Table 2. Optical Density (OD600) Measurements at Different Dilutions

Time (hrs)	OD600 (Undiluted)	OD600 (5×Diluted)	OD600 (10×Diluted)
0	0.1669	0.1599	0.1352
1	0.2157	0.1945	0.1311
2	0.2291	0.2239	0.2020
3	0.2681	0.2274	0.2586
4	0.3097	0.3056	0.2254
5	0.4087	0.3225	0.3405
6	0.4873	0.3760	0.3408
7	0.6579	0.5615	0.4892
8	0.9130	0.7083	0.7429

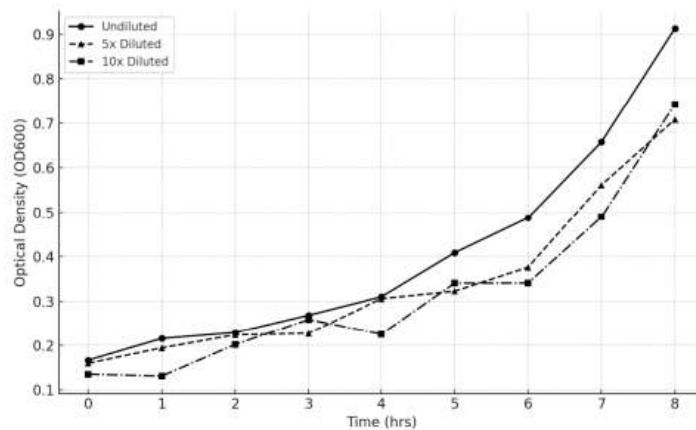


Figure 2. Growth Curve of *Mycobacterium smegmatis* Over an 8-Hour Period in 7H9 Media at 37°C, Comparing Undiluted, 5× Serial Diluted, and 10× Serial Diluted Samples Using Optical Density (OD600) Measurements.

Doubling Time Calculation

The doubling time of *M. smegmatis* was calculated using three distinct methods to ensure the accuracy and robustness of the results.

Growth Curve Analysis: This method involved plotting the OD600 data against time to identify the exponential growth phase. The slope of the growth curve during this phase was used to calculate the doubling time. The exponential growth phase was clearly visible in all three samples, allowing for straightforward calculations. This method provided a baseline estimate of the doubling time under different conditions.

Direct Formula Method: The direct formula method involves calculating the growth rate constant (μ) directly from the change in OD600 between two time points within the exponential growth phase. The doubling time was then derived using the formula $td = \ln(2)/\mu$. This method is useful for quickly estimating the doubling time based on specific time intervals and provides another layer of validation for the results obtained from the other methods.

Linear Regression on Log-Transformed Data: In this approach, the natural logarithm of the OD600 values from the exponential growth phase was plotted against time. A linear regression analysis was then performed on this log-transformed data to determine the growth rate constant (μ). The doubling time was calculated using the formula $td = \ln(2)/\mu$. This method is particularly effective in precisely quantifying the growth rate during the exponential phase, and it generally offers a more accurate estimate of the doubling time compared to the simpler growth curve analysis. Therefore, the results obtained using this method are prioritized for further analysis and discussion.

Table 3. Doubling Times (in hours) for *M. smegmatis* Calculated Using Three Different Methods

Condition	Doubling Time (Growth Curve Analysis)	Doubling Time (Direct Formula Method)	Doubling Time (Linear Regression on Log-Transformed Data)
Undiluted	3.44	3.48	3.35
5× Diluted	3.95	4.13	4.43
10× Diluted	3.44	7.53	5.59

Results and Discussion

This work investigated in detail the growth dynamics of *M. smegmatis*. The central objective of the experiment was to determine the doubling times at various dilution levels by the method of Linear Regression on Log-Transformed Data.

In all samples, we observed a robust phase of exponential growth, particularly pronounced in the undiluted sample, where a rapid increase in optical density was indicative of active cell division. However, the growth curve did not reach a plateau within the 8-hour observation period. This is likely due to the time frame being insufficient, as previous studies have shown that *M. smegmatis* typically requires 20-24 hours for growth to decelerate and enter the stationary phase. Therefore, for future experiments, it is recommended to extend the observation period to over 24 hours to capture the complete growth curve and ensure that the plateau phase is accurately represented.

Further, it is evident that with higher dilutions, the times of doubling do increase, which means that the rate of growth reduces a bit as the concentration of bacteria goes down. For example, for an undiluted sample, the time it takes to double would be around 3.35 hours, which reflects very vigorous growth typical for *M. smegmatis* under optimal conditions. With increased dilution, there was an increase in the doubling time; the 5× diluted sample increased the doubling time to 4.43 hours, while the 10× diluted sample went up to 5.59 hours. The trend that comes out is that even though *M. smegmatis* is still efficient in growth at a lower concentration, probably the reduced density of bacteria might introduce some type of delay in the exponential phase of growth.

This observed increase of the doubling time with the dilution showed the adaptability of *M. smegmatis* in a low-nutrient setting but also posed a limit with regard to how efficiently the bacterium could be kept growing fast under an increasingly diluted setup. This finding thus underscores the importance of initial bacterial concentration when designing experiments or industrial processes that rely on bacterial cultures for optimum growth and productivity. Of the techniques used, utilizing Linear Regression on Log-Transformed Data showed the most reliable for this study. It provides very accurate and robust estimates of doubling time by only considering data from the exponential growth phase where growth is more predictable. This reduces the effect of variability in OD600 readings and lessens the impact of outliers. On the other hand, due to the method's reliance on visual identification of the exponential phase, it may not be that accurate for general growth dynamics. Again, direct formula methods are highly

sensitive to specific time points and, hence less reliable due to the possibility of errors or variations in OD readings.

This growth dynamic of *M. smegmatis* observed here is in good agreement with previously reported growth rates that normally fall between 2.5 and 4 hours under optimal conditions in nutritionally rich media like 7H9 at 37°C. The finding further enhances the reliability of the experimental conditions adopted and indicates that the growth kinetics of *M. smegmatis* is very close to being similar for different initial concentrations. Further, a clear lag and exponential phase can be recognized, which better fits with the established knowledge of the growth behavior of bacteria.

My findings have important implications for industrially relevant applications, particularly in biotechnological processes involving *M. smegmatis*. The implication of the uniformity of the doubling time among all three dilutions at the industrial scale is that variations in initial bacterial concentration will not impact the efficiency of the process. This is of particular interest in the production of recombinant proteins, where the use of *M. smegmatis* is especially important due to its fast growth rate and non-pathogenicity. Optimization of harvest timing in the exponential phase increases yields and productivity, which is directly proportional to a reduction in costs and an increase in efficiency in the industry. Knowledge of the growth kinetics of *M. smegmatis* allows the design of flexible production processes able to adapt to changes in bacterial concentration without loss of productivity.

Materials and Methods

In this study, the bacterial species *Mycobacterium smegmatis* (*Msmeg*) was utilized. *M. smegmatis* is closely related to *Mycobacterium tuberculosis* (MTB), the pathogenic bacterium responsible for tuberculosis, a disease that impacts about one-third of the world's population¹⁹. Investigating the growth of *M. smegmatis* can provide valuable insights that further research on MTB.

The experimental procedure was designed to determine the doubling time of *Msmeg* by measuring the optical density (OD) of the bacterial culture at hourly intervals over an eight-hour period. The bacteria were cultured in 7H9 media and maintained at 37°C in an incubator. The growth medium, 7H9 broth, was prepared according to standard protocols. Additional materials included pipettes with a range of 20 µL to 100 µL, a well plate for OD measurements, a vortex mixer for homogenizing bacterial suspensions, a plate reader for measuring OD, an incubator set at 37°C, appropriate personal protective equipment (PPE), and Eppendorf tubes for sample handling.

The initial OD of the *Msmeg* culture was measured to establish a baseline. The bacterial culture was then diluted to achieve an OD of 0.2, using the following dilution formula:

$$(\text{initial OD})(V_1) = (0.2)(11,000 \mu\text{L})$$

where V_1 represents the volume of the bacterial culture needed.

The diluted culture was aliquoted into a new tube, with a total volume of 11,000 µL. A well plate was prepared by pipetting 200 µL of 7H9 media into the first four wells of the first column. The bacterial culture was vortexed to ensure uniform distribution. Subsequently, 40 µL

of the bacterial culture was mixed with 160 μL of 7H9 media in the same wells, creating a 5 \times dilution. In the second column, 20 μL of the bacterial culture was mixed with 180 μL of 7H9 media, resulting in a 10 \times dilution.

The prepared well plate was inserted into the plate reader, and the initial OD data was recorded. Every hour, an aliquot of 1,100 μL from the incubated bacterial culture was taken, vortexed, and measured for OD. The remaining culture was returned to the incubator. This process was repeated for a total of eight time points. The OD measurements for undiluted, 5 \times diluted, and 10 \times diluted cultures were plotted and analyzed to determine the doubling time.

Works Cited

- Wani, Atif Khurshid, et al. "Microbial Adaptation to Different Environmental Conditions: Molecular Perspective of Evolved Genetic and Cellular Systems." *Archives of Microbiology*, vol. 204, no. 2. Accessed May 20, 2024
<https://doi.org/10.1007/s00203-022-02757-5>.
- Sebastian. *Bacteria: A Very Short Introduction*. OUP Oxford (2013). Accessed June 12, 2024.
- Pelczar, Michael, et al. *Microbiology* (5th ed.). Tata McGraw-Hill (2007). Accessed June 12, 2024
- R. Bailey. Phases of the bacterial growth curve. *ThoughtCo*. Accessed December 14, 2023.
<https://thoughtco.com/bacterial-growth-curve-phases-4172692>
- Gompertz, Benjamin. On the Nature of the Function Expressive of the Law of Human Mortality. (1825). Accessed April 23, 2024.
- Monod, Jaques. The growth of bacterial cultures. *Annual Review of Microbiology*, 3(1), 371-394, (1949). Accessed April 23, 2024.
- Baranyi, Jozsef, et.al. A dynamic approach to predicting bacterial growth in food. *International Journal of Food Microbiology*, 23(3-4), 277-294 (1994). Accessed April 23, 2024.
- Richards. A flexible growth function for empirical use. *Journal of Experimental Botany*, 10(2), 290-301 (1959). Accessed April 23, 2024.
- Buchanan, R.L, et al. "When Is Simple Good Enough: A Comparison of the Gompertz, Baranyi, and Three-Phase Linear Models for Fitting Bacterial Growth Curves." *Food Microbiology*, vol. 14, no. 4, Aug. 1997, pp. 313–326. Accessed June 12, 2024.
<https://doi.org/10.1006/fmic.1997.0125>.
- Zwietering, et.al Modeling of the bacterial growth curve. *Applied and Environmental Microbiology*, 56(6), 1875-1881(1990). Accessed February 30, 2024.
- Madigan, et.al *Brock Biology of Microorganisms* (14th ed.) Pearson (2014). Accessed February 30, 2024.
- Delignette-Muller, M.L. "Relation between the Generation Time and the Lag Time of Bacterial Growth Kinetics." *International Journal of Food Microbiology*, vol. 43, no. 1-2, Aug. 1998, pp. 97–104. Accessed April 3, 2024.
[https://doi.org/10.1016/s0168-1605\(98\)00100-7](https://doi.org/10.1016/s0168-1605(98)00100-7).
- Prasad, Deepika, et al. "Elucidating the Functional Role of Mycobacterium Smegmatis RecX in Stress Response." *Scientific Reports*, vol. 9, no. 1, 29 July 2019, p. 10912,. Accessed 27 Apr. 2023
<https://doi.org/10.1038/s41598-019-47312-3>.
- Cole, S.T, et.al Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 393(6685), 537-544 (1998). Accessed June 12, 2024.

Impact of Technology on Achievement Motivation in Underserved Elementary Students During COVID-19 By Meghna Chellapilla

Underserved elementary school students face significant barriers to accessing quality education, including limited access to technology and the internet. The shift to remote learning during COVID-19 forced schools to adopt various technological solutions, including online classes being held over Zoom and other video conference software. The COVID-19 pandemic only increased the amount of challenges educational institutions encountered, making it essential to explore how existing learning communities and technology usage can bridge the educational gap impacting student populations, especially those in underserved communities. In order to better understand this gap, evaluating which technologies are effective in maintaining or boosting motivation and matriculation among underserved students, especially during the pandemic, helps to identify best practices and areas for improvement. This paper will explore how pre-established learning communities and the use of technology impacted motivation and matriculation in underserved elementary school students during COVID-19.

Underserved student populations refer to groups of students from low-income families and minority groups (Darling-Hammond et al.). In the context of digital divide research, I will explore how limited access to technology disproportionately affected underserved elementary school student populations during COVID-19 online learning, and whether it hindered their educational opportunities and outcomes compared to their peers with better access to digital resources. This includes examining how the lack of necessary devices and reliable internet during the pandemic exacerbated existing educational disparities for students. More specifically, I will explore how these factors impacted achievement motivation overall. In order to do this, I will investigate how the digital divide influenced the exacerbation of learning inequities during COVID-19, how both traditional learning and new technologies were implemented into curriculums, and lastly, how learning communities support underserved student learning.

In order to explore student motivation, I will be using the Heckhausen & Rheinberg's (1980) model and theory for achievement motivation, shown in figure 1.1. This model on achievement motivation defines achievement motivation as “an individual’s motivation to aspire to a certain goal”(*Motivation and Action*, chapter 6). More specifically, I will expand on achievement motivation as it relates to underserved elementary school student populations during the COVID-19 epidemic.

Theme A Learning inequities were exacerbated by COVID

The theme of learning inequities exacerbated by COVID-19 emerges prominently across multiple studies, each highlighting different facets and commonalities of this issue. A central characteristic shared by these studies is the identification of the digital divide as a barrier to equitable education. The articles emphasize that pre-existing disparities in access to technology have long contributed to varied educational outcomes, with limited access hindering students' ability to participate fully in online learning. These inequities were particularly pronounced

among students from low-income households, Black, Indigenous, and People of Color (BIPOC) communities, and rural areas.

Learning inequities have long existed due to varying levels of access to technology in school populations. In a survey of 400 students from diverse educational backgrounds conducted by Afzal et al., there was a stark correlation between limited access to technology and lower educational outcomes (Afzal et al.). This lack of access has resulted in students being unable to fully participate in online classes, leading to disparities in academic performance and engagement even before the pandemic (Afzal et al.).

The survey's findings further illustrated that students with limited access to technology not only struggled academically but also faced socio-emotional challenges. The lack of reliable devices, as well as internet connectivity, hindered students' ability to complete assignments on time, collaborate with peers, and seek help from educators (Afzal et al.). This digital divide also affected their confidence and motivation, as they felt increasingly isolated and left behind compared to their more technologically-equipped classmates (Afzal et al.). Providing adequate technology and internet access is crucial in mitigating these disparities and supporting students' learning, particularly in times of crisis. These pre-existing inequities set the stage for even greater disparities when the COVID-19 pandemic forced a rapid shift to remote learning.

Tager investigated the impact of COVID-19 on technology access in early childhood education. This study found significant disparities in technology access between BIPOC and white children. These disparities, rooted in pre-existing school and residential segregation, were only made worse by the pandemic (Tager). Schools serving predominantly BIPOC communities often lacked the resources to provide adequate technology and internet access, putting these students at a disadvantage when remote learning became the norm (Tager).

COVID-19 magnified these learning inequities by forcing schools to rely heavily on technology for instruction (Tager). Funding inequities in the school system, which were already present before the pandemic, played a critical role in exacerbating the digital divide. Schools in underfunded districts could not supply sufficient devices or ensure reliable internet connections for their students, leading to a wider gap in educational opportunities (Tager). This lack of resources meant that many BIPOC children were unable to participate fully in remote learning, much less have all their educational needs met.

Additionally, the residential divide that contributes to technology disparities became even more problematic during the pandemic. Families in lower-income neighborhoods often face multiple challenges, including overcrowded living conditions and unstable internet access, making it difficult for children to engage in online learning effectively (Tager). This study underscores that the pandemic did not create these inequities, but instead highlighted and intensified them.

A study conducted by Haderlein et al. investigated the impact of COVID-19 on learning, specifically focusing on different American household demographics. The study found significant disparities in access to technology and instructional resources, which were heavily influenced by factors such as race, ethnicity, income, and urban versus rural settings. These

disparities created a significant barrier to equitable education during the pandemic.

This digital divide prevented many students from engaging fully in their education, leading to widening gaps in learning outcomes between different demographic groups (Haderlein et al.). Moreover, the pandemic disrupted services that many families relied on, such as free and reduced-price meals, which were less accessible during school closures (Haderlein et al.). The lack of access to these services added another layer of difficulty for low-income families, further hindering students' ability to focus on their education. The pandemic not only highlighted these existing disparities but also intensified them, making it clear that addressing these inequities is crucial for ensuring all students have the opportunity to succeed in their education (Haderlein et al.).

Tager and Haderlein et al. both investigated similar questions regarding the digital divide and its impact on student learning during the pandemic, emphasizing the role of technology access in exacerbating educational inequities. These studies both highlighted how limited access to technology during the pandemic further widened the gap in educational outcomes among different demographic groups. They focused on the challenges faced by students as schools transitioned to remote learning, highlighting the critical need for equitable access to technology to support student learning and engagement during times of crisis.

In contrast, Afzal et al. gave a broader context for understanding these issues by exploring the digital divide's impact on educational outcomes before the pandemic. This broader perspective highlighted the longstanding nature behind these inequities, and the ways in which the pandemic has exacerbated them further (Afzal et al.). The study emphasized that the digital divide has always been a significant barrier to equitable education, and the pandemic merely intensified these pre-existing challenges. The common thread across these studies is the urgent need to address the digital divide and ensure equitable access to technology for all students, particularly in times of crisis.

The COVID-19 pandemic has forced education to evolve in ways that center on what would have been previously considered non-traditional ways of learning. Through this process, we have identified that a digital divide is present and is a barrier to equitable education. Although we know that learning inequities have long existed due to varying levels of access to technology in school populations, education in the time of COVID-19 has further highlighted that students with insufficient access to technology faced even greater challenges as schools transitioned to online learning. These significant disparities in access to technology and instructional resources were heavily influenced by factors such as race and ethnicity (Tager).

Theme B Leveraging new technology

The onset of the COVID-19 pandemic necessitated a greater reliance on multimedia learning technologies as schools transitioned to online learning environments. Powers et al. examined how the SPARK method could address both the academic and social-emotional needs of students during the height of the pandemic and online learning. This model integrated multimedia tools to create a framework that other educators and school districts could adopt

(Powers et al.). By leveraging these technologies, educators were able to repurpose their instructional practices to meet the challenges posed by remote learning, ensuring that students remained engaged and supported during the uncertain times the pandemic posed (Powers et al.).

The theme of leveraging new technology in education overall reveals various nuances, highlighting the potential of multimedia learning tools to enhance student engagement, motivation, and academic performance. Before the pandemic, educational technologies, including hardware and software tools, were already integrated into classrooms to make lessons more engaging. A study by Hanif and Puspitarini found that these technologies, including hardware and software tools, serve as effective mediums for delivering educational materials (Puspitarini et al). Traditional textbooks often provide limited and unengaging information, leading to student disinterest and pose minor challenges in understanding abstract concepts. While schools are equipped with information and communication technology (ICT) tools such as computers, projectors, and audio sets, these resources were often underutilized in teaching. By incorporating these technologies, educators were able to create more interactive and stimulating learning environments, thereby increasing student motivation and improving learning outcomes—adapting learning environments to be more suitable for educators and students during COVID-19 (Puspitarini et al).

One significant area where multimedia technology demonstrated its effectiveness was in mathematics education. In another study, Shin et al. explored how game technology could improve students' conceptual understanding of math lessons. The findings revealed that educational games significantly enhanced elementary students' mathematics scores, compared to traditional instruction typically given in the classroom (Shin et al.). These games made learning more enjoyable and encouraged active participation in the classroom, which in turn increased student engagement and motivation (Shin et al.). Notably, students who initially struggled with math benefited the most from this approach, highlighting the potential of game technology to support diverse learning needs (Shin et al.).

Furthermore, the benefits of using educational games extended beyond academic performance. Students who used game technology retained mathematical knowledge better over time, instilling a deeper and more long-lasting understanding of concepts (Shin et al.). This long-term retention suggests that multimedia learning tools not only make learning more engaging, but also enhance the overall quality of education as a result. The reliance on these technologies grew as COVID-19 forced both educators and students to adapt to a 100% virtual learning setting (Shin et al.). The usefulness of these technologies demonstrates their critical role in supporting student learning and motivation in an increasingly digital educational landscape.

Hanif et al. and Shin et al. both provided practical insights into the use of learning media in classrooms, highlighting their effectiveness in enhancing student engagement and academic performance. Hanif et al. demonstrated that integrating ICT tools like computers and projectors could make lessons more engaging and diverse, therefore increasing student learning motivation. Similarly, Shin et al. showed that educational games significantly improved students' mathematics scores and retention of knowledge. These studies suggest that pre-established

learning communities equipped with multimedia tools can positively impact student motivation, particularly in underserved populations where traditional teaching methods may fall short. The importance of learning communities will be explored later in this paper.

In contrast, Powers et al. conducted a study in the midst of the pandemic, which presented a theoretical framework for addressing both academic and social-emotional needs through innovative methods during that period of online learning. While it did not provide empirical data, it offered a comprehensive approach that other educators could apply to improve learning outcomes in remote settings. This theoretical model emphasizes the importance of integrating multimedia tools to support student engagement and motivation during uncertain times, which is crucial for maintaining educational continuity in underserved communities. However, it is important to note that although both articles shared good insight on the implementation of technology in elementary school classrooms, none of them specifically address learning in underserved communities.

Furthermore, none of these studies specifically addressed the dual impact of pre-established learning communities and technology on motivation in underserved elementary school students during COVID-19. This gap is significant, because understanding how these factors interact could inform more targeted interventions to support students in the demographic of underserved students.

Subtheme A good deal of teachers are against technology implementation

In a study by Avidov-Ungar et al., researchers delve into the challenges faced by digital leaders in education. These leaders, often tasked with facilitating the use of new technologies in teaching, frequently encounter resistance from teachers who prefer traditional methods (Avidov-Ungar et al.). This resistance highlights a more universal tension within the educational system, where the push for technological integration clashes with pedagogical traditions that have been in practice for decades. Despite the digital leaders' efforts to envision and implement change, their roles are often perceived through the lens of conventional educational roles, complicating their mission to aid their students' learning to the best of their ability (Avidov-Ungar et al.).

The study reveals that a significant number of teachers are reluctant to embrace new technology, stemming from various factors such as comfort with established methods, skepticism about the efficacy of the technology, and concerns about the additional workload associated with learning and implementing new tools (Avidov-Ungar et al.). This hesitancy can create a barrier for digital leaders, who must navigate not only the technical and logistical aspects of technology integration but also the cultural and attitudinal shifts required among their colleagues (Avidov-Ungar et al.). The prevalence of traditional visions of change, as noted in the study, highlights a common challenge: aligning the innovative potential of digital tools with the time-tested practices of educators. This research demonstrates how traditional visions often depict digital leaders as outsiders or disruptors rather than integral members of the teaching community (Avidov-Ungar et al.). Overall, the study suggests that addressing this divide requires

not only strategic planning and professional development but also a concerted effort to reshape the cultural narrative around technology in education (Avidov-Ungar et al.).

Theme C The importance of learning communities

The theme of enhancing student motivation in underserved populations is characterized by various support systems, including familial, educator, and peer support. These studies emphasize the importance of a holistic and inclusive approach to creating a supportive learning environment that can significantly boost student motivation and success. A study conducted by McPartland and Braddock explored barriers to student motivation and strategies to enhance it. The research highlighted that the active involvement of families and communities in the educational process is crucial for supporting student motivation and success (McPartland et al.). This involvement helps create a network of support that reinforces students' efforts and achievements.

Positive peer interactions and relationships also contribute significantly to a supportive learning environment. Consequently, motivation is enhanced, and students are shown to thrive in settings where they feel connected and supported by their peers (McPartland et al.). The study emphasized the importance of these interactions in fostering a sense of belonging and motivation among students, further pointing out the need for training teachers in culturally responsive teaching practices and strategies to support underserved student communities (McPartland et al.). Culturally responsive teaching helps students feel valued and understood, which can significantly increase their motivation in school to participate and learn (McPartland et al.). Schools that promote a positive, culturally responsive atmosphere are more successful in creating an inclusive environment where all students can thrive (McPartland et al.).

During the COVID-19 pandemic, Ambriz et al. found that many students built online learning communities to help each other use technology and increase learning (Ambriz et al.). Learning pods emerged as a popular solution, where small groups of students learned together, providing a mix of education and childcare (Andersson et al.). These pods often utilized resources, like Khan Academy, and supported each other through the challenges of remote learning (Ambriz et al.). Together, these two studies focus on different types of support that enhance student motivation while continuing to uplift the importance of learning communities in a student's academic journey. Ultimately, the themes of familial support, educator support, and online and in-person learning peer support are highlighted as a solid, must-have foundation for students to be supported throughout their learning journey.

Conclusion

The existing literature provides a comprehensive analysis of how multimedia technology and learning communities impact student motivation and educational outcomes, and how existing learning barriers, as well as tools, could have impacted underserved elementary school students during the COVID-19 pandemic. The pandemic underscored the significant role of technology in enhancing learning motivation while highlighting the exacerbation of learning

inequities due to disparities in access to technology. However, there are gaps in this research, such as limited data and insufficient focus on the combined effects of technology and learning communities, that need further exploration.

Existing literature advances our understanding of the role of multimedia technology and learning communities in education. Notable studies demonstrate how multimedia tools can enhance student engagement and motivation while identifying the barriers faced by underserved populations, especially during the shift to remote learning caused by COVID-19. Additionally, it emphasizes the importance of familial, peer, and educator support in fostering student motivation and success, highlighting the multifaceted nature of educational inequities.

However, the literature also identifies significant barriers to the widespread adoption and effectiveness of these technologies, primarily related to access to funding and the cost of technology. These barriers are particularly pronounced in underserved communities, where limited access to technology exacerbates pre-existing educational inequities. The COVID-19 pandemic further highlighted these disparities, as schools' rapid shift to remote learning underscored the critical importance of adequate technological infrastructure, overall leading to widening gaps in learning outcomes among different student populations.

The literature suggests an urgent need to address these barriers to ensure that all students can benefit from the educational advantages of multimedia technology. This involves securing funding and resources to provide adequate technology and internet access, and implementing policies and practices that support the integration of these tools into the curriculum. Addressing these challenges is crucial for mitigating educational disparities and ensuring that multimedia technology can fulfill its potential to enhance learning motivation and outcomes for all students, particularly those in underserved communities.

However, we must recognize that there are several gaps in the current literature. There is a lack of long-term data on the sustained impact of multimedia technology and learning communities, which limits understanding of their prolonged effects on student outcomes both in this paper and beyond it. Long-term studies on the effects of multimedia technology are also needed to assess their sustained impact on student motivation and academic performance. Furthermore, a broader demographic analysis within underserved populations would be essential to ensuring that findings are widely applicable and inclusive of diverse groups.

Future Research

Future research should prioritize conducting long-term studies to assess the sustained impact of multimedia technology and learning communities on student motivation and academic outcomes. While short-term studies have shown positive effects, understanding how these tools influence learning over extended periods is crucial. Long-term research can provide insights into whether the initial motivational and educational gains from multimedia technology persist over time, and how to tweak them as a student progresses into their educational journey.

Additionally, future studies should investigate the combined effects of multimedia technology and pre-established learning communities on student outcomes. Research so far has

often examined these factors in isolation, but understanding their interaction is essential for developing comprehensive educational strategies. By exploring how these elements work together, researchers can develop targeted tools that leverage the strengths of both approaches to support underserved students.

Expanding research to include diverse demographic groups within underserved populations is another critical step. Current studies often focus on specific contexts or groups, which limits the generalizability of their findings. Future research should aim to include a broader range of students from various backgrounds of: socioeconomic statuses, ethnicities/races, and geographical locations. This broader approach will ensure that findings are more inclusive and applicable to a wider audience, addressing the unique challenges faced by different communities and providing more comprehensive solutions to educational inequities.

Finally, exploring policy interventions and funding mechanisms to address the digital divide should be a focal point of future research. Securing adequate resources is essential for providing equitable access to technology and ensuring that all students can benefit from its educational advantages over the course of their education. Researchers should investigate effective funding strategies, such as government initiatives, to support the widespread adoption of multimedia tools in schools. Additionally, examining how policies can promote the sustainable integration of technology into the curriculum will help ensure that these resources are used effectively to enhance learning motivation and outcomes over the long term. By addressing these aspects, future research can contribute to developing a more equitable and effective educational landscape for all students from a diverse set of backgrounds, who each face their own challenges.

Works Cited

- Afzal, A., Khan, S., Daud, S., Ahmad, Z., & Butt, A. (2023). Addressing the digital divide: Access and use of technology in education. *Journal of Social Sciences Review*, 3(2), 883-895.
- Andersson, F. O., & Willems, J. (2024, July). The rise of learning pods: Civil society's expanding role in K-12 education in the United States. In *Nonprofit Policy Forum* (Vol. 15, No. 3, pp. 249-262). De Gruyter.
- Avidov-Ungar, O., Shamir-Inbal, T., & Blau, I. (2022). Typology of digital leadership roles tasked with integrating new technologies into teaching: Insights from metaphor analysis. *Journal of Research on Technology in Education*, 54(1), 92-107.
- Brunstein, J. C., & Heckhausen, H. (2018). Achievement motivation. *Motivation and action*, 221-304.
- Haderlein, S. K., Saavedra, A. R., Polikoff, M. S., Silver, D., Rapaport, A., & Garland, M. (2021). Disparities in educational access in the time of COVID: Evidence from a nationally representative panel of American families. *AERA Open*, 7, 23328584211041350.
- Lee, J. C., Quadlin, N., & Ambriz, D. (2023). Shadow education, pandemic style: Social class, race, and supplemental education during Covid-19. *Research in Social Stratification and Mobility*, 83, 100755.
- McPartland, J., & Braddock, J. H. (1909). A conceptual framework on learning environments and student motivation for language minority and other underserved populations. *Learning*, 3, 06.
- Powers, J. M., Brown, M., & Wyatt, L. G. (2020). SPARK-ing innovation: A model for elementary classrooms as COVID-19 unfolds. *Journal of Professional Capital and Community*, 5(3/4), 307-320.
- Puspitarini, Y. D., & Hanif, M. (2019). Using learning media to increase learning motivation in elementary school. *Anatolian Journal of Education*, 4(2), 53-60.
- Shin, N., Sutherland, L. M., Norris, C. A., & Soloway, E. (2012). Effects of game technology on elementary student learning in mathematics. *British Journal of Educational Technology*, 43(4), 540-560.
- Tager, M. B. (2022). Technology segregation: A pandemic crisis in early childhood education. *International Critical Childhood Policy Studies Journal*, 8(2), 63-79.

Breaking Domestic Chains: A Comparative Analysis of Feminist Ideals in Wollstonecraft and Lu Xun By Xinyue Chen

Abstract

Enlightenment, a male-dominated intellectual movement that began around 1685, worships ideals such as reasoning, liberty, and scientific progress. This movement reshaped European thoughts, on the other hand, China experienced a parallel ideology awakening nearly 300 years later: the May Fourth Movement. Among all the philosophical pursuits, theological criticisms, and aesthetic glorifications, the voice of women. This essay explores the articulation of women empowerment between two of the most influential writers of these periods: Mary Wollstonecraft and Lu Xun.

Mary Wollstonecraft challenges the patriarchal education system and God-centric marriage norms that enslaved women to dependent roles. She advocates for a comprehensive education plan aimed at fostering women's intellectual and moral development, envisioning a “brighter” future where women can achieve independence and equality. In hawkish contrast, Lu Xun cynically exposes the harsh tragedies faced by women who, despite being educated and gaining freedom of love, experience socioeconomic suffocation. There needs to be more than the self-reformation proposed by Wollstonecraft to provide women with secured autonomy. He argues that mere self-reformation, proposed by Wollstonecraft, is insufficient for securing women’s autonomy. Instead, Lu Xun proposes that a broader social revolution is necessary to liberate Chinese society from its feudal mindset before discussing feminism. This mass-centric and socio-economic-driven argument, rooted in his proletariat perspective, differs from the upper-class viewpoint of Wollstonecraft. Both authors provide radical ideas that contribute to the early thoughts of each feminist movement; however, Lu Xun’s approach appears more pragmatic, addressing that women's empowerment is a systemic issue rather than solely on individual progression.

Introduction

Mary Wollstonecraft’s *A Vindication of the Rights of Woman* and Lu Xun’s *Regret for the Past* offer divergent perspectives on women's roles and rights, influenced by their distinct historical and social contexts. *A Vindication of the Rights of Woman* is a seminal work in feminist philosophy and advocacy published in 1792. In this groundbreaking text, Wollstonecraft, who had a tumultuous romantic life with two short-lived relationships, challenges the prevailing notions of her time regarding the roles and education of women. She criticizes the traditional focus of education on making women pleasing to men rather than cultivating their intellectual and moral capacities; proposes that marriage should be based on mutual respect and equality, as a friendship rather than a hierarchical institution; and emphasizes the need for women to be recognized as rational beings capable of making informed decisions. For this article, we mainly focus on Mary’s vision of marriage and her theory of “the correct education” for women in that society.

Lu Xun is widely regarded as one of the most influential Chinese writers of the 20th century. His work is closely associated with the reflection and portrayal of the May Fourth Movement, a cultural and intellectual upheaval that sought to modernize Chinese society from imperial archaic mindset through the promotion of new ideas and the rejection of traditional norms, including those related to gender roles and romantic relationships. One of the notable aspects of this cultural shift was the growing discourse around free love and individual autonomy in romantic relationships. In this historical context, Lu Xun's literature often explores the tragedies that arise from the blind pursuit of new ideas and reforms without a deep understanding of their implications or readiness for their consequences. Many of his works, including *The True Story of Ah Q*, a satirical tragedy about a young peasant who drowns in his blind optimism and his tendency to rationalize his failures as triumphs under the superficial adoption of new ideas without genuine understanding, and his other work *The Medicine*, the story of a man who seeks a cure for his son's tuberculosis through a traditional folk remedy involving human blood, reveal the real-world consequences of clinging to outdated beliefs while pursuing new ideas without fully grasping their implications. For this article, we will focus on Lu Xun's story *Regret for the Past*, which depicts two protagonists, Juansheng and Zijun, who are both well-educated, elope for free love but face tragic outcomes due to economic hardships and societal constraints in China at that time. The ending of these well-educated women differs from the theory presented in Wollstonecraft's *A Vindication of the Rights of Woman*. This essay analyzes the historical and social contexts that shaped the differing perspectives of Wollstonecraft and Lu Xun. Through the lens of class-conflicting theory, it argues that Lu Xun's thoughts, constructed independently of social pressures, offer a more progressive and inclusive vision compared to Wollstonecraft's class-biased approach.

Criticism of Social Norms Toward Women

While Wollstonecraft sought to reform existing norms, Lu Xun addressed the societal expectations reflected in his literature. Their different social backgrounds led to divergent focal points in their works. Wollstonecraft engaged with Enlightenment ideals to critique patriarchal structures confining women. She argued that women's education was profoundly unequal compared to that of men, criticizing a system that prepared women merely for marriage rather than for intellectual and personal growth. She stated, "Teach them in common with man to submit to necessity, instead of trying to make them more pleasing by giving a sex to morals," underscoring her belief that women's education was limited to domestic roles. In contrast, Lu Xun challenged the Confucian emphasis on filial piety and imperial authority, which reinforced hierarchical relationships. He explored the aftermath of "free love" in his literature, embracing it as a symbol of liberation from traditional marriage. In *Regret for the Past*, Lu Xun depicts Juansheng and Zijun, who elope for love but face tragic outcomes due to economic hardships and societal constraints. Zijun's declaration, "I'm my own mistress. None of them has any right to interfere with me," aligns with Wollstonecraft's assertion that women should have power over themselves. However, Lu Xun underscores the importance of financial

stability in sustaining relationships, as seen in Juansheng's lament, "the wretchedness of my present life was largely due to her." Lu Xun was acutely aware that adopting Western ideals without addressing China's unique social realities could lead to further oppression rather than true liberation. His portrayal of educated women reflects the complex social dynamics of his time when there were limited opportunities for women to live independently.

During the May Fourth Movement, many women received Western-style education, leading to a clash between their new ideas and traditional societal structures. Lu Xun's works often depict the tragedy of well-educated women struggling to reconcile their ambitions with societal expectations. Conversely, the suffering of uneducated women is often interpreted subjectively by readers who connect it to the lack of societal progress and personal indifference. For instance, *The New Year's Sacrifice* tells the tragic story of Xianglin's Wife, an uneducated female housemaid who suffers through three marriages, is shamed by villagers, and ultimately commits suicide. She represents uneducated women whose suffering is compounded by societal oppression, inviting readers to reflect on broader social and political stagnation, particularly when her master abandoned her due to her widow identity.

Discussion of Love: Anti-Sentimentalism vs. Criticism Toward Alienation

Wollstonecraft's perspective on love can be characterized by a critique of the prevailing sentimental notions of romance that prioritize emotion over rationality. She argues that a true partnership should be built on mutual respect and equality rather than societal expectations. For example, she states, "I do not wish them [women] to have power over men; but over themselves." This highlights her belief in the necessity of self-possession and autonomy for women in romantic relationships. Wollstonecraft emphasizes the importance of self-reflection and intellectual development, proposing that women should strive for personal growth rather than conform to societal pressures to be pleasing to men.

In contrast, Lu Xun presents a more cynical view of love, particularly in the context of societal constraints. In *Regret for the Past*, the characters Juansheng and Zijun elope, believing in the ideal of free love. However, they are quickly met with the harsh realities of economic hardship, revealing the limitations of their romantic aspirations. Lu Xun's perspective underscores the alienation that arises from the gap between idealized love and the socio-economic realities that dictate individuals' lives. Juansheng laments the burdens of their situation, asserting, "The wretchedness of my present life was largely due to her." This stark realization demonstrates the harsh consequences of romantic ideals in a society where economic conditions severely limit personal freedom and agency.

Discussion on Economic Subjugation: A Form of Slavery

The theme of economic subjugation, viewed by Wollstonecraft as a form of slavery for women, is central to her arguments. She emphasizes the lack of opportunities for women to thrive outside traditional roles, asserting that without financial independence, women are left vulnerable and powerless. For instance, Wollstonecraft critiques the educational system that

prepares women solely for marriage, stating, “Teach them in common with man to submit to necessity, instead of trying to make them more pleasing by giving a sex to morals.” This encapsulates her belief that women must be educated to become self-sufficient and capable of making informed choices.

In contrast, Lu Xun’s works reflect the devastating impact of economic constraints on relationships and individual aspirations. The plight of Juansheng and Zijun in *Regret for the Past* illustrates how societal and economic pressures thwart their attempts at independence and love. The phrase “What Happens after Nora leaves home?” from which text captures Lu Xun’s perspective on the precariousness of women’s autonomy in the face of economic realities. He suggests that when women leave their homes seeking independence, they often face dire consequences, such as prostitution or starvation, as they lack the support and resources to thrive. This critique highlights the systemic barriers that inhibit women from achieving true autonomy.

Reformation vs. Revolution: A Marxist Perspective from Lu Xun

Lu Xun’s focus was on the masses, particularly the proletariat, influenced by Marxist theory. This theory asserts that prioritizing the needs and struggles of the lower classes is essential since such oppression serves as a significant barrier to societal progress.

He envisioned a class-free society from the constraints of imperial rule and the remnants of feudalism, believing that only by dismantling these oppressive structures could a new China emerge. For him, the primary obstacle to women’s autonomy lies in the ignorance stemming from their social backgrounds, rather than merely individual knowledge. He depicted the stagnation of popular thought under imperial politics, suggesting that true progress requires the elimination of these oppressive systems.

In *The True Story of Ah Q*, Lu Xun illustrates how the lower classes are exploited, and how their labor and thoughts remain disconnected from their realities. He argued that only by creating a classless society could individuals reconnect with their labor and pursue meaningful lives. This societal framework does not provide space for women’s autonomy, which further complicates their quest for liberation.

While Wollstonecraft emphasized the necessity of equal education for women, stating that “the education of women has been the same as that of men, except its being more regulated by external manners and accomplishments,” Lu Xun’s perspective suggests that education alone is insufficient. He posited that the liberation of women, and indeed all oppressed groups, requires not only equal rights but also a fundamental restructuring of society’s underlying values and power dynamics.

The difference in the author’s perspectives on feminist autonomy reflects their class gaze and social contexts. Wollstonecraft, as a member of the upper class, primarily addressed issues faced by women of her social standing. Her arguments for a “proper education” and “dignity” were framed within a bourgeois context, making her solutions less applicable to women from lower social classes. In her book, she likened these women to “the feathered race,” drawing an evocative comparison to birds in a cage. She critiques the artificiality and vanity of upper-class

women, suggesting that their education renders them vain and helpless. This stance reveals her pro-bourgeois viewpoint and her critique of the old ruling class. However, her analysis does not extend to a broader, more inclusive consideration of all women, particularly those from lower classes.

Wollstonecraft's personal experiences, particularly her tumultuous romantic life, serve as poignant examples of the consequences of societal stagnation. Her first relationship with Gilbert Imlay began with economic stability, as Imlay was a successful American businessman, providing Wollstonecraft with a degree of financial security. However, this stability proved precarious and ultimately short-lived. As Imlay's interest waned, Wollstonecraft's financial situation became increasingly unstable. Her dependence on him exposed her vulnerability when the relationship deteriorated, highlighting how the lack of financial independence and societal support for women contributes to their oppression and limits their opportunities, even if they have received an education.

Conclusion

In summary, while both authors addressed the oppression of women, their approaches and focal points differed significantly due to their class perspectives. Wollstonecraft's work laid the foundation for Western feminism, emphasizing education and autonomy within the upper class.

In contrast, Lu Xun's writings highlight the importance of addressing broader social and economic inequalities to achieve true progress, reflecting a more inclusive and transformative vision for society. The analysis of their works reveals the complexities of women's autonomy and the need for a nuanced understanding of the socioeconomic factors that shape their experiences.

Work Cited

- Ferguson, S. (2021). The Radical Ideas of Mary Wollstonecraft.
<https://doi.org/10.32920/14637867.v1>
- Lu, H. (1972). The true story of ah q. Foreign language Press.
- Lu, X. (1970). The New Year Sacrifice.
- Lu, X. (1998). Regret for the past: Juansheng's Notes.
- Nara Teriyaki - shoreline, WA. Yelp. (n.d.).
<https://www.yelp.com/biz/nara-chinese-restaurant-shoreline>
- Wollstonecraft, M. (n.d.). A vindication of the rights of woman with strictures on ...
<https://www.earlymoderntexts.com/assets/pdfs/wollstonecraft1792.pdf>
- Zhu, P. (1970, January 1). The anamorphic feminine: History, memory, and woman in Lu Xun's writings. SpringerLink. https://link.springer.com/chapter/10.1057/9781137514738_3

Density in Delhi from 2014-2024 By Angad Singh Miglani

Abstract

Global warming is becoming a more pressing issue day by day. The increase in population has given rise to megacities with vast concrete expanses. One such city is Delhi, India, which is the location for this study. Concretization increases the urban heat island effect, raising land surface temperature. However, various policies are being implemented to counteract this increase in temperature. One major effort is to establish green spaces, i.e. increase vegetation. This research gauges the impact of the Normalized Difference Vegetation Index (NDVI), a metric of vegetation level, on land surface temperature (LST). Using data from Google Earth Engine, I analyzed this relationship between LST and NDVI and discovered a negative correlation between them (with some year-to-year variance in the strength of that relationship). On average, each 0.1 unit increase in NDVI saw a 1.392 °C decrease in land surface temperature. This variability shows that there are also other factors affecting land surface temperature, highlighting the need for greater research into which conditions yield the most effective decrease in LST by adding vegetation to a given geographical area.

Keywords LST, NDVI, Vegetation Density, GEE, Geospatial, Delhi

Introduction

Over the past few years, urbanization has grown worldwide, causing metropolitan areas to expand to accommodate rapidly increasing populations. The United Nations (2018) estimates that 68% of people worldwide will live in cities by 2050, up from 55% in 2018. "Megacities," or urban areas with a population of over 10 million, have emerged due to this expansion (Jin et al., 2021). The Urban Heat Island Effect, or UHI, causes higher land surface temperatures (LST) in big urban centers because of their dense infrastructure, large concrete surfaces, and higher heat outputs (National Geographic, 2022).

Though megacities are now emerging worldwide, India has seen the fastest rate of urbanization, with cities like Delhi, Kolkata, and Mumbai seeing notable population increases of 1% to 2% starting in 2023 (United Nations, 2024). Rapid expansion in these places has increased the amount of concrete cover, worsening the urban heat island effect (Beiser, 2019). The automobile driving, industrial processes and energy consumption of humans have caused metropolitan areas to be warmer than nearby rural areas (Testbook, 2023). Rising urban land surface temperatures have detrimental effects on the environment and human health both now and in the future. In addition to increasing one's risk of heatstroke, these health problems include headaches, skin rashes, and dehydration (Reagan Gyimah et al., 2023).

India has implemented several initiatives to combat rising temperatures. One of these is to increase urban greenery and tree density. To reduce surface temperatures, efforts are being made to expand green areas, such as parks and tree planting (Narayan & Dhindaw, 2021). To create cooler urban environments, Kerala also intends to implement cool pavements composed of

permeable and reflecting materials. These pavements will assist in controlling stormwater runoff and lower surface temperatures. Although there is some green space, such as the expansive parks in Gurugram, the recent construction of parks in more recent areas indicates that efforts to lessen the heat island effect in this Indian city have slowed down (Pati, 2023).

This research paper examines the statistical correlation between Land Surface Temperature and Normalized Difference Vegetation Index in Delhi over 11 years, from 2014-2024. In other words, it aims to see if the density of vegetation and the existence of urban forests, vegetation, or green cover affects the Land Surface Temperature of the city of Delhi. This research will help gauge the effectiveness of urban forests in mitigating the heat island effect.

Literature Review

Studies on the correlation between land surface temperature have been conducted in cities worldwide. Research on the correlation between Land Surface Temperature (LST) and Normalized Difference Vegetation Index (NDVI) has yielded informative results in how vegetation can abet rising temperatures in urban areas. A brief review of LST and NDVI (and other related metrics) is included below for context.

NDVI, or Normalized Difference Vegetation Index, is a widely-used remote sensing metric that quantifies vegetation health and density by analyzing the reflectance of light in the red and near-infrared (NIR) wavelengths (Team Cropin, 2021). It is calculated using the following formula:

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$$

Other remote sensing metrics for vegetation density are also prevalent, like MSAVI, the Modified Soil-Adjusted Vegetation Index, and NDRE, the Normalized Difference Red Edge Index (Liang & Wang, 2020). MSAVI modifies the NDVI calculation to reduce the influence of soil reflectance, making it more suitable for early crop growth stages when the canopy is not fully developed. MSAVI is particularly useful in detecting vegetation health during early growth phases, where NDVI might misinterpret bare soil as low vegetation (Auravant, n.d.). However, since MSAVI was mainly targeted towards agriculture and farm analysis and not the analysis of vegetation in urban areas, it was less suitable for this study. NDRE is a metric that utilizes the red edge band, which is sensitive to chlorophyll content, making it practical for assessing crop health, particularly in later growth stages. In comparison to NDVI, NDRE is more sensitive to changes in chlorophyll. However, since its primary use case lies in crop health and does not assess the general existence of vegetation over a particular area, this metric was also suboptimal for this research study. Given these considerations, NDVI was chosen as the preferred vegetation index for this study.

Land Surface Temperature is a metric calculated by the balance of incoming solar radiation and outgoing thermal radiation, representing the temperature of the Earth's surface as it

would be felt by a human observer. The calculation of Land Surface Temperature can be seen through the following flow chart (Khan et al., 2022):

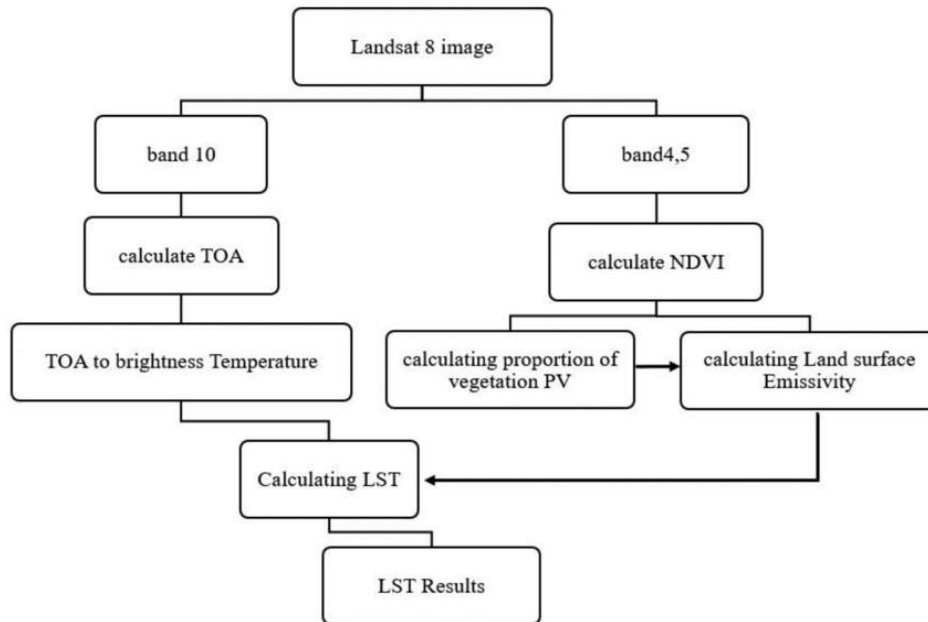


Figure 1. Flow Chart Depicting LST Calculation.

In Figure 1, TOA, or top of atmosphere, refers to the measurements or data collected at the upper boundary of the Earth's atmosphere (Young, 2003).

Even though the calculation of LST uses NDVI, there is no direct mathematical correlation. (The processing of raw data from satellites to Land Surface Temperature will be shown in the 'Data' section of this paper.) This is because of a variety of reasons. First, interactions between vegetation and temperature are often non-linear. For example, as vegetation cover increases, LST may initially decrease, but beyond a certain point, other factors may dominate, leading to an increase in LST despite higher NDVI values (Guha & Govil, 2020). Furthermore, there is spatial and temporal variability as the relationship between NDVI and LST is not static and can change over time and across different geographical locations.

For example, during drought conditions in various North American regions during the summer growing season, vegetation may become stressed, resulting in lower NDVI values. Similarly, LST may increase due to reduced evapotranspiration. In the tropical city of Raipur, India, Subhanil Guha (2021) found that LST values reached their maximum during the pre-monsoon period, while NDVI values decreased over time. Contrastingly, in Bokaro, India, there was a spike in NDVI values due to reduced anthropogenic activities, which allowed for increased vegetation cover. Despite this increase in NDVI, the area also experienced higher LST values. Moreover, LST is calculated as affected by multiple variables beyond vegetation cover, including soil moisture, atmospheric conditions, and land use types (Ullah et al., 2023).

In another prominent study by Guha et al. (2018), the relationship between LST and NDVI within Florence and Naples City, Italy demonstrated a strong negative correlation between these two metrics. Furthermore, research conducted within India, in urban areas such as Shimla , showed a strong correlation between increasing built-up area and land surface temperature (Sunil Jaswal & Paritosh Thakur, 2023). Additional studies also show a negative correlation between LST and NDVI for Delhi, Kathmandu Valley, and Dhaka (Maharjan et al., 2021).

However, no consolidated research for such a correlation across various years in Delhi exists in the current academic world. This is the focus of my study, which is detailed in the remaining section of this paper.

Data

Data Source

The primary dataset used for this analysis is the Landsat 8 Collection 2, Tier 1 Level-2 data from the US Geological Survey (USGS), accessed through Google Earth Engine (GEE). The dataset provides surface reflection data in various spectral bands and thermal infrared data required to calculate land surface temperature. The region of interest, Delhi, was constrained using exact administrative boundaries, ensuring that all operations and calculations were restricted to this specific area. The time frame was kept constant, to the summer months of May to July. Optical bands had a resolution of 30m, but the thermal bands had a resolution of 100m, which were resampled to 30m.

Resampling was conducted for consistency, as by resampling thermal bands to 30 meters, all the analysis, like combining NDVI and LST, is done at the same spatial resolution, avoiding misalignments. It also aided in visualizing LST and NDVI and performing pixel-wise analysis that involves both LST and NDVI, such as the scatter plots shown in the analysis below. The administrative boundary for the region of Delhi was obtained from the FAO GAUL dataset in GEE, specifically *FAO/GAUL/2015/level1*. This provides simplified administrative boundaries globally. The Delhi boundary was extracted by filtering for the exact name "Delhi."

Preprocessing of Landsat 8 Data

Firstly, the raw digital numbers (DN) from the Landsat 8 Collection 2 images were converted to Surface Reflectance and Brightness Temperature values using the standard scaling factors.

- Optical Bands (SR_B2, SR_B3, SR_B4, SR_B5) were scaled using the equation (Digital Earth Africa, 2022):

$$\text{Scaled Value} = DN \times 0.0000275 - 0.2$$

- Thermal Bands (ST_B10) were scaled using the equation:

$$Radiance = DN \times 0.00341802 + 149.0$$

Then, a cloud mask was applied to ensure the accuracy of surface reflectance and temperature data. The QA_PIXEL band was used to mask out both clouds and cloud shadows. Specifically, Bits 3 and 5 in the QA_PIXEL band were used to detect cloud shadow and pixels, respectively. Only clear-sky pixels, where both flags are 0, were retained in the analysis.

Land Surface Emissivity

Emissivity is a key input for calculating LST and is derived using NDVI-based fractional vegetation cover (FV) through the following steps, using maximum and minimum NDVI values calculated across the region.

$$FV = \left(\frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \right)^2$$

Emissivity for each pixel was then calculated using (Sekertekin & Bonafoni, 2020):

$$EM = 0.986 + 0.004 \times FV$$

LST computation

LST was derived from the thermal infrared band using the following equation (Ridho, 2023):

$$LST = \frac{T_b}{1 + (0.00115 \times \frac{T_b}{1.438}) \times \ln(EM)} - 273.15$$

Where:

- T_b is the brightness temperature in Kelvin
- EM is the emissivity derived previously
- The result is in degrees Celsius, as the Kelvin-to-Celsius conversion is applied by subtracting 273.15.

Visualization and Mapping

A true color composite of the Landsat 8 data was generated using bands SR_B4 (red), SR_B3 (green), and SR_B2 (blue). The composite was used to visually verify the area of interest and the data's spatial integrity. Then, NDVI values were visualized on a color gradient from blue (low NDVI) to green (high NDVI), providing a spatial overview of vegetation density in the Delhi region. Lastly, the calculated LST was visualized using a color palette ranging from cooler

blue tones (lower temperatures) to warmer red tones (higher temperatures), with temperatures between 25°C and 50°C.

To analyze the LST data further or use it for additional modeling, the LST, NDVI, and emissivity values were sampled across the region at a spatial resolution of 30 meters and exported to a CSV file using the Google Earth Engine “*Export.table.toDrive*” function. The exported dataset included each sampled point's LST, NDVI, emissivity values, and geometries, which were then processed and presented graphically using Python for further analysis. Matplotlib, Pandas, and Sci-KitLearn libraries were used.

Results

The variance (R^2) from the best-fit line was calculated to quantify correlation. This can be seen in the table below.

The R^2 values obtained from the scatter plots between NDVI and Land Surface Temperature (LST) over the years showcase varying degrees of correlation. The values range from 0.1698 in 2015 to 0.382 in 2022. However, this showed that although there was variance from the mean value, there was a clear, negative correlation between Land Surface Temperature and Vegetation level.

Year	R^2 Value	Gradient
2014	0.2989	-14.28
2015	0.1698	-12.83
2016	0.1997	-11.26
2017	0.3032	-17.10
2018	0.2864	-17.24
2019	0.2843	-17.05
2020	0.3377	-11.55
2021	0.2739	-11.29
2022	0.3820	-15.19
2023	0.1971	-10.21
2024	0.3156	-15.17

The highest R^2 value of 0.382 in 2022, as seen in Fig 2, indicates that in this year, about 38.2% of the variation in LST can be explained by NDVI, representing the best fit among the years analyzed. This suggests that during 2022, vegetation had a more pronounced relationship

in predicting surface temperatures. Similarly, other relatively higher R^2 values are observed in 2020 (0.3377), 2024 (0.3156), and 2017 (0.3033), where the relationship between NDVI and LST is moderately strong. These results show that the variation in vegetation index is directly correlated with the observed LST values.

Furthermore, the table also shows the values of the gradient of the line of best fit for the various years. The average gradient of -13.92 indicates that, on average, a unit increase in NDVI corresponds to a 13.92 °C decrease in Land Surface Temperature (LST). More appropriately, for every 0.1 increase in NDVI, the LST decreases 1.392 °C. While the average gradient summarizes the relationship over time, the year-to-year variability is not evident. Some years showed a stronger relationship with steeper gradients like 2022 and 2023, while others exhibited weaker correlations with shallower slopes, like 2015 and 2016.

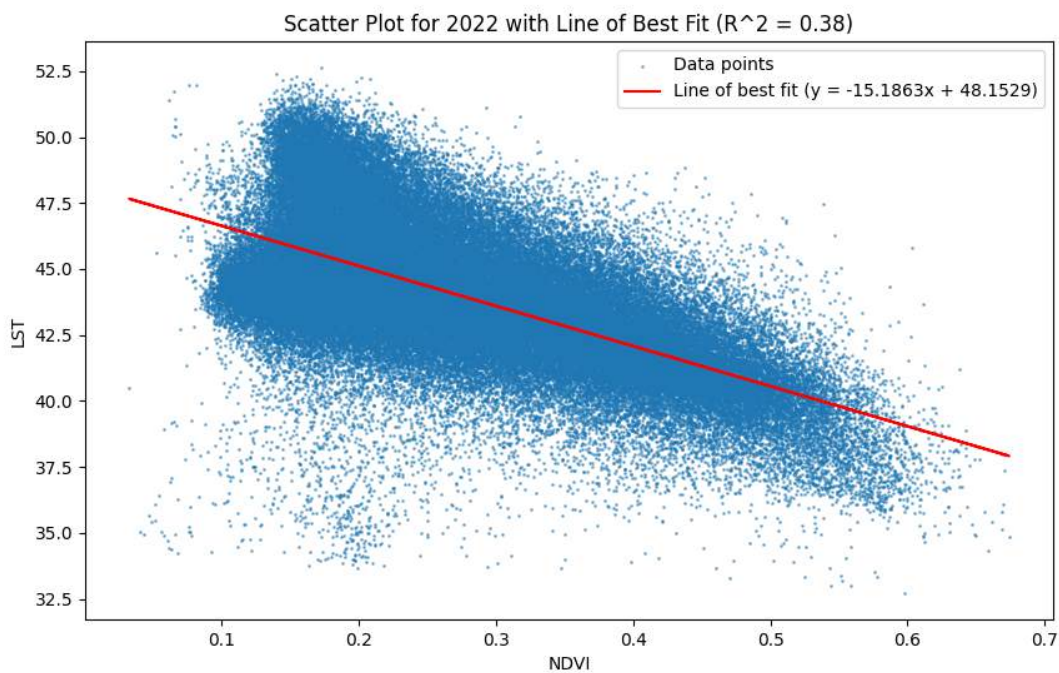


Figure 2. Scatter Plot of LST vs NDVI for 2022 (highest correlation)

On the other hand, years such as 2015 (0.1701), shown in Figure 3, 2023 (0.1971), and 2016 (0.1997) exhibit much lower R^2 values, indicating weaker correlations. In these years, less than 20% of the variance in LST can be explained by NDVI, suggesting that other factors might have had a larger influence on surface temperature. The weaker correlation implies that vegetation health and density, as captured by NDVI, had less importance in predicting surface temperature changes.

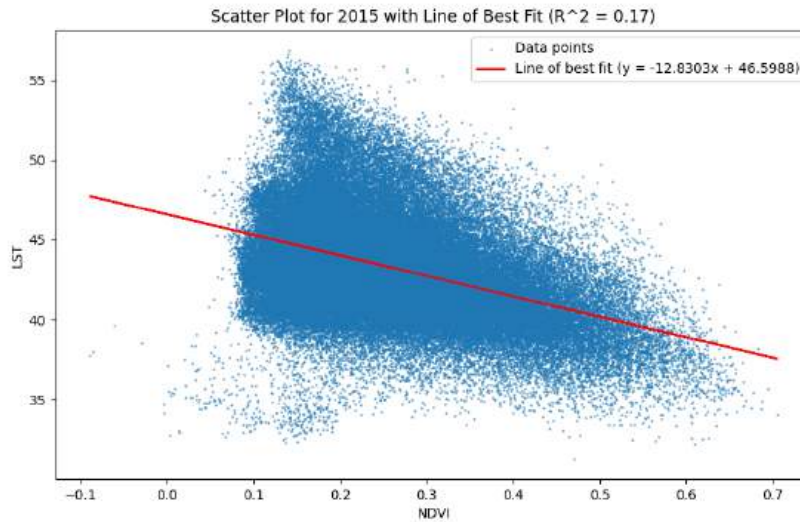


Figure 3. Scatter Plot for LST vs NDVI for 2015 (lowest correlation)

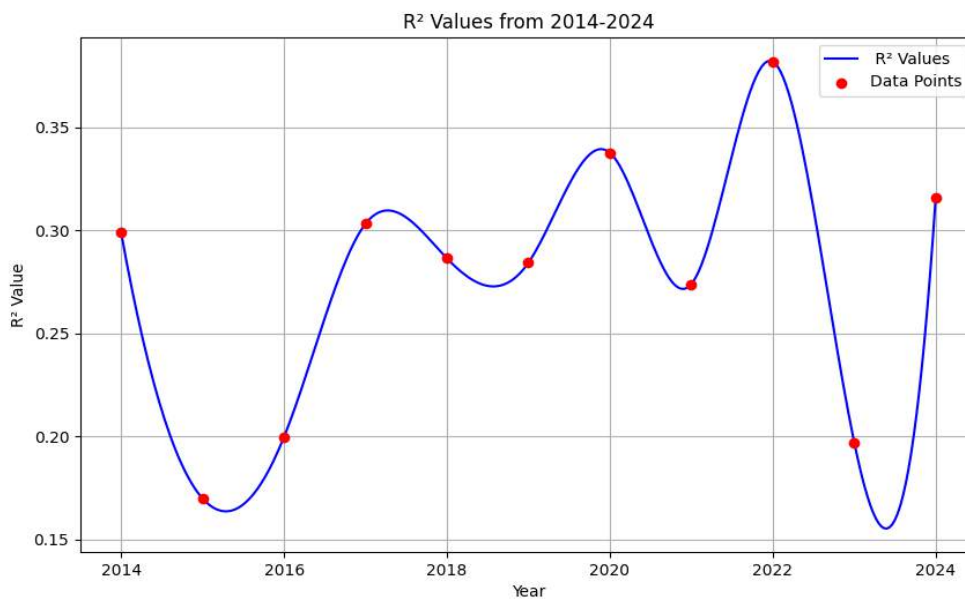


Figure 4. Smooth Curve Plot for R² values from 2014-2024

Discussion

The apparent weakness in the correlation may be due to the size of the datasets. Each annual dataset had roughly 90,000 data points, resulting in noisy values, which could affect the strength of the correlation. The year-to-year variability in both gradients and R² values indicates that the relationship between NDVI and LST is not constant. Some years exhibit a stronger influence of vegetation on surface temperature, while others show weaker correlations.

These fluctuations suggest that while vegetation likely affects LST, other factors alter LST beyond vegetation. Changes in ambient air temperature from car emissions release

pollutants like nitrogen oxides and particulate matter, which can absorb and retain heat in the atmosphere. Increasing rainfall and humidity can cause a cooling effect, as water directly lowers land surface temperature. Moreover, increased pollution levels could trap more heat close to the earth's surface due to the greenhouse effect. Winds can also impact LST in Delhi. The southwesterly monsoon winds have a cooling effect as they bring rain. Furthermore, cloud cover and buildings can cast shadows over the surface, which decreases the amount of sunlight falling on the surface. This would cause a decrease in the land surface temperature.

Limitations and Future Work

The data was very noisy, with over 90,000 data points in each year's dataset. This may have impacted the R^2 value and, thus, the overall interpretation of the relationship between LST and NDVI. Furthermore, there was also a limited temporal scope, as this analysis only examined data from specific months of the year, from May to July, which could limit the generalizability of the results. The resolution of LST and NDVI was also a limiting factor. Improvements in technology could allow for better sampling resolution and thus reduce errors, as small-scale vegetation may not be overlooked.

In the future, cities other than Delhi, such as Mumbai, Bangalore, Chennai, and Hyderabad, could also be studied to see if such a trend is evident in other Indian cities. Furthermore, such analysis could also be conducted in urban areas in cities in other countries, like Rio De Janeiro or Madrid, to understand correlations between LST and NDVI worldwide. The impact of other variables, such as pollution, humidity, and wind, could also be studied to provide a more holistic explanation of the factors affecting NDVI.

Conclusion

This study has successfully shown that there exists a negative relationship between land surface temperature and normalized difference vegetation index in Delhi. It can also be seen that on average, for every 0.1 unit increase in NDVI, LST decreases by 1.392 °C. However, vegetation is not the only factor affecting temperatures as felt by an observer on the ground.

Works Cited

- Auravant. (n.d.). *Vegetation indices and their interpretation: NDVI, GNDVI, MSAVI2, NDRE, and NDWI*. Auravant.
<https://www.auravant.com/en/articles/precision-agriculture/vegetation-indices-and-their-interpretation-ndvi-gndvi-msavi2-ndre-and-ndwi/>
- Beiser, V. (2019, August 20). *Feeling the Heat? Blame Concrete*. Time; Time.
<https://time.com/5655074/concrete-urban-heat/>
- Digital Earth Africa. (2022, February 14). *Landsat Collection 2 Level-2 Surface Reflectance — Digital Earth Africa 2021 documentation*. Docs.digitalearthafrika.org.
https://docs.digitalearthafrika.org/en/latest/data_specs/Landsat_C2_SR_specs.html
- Guha, S. (2021). A long-term monthly assessment of land surface temperature and normalized difference vegetation index using Landsat data. *Urbe. Revista Brasileira de Gestão Urbana*, 13. <https://doi.org/10.1590/2175-3369.013.e20200345>
- Guha, S., & Govil, H. (2020). Land surface temperature and normalized difference vegetation index relationship: a seasonal study on a tropical city. *SN Applied Sciences*, 2(10).
<https://doi.org/10.1007/s42452-020-03458-8>
- Guha, S., Govil, H., Dey, A., & Gill, N. (2018). Analytical study of land surface temperature with NDVI and NDBI using Landsat 8 OLI and TIRS data in Florence and Naples city, Italy. *European Journal of Remote Sensing*, 51(1), 667–678.
<https://doi.org/10.1080/22797254.2018.1474494>
- Jin, Y., Wang, H., Wang, Y., Fry, J., & Lenzen, M. (2021). Material footprints of Chinese megacities. *Resources, Conservation and Recycling*, 174, 105758.
<https://doi.org/10.1016/j.resconrec.2021.105758>
- Khan, R., Li, H., Basir, M., Chen, Y. L., Sajjad, M. M., Haq, I. U., Ullah, B., Arif, M., & Hassan, W. (2022). Monitoring land use land cover changes and its impacts on land surface temperature over Mardan and Charsadda Districts, Khyber Pakhtunkhwa (KP), Pakistan. *Environmental Monitoring and Assessment*, 194(6).
<https://doi.org/10.1007/s10661-022-10072-1>
- Liang, S., & Wang, J. (2020). Estimate of vegetation production of terrestrial ecosystem. *Advanced Remote Sensing*, 581–620. <https://doi.org/10.1016/b978-0-12-815826-5.00015-5>
- Lu, Y., Yue, W., & Huang, Y. (2021). Effects of Land Use on Land Surface Temperature: A Case Study of Wuhan, China. *International Journal of Environmental Research and Public Health*, 18(19), 9987–9987. <https://doi.org/10.3390/ijerph18199987>
- Maharjan, M., Aryal, A., Man Shakya, B., Talchabhadel, R., Thapa, B. R., & Kumar, S. (2021). Evaluation of Urban Heat Island (UHI) Using Satellite Images in Densely Populated Cities of South Asia. *Earth*, 2(1), 86–110. <https://doi.org/10.3390/earth2010006>
- Narayan, P., & Dhindaw, J. (2021, March 19). *Urban Greening for Cooling Cities*. WRI India Ross Center for Sustainable Cities | Helping Cities Make Big Ideas Happen.
<https://www.wricitiesindia.org/content/urban-greening-cooling-cities>

- National Geographic. (2022, May 20). *Urban Heat Island*. Education.nationalgeographic.org. <https://education.nationalgeographic.org/resource/urban-heat-island/>
- Pati, I. (2023, April 16). *Gurgaon: Why your plush new home is in a heat trap*. The Times of India; Times Of India. <https://timesofindia.indiatimes.com/city/gurgaon/gurgaon-why-your-plush-new-home-is-in-a-heat-trap/articleshow/99530682.cms>
- Reagan Gyimah, R., Kwang, C., Agyepong Antwi, R., Morgan Attua, E., Barimah Owusu, A., & Kofi Doe, E. (2023). Trading greens for heated surfaces: Land surface temperature and perceived health risk in Greater Accra Metropolitan Area, Ghana. *The Egyptian Journal of Remote Sensing and Space Sciences*, 26(4), 861–880. <https://doi.org/10.1016/j.ejrs.2023.09.004>
- Ridho, M. (2023, September 22). *Analyzing Land Surface Temperature (LST) with Landsat 8 Data in Google Earth Engine*. Medium; Medium. <https://medium.com/@ridhomuh002/analyzing-land-surface-temperature-lst-with-landsat-8-data-in-google-earth-engine-f4dd7ca28e70>
- Sekertekin, A., & Bonafoni, S. (2020). Land Surface Temperature Retrieval from Landsat 5, 7, and 8 over Rural Areas: Assessment of Different Retrieval Algorithms and Emissivity Models and Toolbox Implementation. *Remote Sensing*, 12(2), 294. <https://doi.org/10.3390/rs12020294>
- Singh, R., Grover, A., & Zhan, J. (2014). Inter-Seasonal Variations of Surface Temperature in the Urbanized Environment of Delhi Using Landsat Thermal Data. *Energies*, 7(3), 1811–1828. <https://doi.org/10.3390/en7031811>
- Sunil Jaswal, & Paritosh Thakur. (2023). Correlation between LST, NDVI and NDBI with reference to Urban Sprawling – A Case Study of Shimla city. *International Journal for Multidisciplinary Research*, 5(6). <https://doi.org/10.36948/ijfmr.2023.v05i06.10258>
- Team Cropin. (2021, December 17). *NDVI- Normalized Difference Vegetation Index*. Wwww.cropin.com. <https://www.cropin.com/blogs/ndvi-normalized-difference-vegetation-index>
- Testbook. (2023, October 27). *Urban Heat Island Effect - Meaning, Causes & Impacts for UPSC!* Testbook. <https://testbook.com/ias-preparation/urban-heat-island-effect>
- Ullah, W., Ahmad, K., Ullah, S., Tahir, A. A., Javed, M. F., Nazir, A., Abbasi, A. M., Aziz, M., & Mohamed, A. (2023). Analysis of the relationship among land surface temperature (LST), land use land cover (LULC), and normalized difference vegetation index (NDVI) with topographic elements in the lower Himalayan region. *Heliyon*, 9(2), e13322. <https://doi.org/10.1016/j.heliyon.2023.e13322>
- United Nations. (2018). *World Urbanization Prospects 2018*. <https://population.un.org/wup/Publications/Files/WUP2018-Highlights.pdf>
- United Nations. (2024). *World Population Prospects*. United Nations. <https://population.un.org/wpp/>
- Young, R. E. (2003). The Galileo probe: how it has changed our understanding of Jupiter. *New Astronomy Reviews*, 47(1), 1–51. [https://doi.org/10.1016/s1387-6473\(02\)00272-5](https://doi.org/10.1016/s1387-6473(02)00272-5)

Exploring the Relationship Between Sleep Disorders and Addiction to Alcohol or Nicotine By Gia Bhat-Roach

Abstract

Thousands of people use and abuse substances every day and this can have potentially detrimental consequences. In particular, substances of abuse can negatively affect sleep, including sleep duration and quality. This impact on sleep can make it extra hard for one with a sleep disorder to function well in society. In this paper, we assess the relationship between alcohol or nicotine abuse and 4 specific sleep disorders: REM Sleep Behavior Disorder (RBD), Obstructive Sleep Apnea (OSA), Narcolepsy (NT1/2), and Restless Leg Syndrome (RLS). Here, we review recent studies assessing the use and abuse of alcohol and nicotine and explore their relationship to the sleep disorders listed above. Throughout the paper, we establish the notion that substance abuse can worsen the symptoms of these disorders as well as add to the strenuous pre-existing effects that one with a sleep disorder may already experience. Sleep disorders are very unknown and underrepresented in research so they are often very difficult to treat, meaning people can struggle with symptoms for many years. The lifestyle choices of people with these disorders, such as drinking or smoking can contribute to the worsening of these disorders and their symptoms. This makes it all the more crucial that we understand the factors that contribute to their development or severity. Awareness about the harmful effects of substance abuse should be raised to the patients with these disorders in order to restore proper sleep quality and better the quality of life for all of these patients.

Keywords Sleep disorders, REM sleep, Gaba receptors, Nicotine, Alcohol, Substance abuse, dopamine

1. Introduction

Alcohol and nicotine are two well-researched psychoactive substances that are notorious for their negative health impacts on the human body. Reports from 2022 show that in the United States, 137.4 million people drank alcohol, 50.9 million people used tobacco products, and 23.5 million people vaped nicotine at the time.¹ These numbers alone depict the wide-reaching influence of substances on the general population. Sleep disorders, while underrepresented in research, are a major factor in many people's lives, with an estimated amount of 50-70 million Americans suffering from an ongoing sleep disorder.² The main objective of this literature review is to compare the effects of alcohol and nicotine on various sleep disorders and observe the way these substances interact with the body and brain to affect sleep (Table 1). Here, we discuss the following sleep disorders: REM sleep behavior disorder (RBD), Narcolepsy (NT1)+(NT2), Obstructive Sleep Apnea (OSA), and Restless Leg Syndrome (RLS), and dive into the cause of each of these sleep disorders, the effects of the disorder on a person, the risks associated with substance abuse, and what substance use translates to for someone with a preexisting sleep disorder.

This information is vital to know because it establishes a relationship between two prevalent areas in people's lives that are often poorly understood. In addition, understanding this research allows us to comprehend how substances specifically affect people with sleep disorders. This is pertinent to know because as one gets older the risks associated with these sleep disorders and/or the development of them become increasingly prominent. To prevent the development or exacerbation of these disorders, proper research must be done into the effects of these substances, and awareness must be brought to the subject. Establishing the connection between these disorders and substance abuse will allow researchers and scientists to counter the effects of these substances and/or raise awareness about substance abuse among people with these disorders; ultimately resulting in better health and better quality of life.

Methods

Search strategy

Studies were identified by searching key phrases on research sites such as pubmed and google scholar. The most common search words were “nicotine”, “alcohol” and “sleep disorders”. The publication dates of the articles are found to be anywhere between 1981 and 2024. All articles have been created by humans and have been reserved to the english language. Abstracts and sources have been looked over to verify merit. Full articles and documents have been examined if agreed to have proper merit.

Study selection

Multiple studies have been incorporated into this paper in effort to ensure unbiased and more accurate data. Some study types include, but are not limited to: Clinical trials, observational studies, case-control studies, and focus group studies. Both sexes are included in each study with the exception of individual studies. With the majority of studies having a wider focus on 18+ year olds (varying on disorder type and demographic).

2. Background on Substances and Sleep Disorders

2.1 Alcohol

Alcohol affects all parts of the brain but has particularly strong effects on the cerebral cortex, limbic system, thalamus, basal forebrain (including basal ganglia), and lastly the hypothalamus. The hypothalamus also is the main area in the brain that controls sleep.³ Alcohol is a sedative, meaning that it depresses the central nervous system because it enters the body, travels through the bloodstream to the brain, and subsequently slows the firing of neurons.^{4,5} This is because of an intake of GABA and, in sequence, an influx of chloride ions (Figure 1). The slowing of these neurons affect the occipital lobe (located in the cerebral cortex), the limbic system, frontal lobes, and the cerebellum. This means that, in turn, alcohol use directly impacts vision, coordination, and judgment.⁹ The thalamus, hypothalamus, and basal ganglia are also the

locations of GABAergic neurons that use melatonin to control sleep, meaning that when alcohol enters the brain and disrupts signaling in these areas, it disturbs the release of melatonin into one's brain and can lead to insomnia or the suppression of REM sleep.^{7,8}

Alcohol use disorder (AUD), also commonly known as alcohol abuse disorder, is a disorder where a person is unable to control the use/abuse of alcohol. In the first stages of AUD, one may experience drinking alone, thinking about alcohol a lot, and blacking out after the usage. In the next stage, one loses control of drinking, and it starts to affect their daily life. In the last stage, alcoholism has reached its chronic point. One only thinks about drinking and forgets about the things that they enjoyed previously. One is also very likely to be suffering from organ failure, heart disease, memory problems and is at a higher risk for cancer.⁹ These separately are unrelated to sleep but when considering the demographic and patient background of most people with sleep disorders this could potentially contribute to health risks. AUDs can also promote insomnia, change sleep architecture, lead to abnormalities in the circadian rhythm, and lead to a short sleeping period, all less than ideal when it comes to sleep disorders.¹⁰ There are, however, ways to stop alcohol dependency, such as therapy or certain prescribed medications.¹¹ AUD is a fairly common disorder, with 29.5 million people (ages 12 and above) suffering from AUD in 2022. This disorder can affect anyone with little to no gender differences. AUD can also affect any age range with no targeted age range or substantial difference (though this probably varies based on worldly location and age of legalized drinking).

2.2 Nicotine

Nicotine (unlike alcohol) is a stimulant rather than a depressant. This means that when nicotine enters the bloodstream, it speeds up the firing of neurons within the brain.¹² The speeding up of the central nervous system causes one to feel more awake and makes one feel like they have more energy.¹³ This is because nicotine binds to acetylcholine receptors. Acetylcholine receptors are excitatory, meaning they send off signals to other neurons.¹⁴ When nicotine binds to these receptors, it takes their place. This means that nicotine speeds up the firing rate of neurons throughout the brain, including in the hypothalamus.¹⁵ The hypothalamus is indirectly connected to the release of adrenaline, so the speed of these receptors stimulates the release of adrenaline.¹⁶ While these effects might seem great at first, they wear off quickly, and during the recession, one is likely to feel tired, grouchy, irritated, anxious, or depressed.¹⁷

Nicotine is a very addictive substance, usually self-administered via smoking. You can find nicotine in cigarettes, vapes, and all tobacco products. Tobacco use is very common and widespread. The Centers for Disease Control and Prevention released an article that estimated 28.3 million U.S. adults were actively using cigarettes in 2021. To further discuss demographics, nicotine dependence via cigarettes is usually found in people over the age of 50.¹⁸ When it comes to gender differences, there is no discriminating factor. The use of these smoking products leaves one at a high risk of developing nicotine dependence.¹⁹ There are also alternative ways to consume nicotine, such as chewing tobacco, pills, or patches. These are less harmful due to lower doses but still put one at risk for nicotine dependence.²⁰

2.3 REM Sleep Behavior Disorder

REM sleep behavior disorder (RBD) is a disorder characterized by a loss of muscle atonia during the REM sleep cycle. Sleep atonia or muscle atonia is the lack of muscle tone that occurs at night during the REM sleep cycle, when the neurotransmitters in the brain signal for lessened muscle tone within the body. This is a normal function of the human body. However, in cases like RBD, the sleep atonia is absent.²¹ This means at night, when the nervous system is supposed to be sending signals to the brain to stop muscle tone production, the muscle tone production actually continues for people with RBD.²² Because of this, people with RBD often act out their dreams in real time because there is no sleep atonia to prevent their muscles from moving, which means they are liable to talking, kicking, punching, arm flailing, thrashing, and more. These actions, if not just annoying, can be harmful to a patient and/or their bed partner.²³ RBD is often triggered during nightmares or action-packed dreams, so it is not uncommon for a person with RBD to fall out of bed or unintentionally harm their partner while still in their sleep state. Another defining characteristic of RBD is the ability to recall dreams well and for prolonged periods of time if woken up during the dream or an RBD episode, whereas a person without RBD usually forgets their dreams after they wake up or partway through the day.²⁴ Although there is no definite way to cure RBD, melatonin or clonazepam supplements can be taken to lessen the symptoms of RBD.²⁵

RBD seems to prevail in men more than women, with over 80% of recorded cases belonging to men, though this might be due to underdiagnosed cases in women.^{26,27,28} RBD cases start typically around age 50+ and happen in the late stages of adulthood. If RBD occurs earlier than the typical demographics, it is listed as early onset REM Sleep Behavior Disorder though this is not as common.²⁹ An additional demographic point is that people with idiopathic RBD are very likely to develop certain neurodegenerative diseases such as Parkinson's, Lewy body dementia, or multiple system atrophy, though the reason for this is currently unknown.³⁰

2.4 Obstructive Sleep Apnea

Obstructive sleep apnea (OSA) is a disorder in which certain muscles in the throat relax and block the airway, causing a person to suddenly wake up during their sleep. This can happen multiple times in one night.³¹ The amount of times one wakes up every hour during the night is known as their apnea hypopnea index (AHI). AHI is a scale used to measure the severity of sleep apnea disorders. In the AHI the greater the number of sleep interruptions per hour the more severe the sleep apnea is. Anything above 5/hour is considered apnea.⁵ These sleep awakenings are considered mild if their AHI is anywhere from 5-15 awakenings per hour. Anywhere from 15-30/hour is considered moderate, and anything above 30/hour is severe.³² The common symptoms of OSA are snoring, daytime sleepiness, and increased hypopneas.³³ Death, though uncommon, is also possible in people with untreated OSA and is more likely to affect older people and people with severe OSA.³⁴ Unlike RBD, obstructive sleep apnea can be caused by various different things, with the most common cause being obesity.³⁵ Some other factors that

can increase the risk of OSA are diabetes mellitus, hypertension, asthma, arthritis, hyperlipidemia, or (imperative to this paper) smoking and/or the use of alcohol.^{36,37} OSA has been seen as a male-predominant disorder in the past, though in current times, it has been found to be increasingly prevalent in women. The reason for OSA being historically underdiagnosed in women may be because of the differences in the way this disorder presents itself. For example, lower AHI in women may generate less cause for concern and can often be mistaken for another disorder, whereas in men, AHI is usually higher, so it's less likely to be mistaken.³⁸ The age range for patients with OSA is a bit unsteady but averaging about 50+ with it affecting older adults most commonly.^{39,40}

There is no real “cure” for OSA because it is a disorder that is based on anatomy; OSA is also not likely to go away permanently. There are, however, some treatments or lifestyle changes that a person with OSA can take to make their OSA more manageable. For example, for some people, nasal surgery might allow them more room in the windpipe and thus lessen their sleep apnea.⁴¹ If the apnea is caused by excessive weight, lifestyle changes and/or weight loss can be made in order to keep the symptoms at bay.⁴² The use of CPAP machines to hold the windpipe open is fairly common too.⁴³ However, there is no one solid treatment method for sleep apnea because it varies from case to case. The treatment of OSA is very critical because untreated OSA has negative effects including higher risk of cardiovascular disease, stroke, metabolic disease, daytime sleepiness (because of the hypopneas), and unintended accidents.⁴⁴ This stresses the need to treat OSA, and is especially relevant when considering that alcohol increases the threat of undiagnosed OSA.

2.5 Narcolepsy

Narcolepsy is an uncommon disorder with an estimated amount of only 0.87-1.21 percent of the world population experiencing narcolepsy type-1.⁴⁵ This disorder can occur at any time and within all age groups but it typically occurs at adolescence. It is found that the younger the age at the start of narcolepsy the more severe the condition will be in the future.⁴⁶ It also has no consistent gender difference.⁴⁷ The number of patients with narcolepsy is fairly small but this may not reflect the actual amounts of patients with narcolepsy because the symptoms of narcolepsy overlap with those of other disorders. Thus it is often misdiagnosed. Some of the disorders that narcolepsy gets mistaken for are depression/anxiety, schizophrenia, OSA, sleep deprivation, or general parasomnias.^{48,49} This could be a potential reason that the narcolepsy diagnoses are so few and far between. There are two different forms of narcolepsy: narcolepsy with cataplexy (NT1) and narcolepsy without cataplexy (NT2). Cataplexy (not to be mistaken with sleep atonia) is the lack of muscle tone during the daytime. Cataplexy is triggered by strong emotions such as laughter, sadness, etc. This is because people with cataplexy are found to lack the cells that produce orexin (also known as hypocretin) this means when they feel a strong emotion and don't have orexin it triggers the part of the brain that creates sleep atonia resulting in loss of muscle tone (most commonly in the face).³ The more common version of narcolepsy, NT1, is when cataplexy is a symptom. The lack of orexins to keep one awake causes a person

with narcolepsy to feel drowsy constantly, even when they have had quality sleep.⁵⁰ NT2 is different, however. Patients with NT2 usually retain somewhat normal levels of orexins, meaning they aren't affected by cataplexy. This is theorized to be because of only partial loss of orexin neurons and could potentially be the result of an outside/head injury, but this type of narcolepsy is significantly under-researched.

The common symptoms for someone with NT1 are abnormal REM sleep, hallucinations, increased amounts of hypopneas, and loss of muscle tone or sudden paralysis (cataplexy) -usually during wake hours-. Opposingly, lack of muscle tone during sleep hours (RBD) can also be a symptom of narcolepsy.^{51,52} The symptoms of NT2 are usually the same as NT1 with the exception of cataplexy.^{53,54} The presence of increased hypopneas as shown with the AHI, is a big factor in the diagnosis of OSA but also appears as a symptom of narcolepsy. When diagnosing patients with OSA doctors usually look for signs of cataplexy so as to not confuse it with narcolepsy.⁵⁵ There is currently no direct cause of narcolepsy. Occasionally it is passed through family genetics, but for the most part it occurs randomly with not much about the origin currently known.⁵⁶ There is no surefire way to cure narcolepsy, but there are certain medications that will help minimize the symptoms of narcolepsy, such as melatonin, antidepressants, and stimulants.⁵⁷

2.6 Restless Leg Syndrome

Restless leg syndrome (RLS) is a sensory-motor disorder that causes one to feel an uncomfortable sensation (usually in the lower legs) that gives one the urge to move.⁵⁸ Restless leg syndrome is triggered by standing still too long or attempting to rest which explains why restless leg syndrome is often worse at night or when the legs are still.⁵⁹ The types of sensations one with RLS might experience are itching, tingling, pricking, burning, electricity, or a pins and needles feeling.⁶⁰ The cause for these provoking sensations is still unknown. Most doctors hypothesize that RLS is a dopamine-related disorder, and there are various drug studies to back up this theory, but nothing is fully confirmed. This underscores the need for more research on this topic.⁵⁹ The demographics of RLS, however, are well known. RLS is a quite common disorder, it is found in 5-10 percent of adults in the USA and 2-4 percent of children. The disorder is more common in women than men and is usually more prevalent as the age gets higher.⁶¹

3. Alcohol and Sleep Disorders

Gamma-aminobutyric acid (GABA) is an amino acid that functions as the main inhibitory neurotransmitter.⁸ Its main function is to lessen activity in neurons. Because GABA is an inhibitory neurotransmitter, it regulates cell signaling. Most relevant here is GABAergic regulation of the locus coeruleus neurons.⁶² The locus coeruleus is a nucleus that connects to the brainstem and other nuclei. Multiple studies have shown that locus coeruleus neurons play a predominant role in the transition between slow wave sleep and REM sleep, as well as playing a significant role in the initiation of the sleep process.^{63,64} This, in turn, means that GABA plays a big role in the regulation of sleep.⁶⁵ When alcohol enters the body it travels into the bloodstream

and circulates to the brain. There, the alcohol also binds to the GABA receptors and serves a similar function which is to depress neuron signaling (Figure 1).⁶⁶ When the GABA leaves the presynaptic terminal, it enters the postsynaptic terminal via the GABA receptors, specifically the GABAA receptor. The GABAA receptor allows negatively charged chloride ions to pass through.⁶² These Cl⁻ ions modulate the neuron network signaling and make the occurrence of an action potential less frequent.⁶⁷ In simple terms, this means that the sedative effects that high doses of alcohol has on the body combined with the excess GABA has lead scientists to reasonably hypothesis that alcohol leads to the suppression of rem sleep.⁶⁸

3.1 REM Sleep Suppression, RBD, and Alcohol

Proper REM sleep is imperative for memory, emotional processing, and healthy brain development.⁶⁹ Therefore, the suppression of REM sleep has many negative effects on the brain. Lack of proper REM sleep causes one to wake up feeling as if they have hardly slept at all. Alcohol interferes with REM sleep and also makes one prone to nightmares, vivid dreams, and restlessness during sleep. More frequent nightmares or vivid dreams can worsen the severity of RBD and worsen the risk of potentially harming oneself or one's partner.⁷⁰ In people without RBD, alcohol can cause RBD-like symptoms. Contrastingly, people who are going through alcohol withdrawal may also develop RBD-like symptoms. While alcohol has not yet been scientifically proven to directly cause RBD, there is enough strong evidence to support that there is a connection between the two. For example, multiple studies have found that people with pRBD are more likely to be heavy drinkers.^{70,71} One study in specific compared light, moderate, and heavy drinkers as well as non-drinkers, and assessed their likelihood of developing pRBD. Results found that moderates and heavy drinkers have a significantly higher likelihood of developing pRBD while there was little to no change in light drinkers and non-drinkers. The study states, "Compared with nondrinkers, current drinkers had a 23% higher likelihood of having pRBD".⁷⁰ In short the consumption of alcohol causes side effects that are known to trigger RBD and make it more harmful to one's self or partner.

3.2 Alcohol and Obstructive Sleep Apnea

As described above, alcohol depresses the central nervous system by intensifying the effects of GABA. As a result of the central nervous system being depressed, heart rate, breathing, and brain activity are all slowed, subsequently leading to the relaxation of muscles all around the body, including the muscles around the back of the throat, including the genioglossus muscle.⁷² Because of the lack of muscle tone in the genioglossus, there is a rise in the likelihood of an upper airway collapse, as well as an increase in the risk of tissue blocking the airway. Both of the previous can lead to OSA-like symptoms or worsen pre-existing OSA.^{73,74} Along with the changes in dopamine and relaxation that alcohol abuse brings, alcohol also leads to changes in the blood vessels of the nose. This is because substances like beer, wine, and ciders are very high in histamine.^{75,76} Histamine is a chemical the body releases as an allergic response, and can cause sneezing and congestion in the nose. This is bad for a person with OSA because it blocks the

airway and causes breathing resistance, as well as snoring and mouth breathing. All of these factors lead to an increased risk of developing obstructive sleep apnea because they limit the amount of air one can breathe and make one more prone to airway collapse or particle obstruction.⁷⁷ In fact, studies have shown that drinking alcohol increases your likelihood of developing obstructive sleep apnea by 25%.⁷⁸ In a separate study, 11,859 patients were questioned for age, sex, BMI, high blood pressure, and other standard background questions, as well as their relationship to alcohol. This data was then used to observe the connection between alcohol abuse disorders and the likelihood of OSA. The study found that patients who had a pre-existing alcohol use disorder were more likely to be at risk of OSA than those who did not (aOR 2.45, 95% CI 2.04-2.95).⁷⁹ To summarize alcohol increases the risk of airway obstructions and congestion which can cause OSA. Alcohol has also been proven to cause OSA.

3.2.1 Hypoventilation and the threshold period

Another, more imperative mechanism is that the consumption of alcohol increases the threshold period from when one's breathing stops to when they gasp for air again. In a person with OSA this threshold period is normally 20-30 seconds and up to five times an hour for adults, or one time per hour for children.²⁰ But because alcohol causes breathing to slow down, it lengthens this threshold period. This may seem like very redundant information but the potential risk factors are highlighted when a particle obstructs the airpipe of someone with OSA. This is because alcohol slows down one's reaction time, and when a person with OSA stops breathing because of an obstruction in their airpipe (as they commonly do) the body is often unable to react and gasp for air fast enough, leading to a drop in one's blood oxygen level. This sudden drop is known as hypoventilation. Hypoventilation causes one's carbon dioxide levels to build up, and when this extra CO₂ mixes with water, it forms carbonic acid.⁸⁰ While carbonic acid is not inherently bad, in cases of people with OSA, it can be potentially lethal. This is because the body is not gasping for air fast enough, so the built-up carbonic acid is not leaving the body. This poses a risk for respiratory acidosis, which is a condition that happens when the body cannot expel carbon dioxide and instead causes the bodily fluids (including the blood) to become acidic. This built-up carbonic acid primarily affects oxygen by lowering oxygen levels in the blood (and by association, the respiratory system); hence the name respiratory acidosis.⁸⁰ If one can't breathe quickly enough, these low levels of oxygen may result in death. In short alcohol increases the threshold period posing the risk of death as well as the risk of developing respiratory acidosis.

3.2.2 Alcohol and Central Sleep Apnea

Central sleep apnea (CSA) is a less common sleep disorder that is very closely related to OSA. The main differentiating factor between the two disorders is that OSA is generally caused by a particle or obstruction physically blocking one's airway, while CSA is caused by the brain not sending proper signals to the muscles that control breathing, such as the diaphragm and inspiratory rib cage muscles.⁸¹ This means patients with CSA and alcohol abuse will still

experience longer threshold periods and the risk of hypoventilation, but the root of the problem stems from alcohol slowing the central nervous system, causing reaction time to slow. CSA, while sharing most of the underlying symptoms of OSA, is still important to mention as a unit in order to understand the difference in where these disorders stem from and the different way alcohol affects both.

3.3 Alcohol and Narcolepsy

While alcohol abuse is not known to cause narcolepsy, it still has many effects on narcolepsy and its symptoms. We have already established that alcohol has very soporific effects so it's no surprise that when the sedative effects of alcohol mix with the already demanding symptoms of narcolepsy, the combination creates extreme exhaustion. This is due to the fact that alcohol further depresses the central nervous system, leading to what is known as "alcohol-induced narcolepsy".⁶ Symptoms of alcohol-induced narcolepsy include uneven sleep, waking up during the night, falling asleep while doing tasks, and cataplexy.⁸³ Because the central nervous system is depressed, the production of GABA increases (similarly to RBD), which causes a person to fall asleep faster than normal. This in turn causes one to spend more time in slow-wave sleep and less time in REM sleep, resulting in frequent sleep interruptions and ultimately worse sleep quality. This poor quality of sleep at night can worsen narcolepsy daytime symptoms. With this, vivid dreams, paralysis, and hallucinations are also possible.⁸⁴ While alcohol and narcolepsy are not lethal on their own, alcohol combined with narcolepsy medication can make the medication less effective or in some cases have life-threatening side effects (varying based on medication).⁸⁵ In fact, the FDA has issued a warning against the use of alcohol and any drugs that suppress the central nervous system, such as Xyrem (a treatment for narcolepsy with cataplexy). This is because the combination increases the chance of depressed breathing, loss of consciousness, comas, and in some cases death.⁸⁶ The depressed breathing that Xyrem and alcohol cause is also a big risk factor for people with OSA because it can lead to respiratory depression. To recap, when the effects of alcohol are mixed with the exhaust of narcolepsy it makes it very hard for the patient to fight off sleep, the patient also gets lower quality sleep because of alcohol's effects making it hard to stay awake and a poor quality sleep when one does sleep.

3.4 Alcohol and Restless Leg Syndrome

When alcohol exits the body of someone with alcohol use disorder the nervous system becomes hyperactive.⁸⁷ This is because when a heavy drinker consumes alcohol daily, the brain becomes used to the lessened stimulation. This means that when the alcohol leaves the body, the body makes an attempt to readjust to the regular amount of stimulation and, in turn, causes hyperactivity in the nervous system.⁸⁸ This hyperactivity can lead to alcohol shakes or tremors in someone who already deals with RLS. Cold turkey, otherwise known as the sudden withdrawal from the use of a substance, can also cause alcohol shakes and tremors, which can worsen RLS and lower the quality of REM sleep because the body feels constantly stimulated unless the legs

are moving.⁸⁹ Alcohol in the body generally intensifies the uncomfortable sensations that one feels and makes it harder for someone with RLS or any sleep disorder to get a good quality sleep without it affecting their other symptoms.

4. Nicotine and Sleep Disorders

4.1 Nicotine and REM sleep

Nicotine is a stimulant, which means that it increases alertness and subsequently makes it harder to fall asleep.⁹⁰ When someone uses nicotine before sleeping the nicotine stimulates certain neurotransmitters such as dopamine and creates an influx. The neurotransmitter dopamine is a precursor to the synthesis of norepinephrine (also called noradrenaline) which is a neurotransmitter that controls arousal functions, attention, and cognitive functions.⁹¹ The production of norepinephrine is high in the morning, and as the night progresses, the production slows, meaning there are very low levels during NREM and REM sleep.⁹² Nonetheless, when nicotine stimulates the production of dopamine at night, it also increases the levels of norepinephrine, causing one to stay awake or have trouble falling asleep. In addition to dopamine, nicotine also stimulates the neurotransmitter serotonin. Serotonin affects wakefulness and sleep onset and is the main contributor to preventing REM sleep.^{93,94} It functions similarly to norepinephrine in that its overproduction prevents sleep. When both of these neurotransmitters are stimulated together it causes the suppression of REM sleep. REM sleep deprivation is harmful in many ways, predominantly affecting mood, memory processing, and brain development.⁹⁵

4.1.1 Nicotine and REM Sleep Behavior Disorder

As mentioned previously, the origin of RBD is significantly understudied. Surveys show that the cause of 55% of RBD cases are still unknown.^{96,97} While there are very few studies showing that nicotine is the origin of RBD, there have been a few studies showing that patients with RBD are more likely to be smokers. One of these papers uses statistics from various polysomnograms (PSG) that show smoking is more common in people with RBD.⁹⁸ Another study highlights the work of the international REM Sleep Behavior Disorder Study group in a study with 347 controls. The outcome was that more of the RBD patients had a history of smoking than any of the non-RBD 347 controls.⁹⁹ The debate on whether or not this information is significant varies from study to study. Some studies have found that the evidence of smoking being connected to RBD is too insignificant to pose a real threat. Overall this is a topic that scientists need to do further research on in order to draw any meaningful conclusions.

4.2 Nicotine and Obstructive Sleep Apnea: Sleep Architecture and Neuromuscular Functions

Sleep architecture is the normal structural organization of sleep, including NREM sleep stages 1, 2, 3, and 4 and REM sleep.¹⁰⁰ Smoking and nicotine abuse can change the body's sleep

architecture as well as neuromuscular functions.¹⁰¹ When nicotine enters the body, it stimulates the release of dopamine (as described in the previous paragraphs) which causes an increase in sleep latency. Sleep latency is the amount of time one takes to fall asleep. The imbalance of dopamine that nicotine causes can also lead to increased sleep fragmentation. Sleep fragmentation is exactly what it sounds like; it is when a person experiences fragmented sleep due to hypopneas or awakenings during the night.¹⁰² Sleep fragmentation is thought to contribute to the worsening of OSA.¹⁰³ This is because both sleep latency and sleep fragmentation lead to poor sleep quality, poor sleep quality means that the muscles are doing less protein synthesis, and in turn, this makes the muscles unable to function and worsens dilator muscle activity.¹⁰⁴ A dilator muscle is any muscle that expands or widens a body part,^{105,106} in this case, the throat muscles. It is thought that because the activity of these muscles is being slowed, these muscles are not holding the airway open as much, and this can lead to an increase in airway collapses. While this may not affect all people with OSA, when you consider the demographics of patients with OSA -usually older and have an impaired upper airway - it is very likely that they use their dilator muscles as their main means to keep their airway open, meaning they are very susceptible to coughing, choking, or even death if their airway collapses. In short, nicotine increases sleep latency and causes sleep fragmentation. These both worsen dilator muscle activity and make the airway more likely to collapse.

4.2.1 OSA and the nicotine rebound effect

The rebound effect is when symptoms of a drug or medication are absent during usage, but as one withdraws from the substance, the symptoms come back. This is seen in OSA because nicotine (if used over an extended period of time) actually decreases the amount of dopamine in one's body. The rebound effect kicks in during withdrawal from nicotine use, and the body creates more dopamine in response. This extra dopamine is used in the same manner as described previously and can worsen OSA.¹⁰⁷ The rebound effect can also make recovering people more inclined to smoke again to avoid dealing with the same symptoms they were trying to alleviate in the first place.¹⁰⁸

4.2.2. Inflammation and OSA

Because nicotine is a stimulant, when it enters the body it stimulates certain white blood cells such as leukocytes, neutrophils, lymphocytes, and monocytes.¹⁰⁹ Neutrophils, specific to our case, are the white blood cells that help fight off infection.¹¹⁰ When nicotine enters the body, neutrophils are stimulated by nicotine and create cytokines.¹¹¹ Cytokines are proteins that control the progression of immune system cells and blood system cells. They create these cells to fight off infections and protect the body from harmful pathogens. There are two kinds of cytokines: anti-inflammatory and pro-inflammatory. In the case of OSA, pro-inflammatory cytokines cause inflammation and swelling around the body.¹¹² Inflammation causes the uvula, nose, and throat to swell, which reduces the amount of space in the airway and makes one vulnerable to choking when particle obstruction occurs and/or causes disruptions in the airway.¹¹³

4.3 RLS and Nicotine:

RLS is hypothesized by medical professionals to be a dopamine-related disorder. Dopamine is used in the basal ganglia to create the purposeful movement of the body by controlling movement via complex actions on striatal neurons. Striatal neurons are clusters of nuclei that are in charge of decision-making functions such as motor control.¹¹⁴ These striatal neurons control the direct and indirect motor pathways within the basal ganglia. This means they affect the result of the striato-pallidal complex to the thalamus and then from the thalamus to the brainstem.^{115,116} The combination of all these parts is how dopamine usually is used to create voluntary movement. However, when nicotine enters the body, it stimulates central nicotinic acetylcholine receptors (nAChRs) and initiates the release of various neurotransmitters within the brain, including dopamine.¹¹⁷ Because there is a sudden influx of dopamine it creates an imbalance within the brain. While there is no direct linkage between nicotine and the cause of RLS, it is known that excessive dopamine can cause dyskinesia (a similar disorder to RLS), and people with RLS may experience worsened symptoms and/or dyskinetic movements. More research is yet to be done to prove the connection between the two but the involvement of dopamine in both RLS and dyskinesia might prove that there is a connection between RLS and nicotine.

5. Conclusion

Overall, alcohol and nicotine play a huge role in the functioning of the human body and central nervous system. Since alcohol is a depressant and nicotine is a stimulant, they affect the body and brain in very different ways, but both can have effects on sleep disorders. Here, we have explored the problematic effects that these substances have on a patient who is already struggling with these disorders. We analyzed the effects of alcohol and nicotine on 4 main sleep disorders (Table 1). In short, when alcohol enters the body, it leads to the suppression of REM sleep. This is pertinent to RBD because when someone with RBD lacks REM sleep, they are more prone to nightmares or vivid dreams. Because alcohol is a depressant, it also increases the signaling of the neurotransmitter GABA that controls sleep. This means people with narcolepsy experience extreme exhaustion on top of their usual symptoms. Alcohol being a depressant is also a huge factor in OSA as alcohol causes one to relax physically. This can lead to the expansion of muscles within the throat and lead to exaggerated symptoms of OSA. Nicotine can also lead to a worse AHI. In people with RLS, both alcohol and nicotine use and withdrawal can lead to shakes and tremors. .

In conclusion, there is evidence that alcohol, nicotine, and the general abuse of substances have a net negative impact on a person with a sleep disorder. The use of these substances is proven to worsen someone's REM sleep cycle and lead to a worse quality of sleep meaning, less energy during the day and subsequently lower quality of life. Knowing the connection between these substances of abuse and sleep disorders is important because thousands of people suffer from sleep disorders and as one gets older, the risks of developing one of these sleep disorders increases significantly. Research on the connection between these

common substances of abuse and these disorders can provide knowledge that will improve the quality of life for people who currently have a sleep disorder. A better understanding of these relationships can also prevent worsening of sleep disorders via awareness of the likelihood of substances of abuse to worsen these disorders.

Works Cited

- (1) Highlights for the 2022 National Survey on Drug Use and Health.
- (2) Institute of Medicine (US) Committee on Sleep Medicine and Research. *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*; Colten, H. R., Altevogt, B. M., Eds.; The National Academies Collection: Reports funded by National Institutes of Health; National Academies Press (US): Washington (DC), 2006.
- (3) Huff, T.; Mahabadi, N.; Tadi, P. Neuroanatomy, Visual Cortex. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2024.
- (4) Mukherjee, S. Alcoholism and Its Effects on the Central Nervous System. *Curr. Neurovasc. Res.* **2013**, *10* (3), 256–262. <https://doi.org/10.2174/15672026113109990004>.
- (5) Oscar-Berman, M.; Marinkovic, K. Alcoholism and the Brain: An Overview. *Alcohol Res. Health* **2003**, *27* (2), 125–133.
- (6) Oscar-Berman, M.; Marinkovic, K. Alcoholism and the Brain: An Overview. *Alcohol Res. Health* **2003**, *27* (2), 125–133.
- (7) Chakravorty, S.; Chaudhary, N. S.; Brower, K. J. Alcohol Dependence and Its Relationship with Insomnia and Other Sleep Disorders. *Alcohol. Clin. Exp. Res.* **2016**, *40* (11), 2271–2282. <https://doi.org/10.1111/acer.13217>.
- (8) Allen, M. J.; Sabir, S.; Sharma, S. GABA Receptor. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2023.
- (9) Ellis, R. R. *Alcohol Use Disorder*. WebMD. <https://www.webmd.com/mental-health/addiction/what-is-alcohol-abuse> (accessed 2024-01-23).
- (10) He, S.; Hasler, B. P.; Chakravorty, S. Alcohol and Sleep-Related Problems. *Curr. Opin. Psychol.* **2019**, *30*, 117–122. <https://doi.org/10.1016/j.copsyc.2019.03.007>.
- (11) Stokłosa, I.; Więckiewicz, G.; Stokłosa, M.; Piegza, M.; Pudło, R.; Gorczyca, P. Medications for the Treatment of Alcohol Dependence—Current State of Knowledge and Future Perspectives from a Public Health Perspective. *Int. J. Environ. Res. Public Health* **2023**, *20* (3), 1870. <https://doi.org/10.3390/ijerph20031870>.
- (12) Schilström, B.; Rawal, N.; Mamedi-Engvall, M.; Nomikos, G. G.; Svensson, T. H. Dual Effects of Nicotine on Dopamine Neurons Mediated by Different Nicotinic Receptor Subtypes. *Int. J. Neuropsychopharmacol.* **2003**, *6* (1), 1–11. <https://doi.org/10.1017/S1461145702003188>.
- (13) Hendler, R. A.; Ramchandani, V. A.; Gilman, J.; Hommer, D. W. Stimulant and Sedative Effects of Alcohol. *Curr. Top. Behav. Neurosci.* **2013**, *13*, 489–509. https://doi.org/10.1007/7854_2011_135.
- (14) Sam, C.; Bordoni, B. Physiology, Acetylcholine. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2024.
- (15) Calarco, C. A.; Picciotto, M. R. Nicotinic Acetylcholine Receptor Signaling in the Hypothalamus: Mechanisms Related to Nicotine's Effects on Food Intake. *Nicotine Tob. Res.* **2019**, *22* (2), 152–163. <https://doi.org/10.1093/ntr/ntz010>.

- (16) Smith, S. M.; Vale, W. W. The Role of the Hypothalamic-Pituitary-Adrenal Axis in Neuroendocrine Responses to Stress. *Dialogues Clin. Neurosci.* **2006**, *8* (4), 383–395.
- (17) McLaughlin, I.; Dani, J. A.; De Biasi, M. Nicotine Withdrawal. *Curr. Top. Behav. Neurosci.* **2015**, *24*, 99–123. https://doi.org/10.1007/978-3-319-13482-6_4.
- (18) Han, B.; Einstein, E. B.; Compton, W. M. Patterns and Characteristics of Nicotine Dependence Among Adults With Cigarette Use in the US, 2006-2019. *JAMA Netw. Open* **2023**, *6* (6), e2319602. <https://doi.org/10.1001/jamanetworkopen.2023.19602>.
- (19) Benowitz, N. L.; Hukkanen, J.; Jacob, P. Nicotine Chemistry, Metabolism, Kinetics and Biomarkers. *Handb. Exp. Pharmacol.* **2009**, No. 192, 29–60. https://doi.org/10.1007/978-3-540-69248-5_2.
- (20) Wadgave, U.; Nagesh, L. Nicotine Replacement Therapy: An Overview. *Int. J. Health Sci.* **2016**, *10* (3), 425–435.
- (21) Brown, R. E.; Basheer, R.; McKenna, J. T.; Strecker, R. E.; McCarley, R. W. Control of Sleep and Wakefulness. *Physiol. Rev.* **2012**, *92* (3), 1087–1187. <https://doi.org/10.1152/physrev.00032.2011>.
- (22) Figorilli, M.; Lanza, G.; Congiu, P.; Lecca, R.; Casaglia, E.; Mogavero, M. P.; Puligheddu, M.; Ferri, R. Neurophysiological Aspects of REM Sleep Behavior Disorder (RBD): A Narrative Review. *Brain Sci.* **2021**, *11* (12), 1588. <https://doi.org/10.3390/brainsci11121588>.
- (23) Hu, M. T. REM Sleep Behavior Disorder (RBD). *Neurobiol. Dis.* **2020**, *143*, 104996. <https://doi.org/10.1016/j.nbd.2020.104996>.
- (24) Boeve, B. F. REM Sleep Behavior Disorder: Updated Review of the Core Features, the RBD-Neurodegenerative Disease Association, Evolving Concepts, Controversies, and Future Directions. *Ann. N. Y. Acad. Sci.* **2010**, *1184*, 15–54. <https://doi.org/10.1111/j.1749-6632.2009.05115.x>.
- (25) McGrane, I. R.; Leung, J. G.; St Louis, E. K.; Boeve, B. F. Melatonin Therapy for REM Sleep Behavior Disorder: A Critical Review of Evidence. *Sleep Med.* **2015**, *16* (1), 19–26. <https://doi.org/10.1016/j.sleep.2014.09.011>.
- (26) Okura, M.; Taniguchi, M.; Sugita, H.; Ohi, M.; Tachibana, N. [Demographic characteristics of RBD patients at a sleep center--with special emphasis on neurodegenerative diseases as the background condition]. *Brain Nerve Shinkei Kenkyu No Shinpo* **2007**, *59* (11), 1265–1271.
- (27) Wing, Y. K.; Lam, S. P.; Li, S. X.; Yu, M. W. M.; Fong, S. Y. Y.; Tsoh, J. M. Y.; Ho, C. K. W.; Lam, V. K. H. REM Sleep Behaviour Disorder in Hong Kong Chinese: Clinical Outcome and Gender Comparison. *J. Neurol. Neurosurg. Psychiatry* **2008**, *79* (12), 1415–1416. <https://doi.org/10.1136/jnnp.2008.155374>.
- (28) Bodkin, C. L.; Schenck, C. H. Rapid Eye Movement Sleep Behavior Disorder in Women: Relevance to General and Specialty Medical Practice. *J. Womens Health* **2009**, *18* (12), 1955–1963. <https://doi.org/10.1089/jwh.2008.1348>.
- (29) Ju, Y.-E. S. Rapid Eye Movement Sleep Behavior Disorder in Adults Younger than 50

- Years of Age. *Sleep Med.* **2013**, *14* (8), 768–774.
<https://doi.org/10.1016/j.sleep.2012.09.026>.
- (30) Dauvilliers, Y.; Schenck, C. H.; Postuma, R. B.; Iranzo, A.; Luppi, P.-H.; Plazzi, G.; Montplaisir, J.; Boeve, B. REM Sleep Behaviour Disorder. *Nat. Rev. Dis. Primer* **2018**, *4* (1), 1–16. <https://doi.org/10.1038/s41572-018-0016-5>.
- (31) Prisant, L. M.; Dillard, T. A.; Blanchard, A. R. Obstructive Sleep Apnea Syndrome. *J. Clin. Hypertens. Greenwich Conn* **2006**, *8* (10), 746–750.
<https://doi.org/10.1111/j.1524-6175.2006.888139.x>.
- (32) Goyal, M.; Johnson, J. Obstructive Sleep Apnea Diagnosis and Management. *Mo. Med.* **2017**, *114* (2), 120–124.
- (33) Slowik, J. M.; Sankari, A.; Collen, J. F. Obstructive Sleep Apnea. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2024.
- (34) Marshall, N. S.; Wong, K. K. H.; Liu, P. Y.; Cullen, S. R. J.; Knuiaman, M. W.; Grunstein, R. R. Sleep Apnea as an Independent Risk Factor for All-Cause Mortality: The Busselton Health Study. *Sleep* **2008**, *31* (8), 1079–1085.
- (35) Tuomilehto, H.; Seppä, J.; Uusitupa, M. Obesity and Obstructive Sleep Apnea--Clinical Significance of Weight Loss. *Sleep Med. Rev.* **2013**, *17* (5), 321–329.
<https://doi.org/10.1016/j.smr.2012.08.002>.
- (36) Mitra, A. K.; Bhuiyan, A. R.; Jones, E. A. Association and Risk Factors for Obstructive Sleep Apnea and Cardiovascular Diseases: A Systematic Review. *Diseases* **2021**, *9* (4), 88. <https://doi.org/10.3390/diseases9040088>.
- (37) Thompson, C.; Legault, J.; Moullec, G.; Baltzan, M.; Cross, N.; Dang-Vu, T. T.; Martineau-Dussault, M.-È.; Hanly, P.; Ayas, N.; Lorrain, D.; Einstein, G.; Carrier, J.; Gosselin, N. A Portrait of Obstructive Sleep Apnea Risk Factors in 27,210 Middle-Aged and Older Adults in the Canadian Longitudinal Study on Aging. *Sci. Rep.* **2022**, *12* (1), 5127. <https://doi.org/10.1038/s41598-022-08164-6>.
- (38) Geer, J. H.; Hilbert, J. Gender Issues in Obstructive Sleep Apnea. *Yale J. Biol. Med.* **2021**, *94* (3), 487–496.
- (39) Ghavami, T.; Kazeminia, M.; Ahmadi, N.; Rajati, F. Global Prevalence of Obstructive Sleep Apnea in the Elderly and Related Factors: A Systematic Review and Meta-Analysis Study. *J. Perianesthesia Nurs. Off. J. Am. Soc. PeriAnesthesia Nurses* **2023**, *38* (6), 865–875. <https://doi.org/10.1016/j.jopan.2023.01.018>.
- (40) Punjabi, N. M. The Epidemiology of Adult Obstructive Sleep Apnea. *Proc. Am. Thorac. Soc.* **2008**, *5* (2), 136–143. <https://doi.org/10.1513/pats.200709-155MG>.
- (41) Pavwoski, P.; Shelgikar, A. V. Treatment Options for Obstructive Sleep Apnea. *Neurol. Clin. Pract.* **2017**, *7* (1), 77–85. <https://doi.org/10.1212/CPJ.0000000000000320>.
- (42) Treatment of Obstructive Sleep Apnea. In *InformedHealth.org [Internet]*; Institute for Quality and Efficiency in Health Care (IQWiG), 2019.
- (43) *CPAP - CPAP | NHLBI, NIH*. <https://www.nhlbi.nih.gov/health/cpap> (accessed 2024-03-29).

- (44) Knauert, M.; Naik, S.; Gillespie, M. B.; Kryger, M. Clinical Consequences and Economic Costs of Untreated Obstructive Sleep Apnea Syndrome. *World J. Otorhinolaryngol. - Head Neck Surg.* **2015**, *1* (1), 17–27. <https://doi.org/10.1016/j.wjorl.2015.08.001>.
- (45) Spruyt, K. Narcolepsy Presentation in Diverse Populations: An Update. *Curr. Sleep Med. Rep.* **2020**, *6* (4), 239–250. <https://doi.org/10.1007/s40675-020-00195-7>.
- (46) Dauvilliers, Y.; Montplaisir, J.; Molinari, N.; Carlander, B.; Ondze, B.; Besset, A.; Billiard, M. Age at Onset of Narcolepsy in Two Large Populations of Patients in France and Quebec. *Neurology* **2001**, *57* (11), 2029–2033. <https://doi.org/10.1212/wnl.57.11.2029>.
- (47) Won, C.; Mahmoudi, M.; Qin, L.; Purvis, T.; Mathur, A.; Mohsenin, V. The Impact of Gender on Timeliness of Narcolepsy Diagnosis. *J. Clin. Sleep Med. JCSM Off. Publ. Am. Acad. Sleep Med.* **2014**, *10* (1), 89–95. <https://doi.org/10.5664/jcsm.3370>.
- (48) Shen, Z.; Shuai, Y.; Mou, S.; Shen, Y.; Shen, X.; Yang, S. Case Report: Cases of Narcolepsy Misdiagnosed as Other Psychiatric Disorders. *Front. Psychiatry* **2022**, *13*, 942839. <https://doi.org/10.3389/fpsy.2022.942839>.
- (49) Dunne, L.; Patel, P.; Maschauer, E. L.; Morrison, I.; Riha, R. L. Misdiagnosis of Narcolepsy. *Sleep Breath. Schlaf Atm.* **2016**, *20* (4), 1277–1284. <https://doi.org/10.1007/s11325-016-1365-5>.
- (50) Tsuneki, H.; Wada, T.; Sasaoka, T. Chronopathophysiological Implications of Orexin in Sleep Disturbances and Lifestyle-Related Disorders. *Pharmacol. Ther.* **2018**, *186*, 25–44. <https://doi.org/10.1016/j.pharmthera.2017.12.010>.
- (51) Lividini, A.; Pizza, F.; Filardi, M.; Vandi, S.; Ingravallo, F.; Antelmi, E.; Bruni, O.; Cosentino, F. I. I.; Ferri, R.; Guarnieri, B.; Marelli, S.; Ferini-Strambi, L.; Romigi, A.; Bonanni, E.; Maestri, M.; Terzaghi, M.; Manni, R.; Plazzi, G. Narcolepsy Type 1 Features across the Life Span: Age Impact on Clinical and Polysomnographic Phenotype. *J. Clin. Sleep Med. JCSM Off. Publ. Am. Acad. Sleep Med.* **2021**, *17* (7), 1363–1370. <https://doi.org/10.5664/jcsm.9198>.
- (52) Nightingale, S.; Orgill, J. C.; Ebrahim, I. O.; de Lacy, S. F.; Agrawal, S.; Williams, A. J. The Association between Narcolepsy and REM Behavior Disorder (RBD). *Sleep Med.* **2005**, *6* (3), 253–258. <https://doi.org/10.1016/j.sleep.2004.11.007>.
- (53) Evans, R.; Kimura, H.; Alexander, R.; Davies, C. H.; Faessel, H.; Hartman, D. S.; Ishikawa, T.; Ratti, E.; Shimizu, K.; Suzuki, M.; Tanaka, S.; Yukitake, H.; Dauvilliers, Y.; Mignot, E. Orexin 2 Receptor-Selective Agonist Danavorexton Improves Narcolepsy Phenotype in a Mouse Model and in Human Patients. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119* (35), e2207531119. <https://doi.org/10.1073/pnas.2207531119>.
- (54) Malter, M.; Neuneier, J.; Triller, A.; Kallweit, U. [Narcolepsy in adults: Definition, etiology and treatment]. *Fortschr. Neurol. Psychiatr.* **2021**, *89* (3), 103–113. <https://doi.org/10.1055/a-1244-2612>.
- (55) Sansa, G.; Iranzo, A.; Santamaria, J. Obstructive Sleep Apnea in Narcolepsy. *Sleep Med.* **2010**, *11* (1), 93–95. <https://doi.org/10.1016/j.sleep.2009.02.009>.

- (56) Information (US), N. C. for B. Narcolepsy. In *Genes and Disease [Internet]*; National Center for Biotechnology Information (US), 1998.
- (57) Delzell, E. *Finding the Right Narcolepsy Medication*. WebMD. <https://www.webmd.com/sleep-disorders/narcolepsy-medication> (accessed 2024-01-19).
- (58) Mansur, A.; Castillo, P. R.; Rocha Cabrero, F.; Bokhari, S. R. A. Restless Legs Syndrome. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2024.
- (59) Ekblom, K.; Ulfberg, J. Restless Legs Syndrome. *J. Intern. Med.* **2009**, *266* (5), 419–431. <https://doi.org/10.1111/j.1365-2796.2009.02159.x>.
- (60) Karroum, E. G.; Golmard, J.-L.; Leu-Semenescu, S.; Arnulf, I. Sensations in Restless Legs Syndrome. *Sleep Med.* **2012**, *13* (4), 402–408. <https://doi.org/10.1016/j.sleep.2011.01.021>.
- (61) Hening, W.; Allen, R. P.; Tenzer, P.; Winkelmann, J. W. Restless Legs Syndrome: Demographics, Presentation, and Differential Diagnosis. *Geriatrics* **2007**, *62* (9), 26–29.
- (62) Mihic, S. J.; Harris, R. A. GABA and the GABAA Receptor. *Alcohol Health Res. World* **1997**, *21* (2), 127–131.
- (63) Van Egroo, M.; Koshmanova, E.; Vandewalle, G.; Jacobs, H. I. L. Importance of the Locus Coeruleus-Norepinephrine System in Sleep-Wake Regulation: Implications for Aging and Alzheimer’s Disease. *Sleep Med. Rev.* **2022**, *62*, 101592. <https://doi.org/10.1016/j.smr.2022.101592>.
- (64) Aston-Jones, G.; Bloom, F. E. Activity of Norepinephrine-Containing Locus Coeruleus Neurons in Behaving Rats Anticipates Fluctuations in the Sleep-Waking Cycle. *J. Neurosci. Off. J. Soc. Neurosci.* **1981**, *1* (8), 876–886. <https://doi.org/10.1523/JNEUROSCI.01-08-00876.1981>.
- (65) Park, S.; Kang, I.; Edden, R. A. E.; Namgung, E.; Kim, J.; Kim, J. Shorter Sleep Duration Is Associated with Lower GABA Levels in the Anterior Cingulate Cortex. *Sleep Med.* **2020**, *71*, 1–7. <https://doi.org/10.1016/j.sleep.2020.02.018>.
- (66) Mihic, S. J.; Ye, Q.; Wick, M. J.; Koltchine, V. V.; Krasowski, M. D.; Finn, S. E.; Mascia, M. P.; Valenzuela, C. F.; Hanson, K. K.; Greenblatt, E. P.; Harris, R. A.; Harrison, N. L. Sites of Alcohol and Volatile Anaesthetic Action on GABA(A) and Glycine Receptors. *Nature* **1997**, *389* (6649), 385–389. <https://doi.org/10.1038/38738>.
- (67) Wilson, C. S.; Mongin, A. A. The Signaling Role for Chloride in the Bidirectional Communication between Neurons and Astrocytes. *Neurosci. Lett.* **2019**, *689*, 33–44. <https://doi.org/10.1016/j.neulet.2018.01.012>.
- (68) Colrain, I. M.; Nicholas, C. L.; Baker, F. C. Alcohol and the Sleeping Brain. *Handb. Clin. Neurol.* **2014**, *125*, 415–431. <https://doi.org/10.1016/B978-0-444-62619-6.00024-0>.
- (69) Goldstein, A. N.; Walker, M. P. The Role of Sleep in Emotional Brain Function. *Annu. Rev. Clin. Psychol.* **2014**, *10*, 679–708. <https://doi.org/10.1146/annurev-clinpsy-032813-153716>.
- (70) Ma, C.; Pavlova, M.; Li, J.; Liu, Y.; Sun, Y.; Huang, Z.; Wu, S.; Gao, X. Alcohol Consumption and Probable Rapid Eye Movement Sleep Behavior Disorder. *Ann. Clin.*

- Transl. Neurol.* **2018**, 5 (10), 1176–1183. <https://doi.org/10.1002/acn3.630>.
- (71) Xiang, Y.; Zhou, X.; Huang, X.; Zhou, X.; Zeng, Q.; Zhou, Z.; Xu, Q.; Liu, Z.; Sun, Q.; Tan, J.; Yan, X.; Tang, B.; Zhang, X.; Guo, J. The Risk Factors for Probable REM Sleep Behavior Disorder: A Case-Control Study. *Sleep Med.* **2023**, 110, 99–105. <https://doi.org/10.1016/j.sleep.2023.08.007>.
- (72) Krol, R. C.; Knuth, S. L.; Bartlett, D. Selective Reduction of Genioglossal Muscle Activity by Alcohol in Normal Human Subjects. *Am. Rev. Respir. Dis.* **1984**, 129 (2), 247–250.
- (73) Grace, K. P.; Hughes, S. W.; Horner, R. L. Identification of the Mechanism Mediating Genioglossus Muscle Suppression in REM Sleep. *Am. J. Respir. Crit. Care Med.* **2013**, 187 (3), 311–319. <https://doi.org/10.1164/rccm.201209-1654OC>.
- (74) Alzoubaidi, M.; Mokhlesi, B. Obstructive Sleep Apnea during REM Sleep: Clinical Relevance and Therapeutic Implications. *Curr. Opin. Pulm. Med.* **2016**, 22 (6), 545–554. <https://doi.org/10.1097/MCP.0000000000000319>.
- (75) Garai, G.; Dueñas, M. T.; Irastorza, A.; Martín-Alvarez, P. J.; Moreno-Arribas, M. V. Biogenic Amines in Natural Ciders. *J. Food Prot.* **2006**, 69 (12), 3006–3012. <https://doi.org/10.4315/0362-028x-69.12.3006>.
- (76) Visciano, P.; Schirone, M. Update on Biogenic Amines in Fermented and Non-Fermented Beverages. *Foods* **2022**, 11 (3), 353. <https://doi.org/10.3390/foods11030353>.
- (77) Taylor-Clark, T.; Sodha, R.; Warner, B.; Foreman, J. Histamine Receptors That Influence Blockage of the Normal Human Nasal Airway. *Br. J. Pharmacol.* **2005**, 144 (6), 867–874. <https://doi.org/10.1038/sj.bjp.0706118>.
- (78) Simou, E.; Britton, J.; Leonardi-Bee, J. Alcohol and the Risk of Sleep Apnoea: A Systematic Review and Meta-Analysis. *Sleep Med.* **2018**, 42, 38–46. <https://doi.org/10.1016/j.sleep.2017.12.005>.
- (79) Ko, J.; Lim, J. H.; Kim, D. B.; Joo, M. J.; Jang, Y. S.; Park, E.-C.; Shin, J. Association between Alcohol Use Disorder and Risk of Obstructive Sleep Apnea. *J. Sleep Res.* **2023**, e14128. <https://doi.org/10.1111/jsr.14128>.
- (80) Patel, S.; Sharma, S. Respiratory Acidosis. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2024.
- (81) Rana, A. M.; Sankari, A. Central Sleep Apnea. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2024.
- (82) Costardi, J. V. V.; Nampo, R. A. T.; Silva, G. L.; Ribeiro, M. A. F.; Stella, H. J.; Stella, M. B.; Malheiros, S. V. P. A Review on Alcohol: From the Central Action Mechanism to Chemical Dependency. *Rev. Assoc. Medica Bras. 1992* **2015**, 61 (4), 381–387. <https://doi.org/10.1590/1806-9282.61.04.381>.
- (83) Wang, X. Narcolepsy Induced by Chronic Heavy Alcohol Consumption: A Case Report. *Shanghai Arch. Psychiatry* **2012**, 24 (5), 294–296. <https://doi.org/10.3969/j.issn.1002-0829.2012.05.010>.
- (84) Soyka, M. [Alcohol-induced hallucinosis. Clinical aspects, pathophysiology and therapy].

- Nervenarzt* **1996**, 67 (11), 891–895. <https://doi.org/10.1007/s001150050069>.
- (85) *What's the Link Between Alcohol and Narcolepsy?* Healthline. <https://www.healthline.com/health/narcolepsy-and-alcohol> (accessed 2023-10-24).
- (86) Research, C. for D. E. and. FDA Drug Safety Communication: Warning against Use of Xyrem (Sodium Oxybate) with Alcohol or Drugs Causing Respiratory Depression. *FDA* **2019**.
- (87) Valenzuela, C. F. Alcohol and Neurotransmitter Interactions. *Alcohol Health Res. World* **1997**, 21 (2), 144–148.
- (88) Finn, D. A.; Crabbe, J. C. Exploring Alcohol Withdrawal Syndrome. *Alcohol Health Res. World* **1997**, 21 (2), 149–156.
- (89) Park, Y.-M.; Park, H. K.; Kim, L.; Lee, H.-J.; Kang, S.-G. Acute-Withdrawal Restless Legs Syndrome Following Abrupt Cessation of Short-Term Tramadol. *Psychiatry Investig.* **2014**, 11 (2), 204–206. <https://doi.org/10.4306/pi.2014.11.2.204>.
- (90) Napierała, M.; Bogusiewicz, J.; Enko, J.; Florek, E. [Nicotine and caffeine: influence on dopaminergic transmission]. *Przegl. Lek.* **2016**, 73 (10), 791–794.
- (91) Hussain, L. S.; Reddy, V.; Maani, C. V. Physiology, Noradrenergic Synapse. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2023.
- (92) Hayat, H.; Regev, N.; Matosevich, N.; Sales, A.; Paredes-Rodriguez, E.; Krom, A. J.; Bergman, L.; Li, Y.; Lavigne, M.; Kremer, E. J.; Yizhar, O.; Pickering, A. E.; Nir, Y. Locus Coeruleus Norepinephrine Activity Mediates Sensory-Evoked Awakenings from Sleep. *Sci. Adv.* **2020**, 6 (15), eaaz4232. <https://doi.org/10.1126/sciadv.aaz4232>.
- (93) Awtry, T. L.; Werling, L. L. Acute and Chronic Effects of Nicotine on Serotonin Uptake in Prefrontal Cortex and Hippocampus of Rats. *Synap. N. Y. N* **2003**, 50 (3), 206–211. <https://doi.org/10.1002/syn.10259>.
- (94) Monti, J. M. Serotonin Control of Sleep-Wake Behavior. *Sleep Med. Rev.* **2011**, 15 (4), 269–281. <https://doi.org/10.1016/j.smr.2010.11.003>.
- (95) *REM Sleep Revealed: Enhance Your Sleep Quality*. Sleep Foundation. <https://www.sleepfoundation.org/stages-of-sleep/rem-sleep> (accessed 2023-12-18).
- (96) *REM Sleep Disorders – The Center for Sleep Medicine*. <https://www.sleepmedcenter.com/rem-sleep-disorders/> (accessed 2023-12-30).
- (97) Contributors, W. E. *REM Sleep Behavior Disorder*. WebMD. <https://www.webmd.com/sleep-disorders/rem-sleep-behavior-disorder> (accessed 2023-12-04).
- (98) Yao, C.; Fereshtehnejad, S.-M.; Keezer, M. R.; Wolfson, C.; Pelletier, A.; Postuma, R. B. Risk Factors for Possible REM Sleep Behavior Disorder. *Neurology* **2019**, 92 (5), e475–e485. <https://doi.org/10.1212/WNL.0000000000006849>.
- (99) Postuma, R. B.; Montplaisir, J. Y.; Pelletier, A.; Dauvilliers, Y.; Oertel, W.; Iranzo, A.; Ferini-Strambi, L.; Arnulf, I.; Hogl, B.; Manni, R.; Miyamoto, T.; Mayer, G.; Stiasny-Kolster, K.; Puligheddu, M.; Ju, Y.; Jennum, P.; Sonka, K.; Santamaria, J.; Fantini, M. L.; Zucconi, M.; Leu-Semenescu, S.; Frauscher, B.; Terzaghi, M.; Miyamoto, M.;

- Unger, M. M.; Cochen De Cock, V.; Wolfson, C. Environmental Risk Factors for REM Sleep Behavior Disorder. *Neurology* **2012**, *79* (5), 428–434. <https://doi.org/10.1212/WNL.0b013e31825dd383>.
- (100) Colten, H. R.; Altevogt, B. M.; Research, I. of M. (US) C. on S. M. and. Sleep Physiology. In *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*; National Academies Press (US), 2006.
- (101) Krishnan, V.; Dixon-Williams, S.; Thornton, J. D. Where There Is Smoke... There Is Sleep Apnea. *Chest* **2014**, *146* (6), 1673–1680. <https://doi.org/10.1378/chest.14-0772>.
- (102) Morrell, M. J.; Finn, L.; Kim, H.; Peppard, P. E.; Badr, M. S.; Young, T. Sleep Fragmentation, Awake Blood Pressure, and Sleep-Disordered Breathing in a Population-Based Study. *Am. J. Respir. Crit. Care Med.* **2000**, *162* (6), 2091–2096. <https://doi.org/10.1164/ajrccm.162.6.9904008>.
- (103) Kimoff, R. J. Sleep Fragmentation in Obstructive Sleep Apnea. *Sleep* **1996**, *19* (9 Suppl), S61-66. https://doi.org/10.1093/sleep/19.suppl_9.s61.
- (104) Lamon, S.; Morabito, A.; Arentson-Lantz, E.; Knowles, O.; Vincent, G. E.; Condo, D.; Alexander, S. E.; Garnham, A.; Paddon-Jones, D.; Aisbett, B. The Effect of Acute Sleep Deprivation on Skeletal Muscle Protein Synthesis and the Hormonal Environment. *Physiol. Rep.* **2021**, *9* (1), e14660. <https://doi.org/10.14814/phy2.14660>.
- (105) Blumen, M.; Chabolle, F.; Rabischong, E.; Rabischong, P.; Frachet, B. [Dilator muscles of the pharynx and their implication in the sleep apnea syndrome of obstructive type. Review of the literature]. *Ann. Oto-Laryngol. Chir. Cervico Faciale Bull. Soc. Oto-Laryngol. Hopitaux Paris* **1998**, *115* (2), 73–84.
- (106) Sériès, F.; Cormier, Y.; La Forge, J. [Physiopathology of obstructive sleep apneas]. *Rev. Mal. Respir.* **1989**, *6* (5), 397–407.
- (107) Teixeira, M. Z. Rebound Effect of Modern Drugs: Serious Adverse Event Unknown by Health Professionals. *Rev. Assoc. Medica Bras. 1992* **2013**, *59* (6), 629–638. <https://doi.org/10.1016/j.ramb.2013.05.003>.
- (108) Fotuhi, O.; Fong, G. T.; Zanna, M. P.; Borland, R.; Yong, H.-H.; Cummings, K. M. Patterns of Cognitive Dissonance-Reducing Beliefs among Smokers: A Longitudinal Analysis from the International Tobacco Control (ITC) Four Country Survey. *Tob. Control* **2013**, *22* (1), 52–58. <https://doi.org/10.1136/tobaccocontrol-2011-050139>.
- (109) Pedersen, K. M.; Çolak, Y.; Ellervik, C.; Hasselbalch, H. C.; Bojesen, S. E.; Nordestgaard, B. G. Smoking and Increased White and Red Blood Cells. *Arterioscler. Thromb. Vasc. Biol.* **2019**, *39* (5), 965–977. <https://doi.org/10.1161/ATVBAHA.118.312338>.
- (110) Amulic, B.; Cazalet, C.; Hayes, G. L.; Metzler, K. D.; Zychlinsky, A. Neutrophil Function: From Mechanisms to Disease. *Annu. Rev. Immunol.* **2012**, *30*, 459–489. <https://doi.org/10.1146/annurev-immunol-020711-074942>.
- (111) Rosales, C. Neutrophil: A Cell with Many Roles in Inflammation or Several Cell Types? *Front. Physiol.* **2018**, *9*, 113. <https://doi.org/10.3389/fphys.2018.00113>.
- (112) Dinarello, C. A. Proinflammatory Cytokines. *Chest* **2000**, *118* (2), 503–508.

- <https://doi.org/10.1378/chest.118.2.503>.
- (113) Kim, K. S.; Kim, J. H.; Park, S. Y.; Won, H.-R.; Lee, H.-J.; Yang, H. S.; Kim, H. J. Smoking Induces Oropharyngeal Narrowing and Increases the Severity of Obstructive Sleep Apnea Syndrome. *J. Clin. Sleep Med. JCSM Off. Publ. Am. Acad. Sleep Med.* **2012**, *8* (4), 367–374. <https://doi.org/10.5664/jcsm.2024>.
- (114) Bamford, I. J.; Bamford, N. S. The Striatum’s Role in Executing Rational and Irrational Economic Behaviors. *Neurosci. Rev. J. Bringing Neurobiol. Neurol. Psychiatry* **2019**, *25* (5), 475–490. <https://doi.org/10.1177/1073858418824256>.
- (115) Marsden, C. D. Dopamine and Basal Ganglia Disorders in Humans. *Semin. Neurosci.* **1992**, *4* (2), 171–178. [https://doi.org/10.1016/1044-5765\(92\)90015-T](https://doi.org/10.1016/1044-5765(92)90015-T).
- (116) Schwarz, P. B.; Peever, J. H. Dopamine Triggers Skeletal Muscle Tone by Activating D1-like Receptors on Somatic Motoneurons. *J. Neurophysiol.* **2011**, *106* (3), 1299–1309. <https://doi.org/10.1152/jn.00230.2011>.
- (117) Benowitz, N. L. Pharmacology of Nicotine: Addiction, Smoking-Induced Disease, and Therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **2009**, *49*, 57–71. <https://doi.org/10.1146/annurev.pharmtox.48.113006.094742>.

The Harmonizing Role of Socially Beneficial Finance on Pakistan-India Relations and the Prospect of Peace By Sheryar Afridi

1.1 Abstract

This paper presents an innovative approach to Pakistan and India relations, focusing on the impacts of social impact finance. These two countries, which are also nuclear-powered nations, have had a history marred with conflict, territorial claims, and suspicion. Nevertheless, there is room for economic interactions, especially for peace. This research covers the impact of social impact finance policies – such as investments looking for social and environmental dividends alongside financial ones on the relationships and national peace-building efforts. Social inputs heightened the opportunities created through inter or intra-dependency in areas like healthcare, education, and renewable energy suffices. The research employs a qualitative method through case studies and available literature to support the argument that trade can help soothe socio-economic tensions and promote peace among the countries. Evaluation of policies aimed at deepening cross-border economic relations, dealing with corruption issues, and integrating international assistance in pursuing peace are also proposed.

1.2 Introduction

Pakistan and India have had their dramatic share of the history since their partition in 1947, which is characterized by several wars, territorial claims, and political conflicts, respectively. Perhaps the most contentious issue component of disharmony is Kashmir, a territory subject to protracted territorial disputes and military confrontations (Rid, 2020). War and violence, despite the lack of complete resolution in many instances, are the only recourse between the two nations. Repeated failures of meetings and compromises have doomed both countries to the brink of sustainable peace: deep distrust. Though diplomatic negotiations and nonmilitary treaties have proven futile in solving underlying problems, political and economic cooperation is a different solution.

The concept of economic dependence between countries has been around for quite some time, and with it, the attempts to limit the probability of war occurring between them. Instead, shared interests erode the propensity for aggression among nations. In the India-Pakistan situation, there is social impact financing, which is brought in for the benefit of society. This type of funding aims to tackle and resolve socio-economic problems through financial returns, recognizing the importance of peace (Khan, Rani & Khan, 2023). This research paper addresses providing social impact finance in Pakistan and India as a basis for both countries' positive joint economic development and overcoming socio-economic issues in the healthcare, education, and renewable energy sectors, which sincerely require collaborative efforts.

2.0 Background

Hostility between Pakistan and India has its roots in the British-India partition that led to violence, displacement of people, and the creation of the two countries. This division reinforced

bitter relations between the two nations, especially the western region of Kashmir, which has been and remains a point of contention. The two nations have engaged in a handful of wars over the area but remain tense militarily and diplomatically (Adnan & Fatima, 2020). Other underlying factors such as religion, nationalism, cross-border militant groups, and the struggle for hegemony have aggravated the Kashmir dispute.

Although there has been a political crunch and security concerns that led to conflicts, there have also been some periods of peace that have been aided by trade and cultural exchange. On the contrary, most political endeavors have been politically motivated and subjected to security issues. An example of the controversy is that although their relations warmed up briefly in the early 2000s, the potential for cross-border commuting holding and trade between Pakistan and India has remained dormant (Haider & Azad, 2021). Whenever economic collaboration is strategically applied, it may create other interdependences that would lower the chances of conflict. Cross-border economic cooperation and interdependence have worked in many conflict regions, including Northern Ireland, where such economic linkages were central to stemming violence (Azam et al., 2024).

On the other hand, social impact finance makes investments intending to generate a social impact and an economic return. A balanced development that involves social impact and finance and prevents funds on both scales can also expand the commercial and cooperative capabilities of industries crucial for the growth of one or both countries. Such an approach addresses the primary drivers and initiators of explosive factors such as poverty, inequality, and lack of essential services. Also, it builds peace in the long run through economic cooperation.

3.0 Methodology

This study utilizes a qualitative methodology to examine the role of social impact finance in peacebuilding between Pakistan and India. The methodology draws insights from literature and other sources, including case studies and data analysis, to understand how social impact finance can lead to increased economic interactions between the two countries.

3.1 Literature Review

The first section surveys the existing literature on Pakistan-India relations, peace processes, and economics to form the base of the study. The literature investigates past studies on trade across borders and its efficacy in calming tensions and social impact finance for development (Shukla, 2020).

3.2 Case Studies

Among the cases reviewed, the author of this paper addresses the case of the Aman ki Asha movement, which sought to mobilize civil engagements through people's diplomacy between the Indian and Pakistani peoples and other case studies about social impact finance initiatives from other war-torn countries. These cases show how social impact investments and economic integration may be harnessed to alleviate conflict (Rid, 2020).

3.3 Data Presentation

Trade and investment data between India and Pakistan are evaluated to highlight the fields in which social impact finance would have the most positive influence. The analysis demonstrates how merging research, policy, and the new economy can address fundamental issues for humanity. (Khan et al, 2023)

4.0 Results

The research stresses that social impact finance can help build more economic ties between Pakistan and India, making it more conducive for connecting people. However, specific vital sectors, such as healthcare, education, and renewable energy, offer great potential for the development of investment opportunities that transcend borders and address common challenges. Healthcare, education, and renewable energy are important sectors that will bring overall socio-economic development and investment, thus promoting long-lasting relations between the two nations.

Nevertheless, encouraging economic development has been challenging for Pakistan and India, especially in the rural and remote regions where proper healthcare services are scarce. Healthcare-focused initiatives financed through social impact bonds could help improve the medical system concerning the available facilities and medical attention and lead to the mitigation of health in the population. This would not only foster job creation because people will be hired for the specific healthcare projects taken up but also improve the population's living standards; hence, goodwill between the two countries diminishes the chances of conflict (Haider & Azad, 2021). The scope for collaboration in healthcare is the possibility of exchanging this knowledge and such technologies and innovations with those of other nations.

Education is another of the areas in which social impact finance could be utilized, and in this case, transform the whole sector. The case is similar for India, as large portions of the country have limited geographical access to sufficient educational facilities. Such cross-border education-related investments will be significant since they will reduce the literacy gap by executing projects that provide quality education and allow people to become more economically mobile. Therefore, by putting money into education, the two nations can reduce poverty levels, have an educated and professional workforce, and have economic advancement in the long run (Misson, 2022). Education, in addition to improving understanding and trust orientation, could aid in resolving such issues where a nation seeks other nations' help to create peace.

In addition, the sphere of renewable energy suggests new avenues of collaboration. Energy shortages plague both nations, and they are forced to depend mainly on fossil fuels, which are detrimental to the environment. There is a possibility of facilitating social impact investment on such projects in renewable energy as joint enterprises in solar or wind energy generation for both countries (Aslam, 2022). Such measures would not only mitigate emissions but also surmount the issue of external dependency and hence improve the two countries' energy security. Environmental sustainability could become a common interest that promotes regional cooperation and mitigates conflict over-dependence on natural resources.

The study also stresses the relevance of a transparent, accountable, and inclusive approach to development in cross-border business collaboration. The distinctive characteristic of social impact investments is their inherent need for an effective monitoring and evaluation mechanism to guarantee the investments' desired financial and social outcomes. This could complement the independent mechanisms, which raise the level of confidence from investors that enabling conditions are put in place to facilitate the optimal management of cross-border projects. Moreover, such businesses in other countries must adopt a social impact finance approach, which prioritizes disadvantaged social groups, to ensure that equality of the benefits of economic interaction is attained (Riaz et al., 2024). Addressing the social investment void can reduce the economic conditions that generally lie within violent conflicts.

5.0 Discussion

While the path between social impact finance and the creation of peace within South Asia, especially between India and Pakistan, is vividly clear, many challenges need to be tackled if these initiatives materialize. Political risk factors are still based on historical, past, and present political tensions and security concerns that hamper efforts at cooperation. Notably, social impact finance reverses this unfortunate trend as it seeks to promote bottom-up concepts close to the people without necessarily being engineered through the central authority. These efforts, particularly those focusing on development and inclusiveness, can help generate trust and collaboration among the people even in the absence of formal political agreements.

Among others, one of the most significant barriers associated with economic collaboration between Pakistan and India is the high distrust between the two countries' governments. Such attempts aimed at developing economic relations have, time and gains, been thwarted by security issues and political tensions stemming from the Kashmir dispute (Naseer, 2024). Therefore, Such risk must be remedied through the need to make economic relations and cooperation unaffected by politics and ensure that political pressures do not affect social impact investments. This can be done through liberalization of the funding and letting in international lending bodies and aid agencies that can give money and skills to back up regional projects (Khan et al., 2024).

Another issue that needs urgent attention is the difference in the economies of both countries, more so Pakistan and India. In this regard, India could reap more significant benefits from the cross-border movement of goods and capital. To overcome this, it is suggested that social impact finance projects be concentrated on the less developed areas in both countries, especially Pakistan, to prevent the inequitable distribution of the gains of economic cooperation (Riaz et al., 2024). Social finance could mitigate the economic disparity between the two countries by fostering the purpose of social justice by reaching out to discriminated and underprivileged sections of society.

India and Pakistan also face grave drawbacks in the countries' governance systems. Corruption, inefficiency, and Leadership interventionism are recurrent problems in Pakistan and India and may pose barriers to developing cross-border economies. To reduce these prospects,

accountability, and transparency are the primary goals. Ventures to raise social impact finance should include adequate structures for accountability, such as having an independent agency in charge of accountability for the impact of the funds allocated and used and the intended societal impact (Ashraf, 2020). Once these frameworks are put in place, both countries would be able to measure the achievement of social impact investment and the returns it brings to the country's people.

The importance of international organizations in advancing social impact finance initiatives is beyond describable limitations. Put differently, countries with the support of the World Bank, Asian Development Bank, and UN can also be practical for offshore work's financing, technical, and managerial aspects (Haider & Azad, 2021). These agencies can also be neutral third parties in conflict resolution situations to ensure that the interdependency of the nations affected remains unchanged and exceptionally healthy growth continues to rise irrespective of the prevailing political temperatures. With the involvement of external organizations in the projects, cross-border risks will be minimized by Pakistan and India so that a conducive factor for economic interaction will be enhanced.

Private sector participation is another important factor that leads to the effectiveness of social impact finance initiatives. Private investors both local and foreign can augment considerable economic growth through social impact investment; however, this is also restrained by significant factors on the exogenous circumstances in both Pakistan and India, which is the improvement of the local investment climate. This also means policies that liberalize the economy in favor of the private sector—reducing administrative barriers as barriers to entry investing capital, mainly foreign, in social factors, improving the senior levels of the management hierarchy or integrity, and providing protections for investment laws. Both nations, Nepal and India, can create a complete environment to privately hold capital towards the program to usher social impact investing and economic growth.

Last but not least, the human contact aspect of diplomacy must be promoted to consolidate the efforts towards economic cooperation. Projects such as Aman ki Asha, which focuses on exchanges between the citizens of India and Pakistan, have emerged as an effective tool for combating stereotypes and fostering understanding (Rid, 2020). Implementing such projects more widely, especially supporting them with social impact finance projects, can make it possible to level peace-building work by facilitating communication and interactions on the grassroots level. People-to-people diplomacy can also effectively clarify the case for cross-border economic collaboration and social impact finance.

6.0 Conclusion and Recommendations

In the case of India and Pakistan, social impact finance is a catalyst that can be used for conflict resolution through the economic integration of these two nations and by addressing their common socio-economic challenges. By directing investments into sectors like health care, education, or renewable energy, both countries will reap rewards that will decrease the chances of war and enhance enduring peace. The study's findings suggest that social impact finance is a

viable means to peace once the aspects of participation and fostering the local economy take center stage.

Still, several barriers need to be dealt with for social impact finance to be helpful in the promotion of peace. Political divisions, economic imbalances, demographics, and governance remain core barriers to transnational integration. To address these problems, both countries must ensure that their business ventures are conducted in an adhesive manner, emphasizing transparency and inclusion. International structures can assist in this process, including social impact funds that seek to promote investments in such structures, guaranteeing the satisfactory execution and efficiency of the investments made.

6.1 Recommendations

6.1.1 Expand Cross-Border Social Impact Investments

The governments of both Pakistan and India ought to give importance to the development of cross-border social impact initiatives, particularly in the sectors of health, education, and renewable energy. Improving these sectors is essential because the potential for development exists within the economies and because they seek to tackle global challenges and achieve equitable development. Offering tax holidays and subsidies and lowering trade restrictions to facilitate private sector investment in these subsectors would benefit the government in both countries.

6.1.2 Create a Social Impact Finance Fund

There is a need to develop an exclusive social impact finance fund that would be raised and pooled from international organizations such as the World Bank, Asian Development Bank, individual private investors, etc., for the funding of cross-border projects for India and Pakistan. Such a fund could accomplish both goals by investing in sectors that would remove political and security barriers while disbanding many securities and profitability, such as education, healthcare, and renewable energy sources. The fund should focus on investments in development-tarred, conflict-torn regions to pool resources and benefits adequately.

6.1.3 Teach Transparency & Accountability

Social impact finance projects must not be mismanaged, abused, or subjected to political interference, and in order to achieve this, the two countries must enforce the legal frameworks and anti-corruption measures even further. Create external organizations to check social impact projects on their ability to deliver profit and the targeted level of social impact on a case-by-case basis. In this way, they shall strengthen the protection of the investors and promote the viability of cross-border investment in the long run.

6.1.4 Politicization

Economic cooperation between Pakistan and India must be depoliticized to avoid constant interruptions due to contentious politics. Many such activities enhance trust and communication at the state and society levels, such as confidence-building measures, which include joint research, scholarly exchanges, and cultural exchange. Such measures should go hand in hand with the social impact finance projects to promote such peacebuilding efforts.

6.1.5 Invite International Organizations

International organizations, including the United Nations, World Bank, and the Asian Development Bank, should become significant players in coordinating and managing social financing projects between Pakistan and India, as well as supporting their cross-border social finance projects. They can act as finance risk guard, conduct sensitive negotiations, and provide technical input during conflict. Their participation would provide an independent environment where other social impact finance programs operate, regardless of political relations between the nations.

6.1.6 Foster People-to-People Diplomacy

People projects such as Aman ki Asha that seek to promote amicable relations between the people of India and Pakistan can also serve the objectives of social investing. These programs operate by dispelling cultural misconceptions, facilitating understanding, and promoting conversation among the populations of both nations. Peace and economic relations require adequate diplomacy from the bottom, and such diplomacy can be built through grassroots initiatives.

6.1.7 Develop Renewable Energy Partnerships

Given their environmental and energy concerns, Pakistan and India should pursue partnerships in the renewable energy sector. There are opportunities to raise solar and wind financing for the benefit of both countries through social impact investment. In addition to meeting the demand for clean and renewable energy, these projects will also help enhance cooperation in areas of mutual interest. Clean energy projects can help countries like Pakistan and India reduce their reliance on fossil fuels, combat global warming, and foster further cooperation in the future.

6.1.8 Focus on Marginalized Regions

Social impact finance policies, plans, and strategies should target social impact investments in the more neglected and underdeveloped areas of Pakistan and India, especially the rural regions. Thus, targeting the region and social impact investments can help alleviate regional disparities, which are often a cause for social tension and strife. Inclusive growth, which seeks to better the conditions of low-income people, also goes towards long-term peace and stability.

6.2 Conclusion

The concept of social impact finance presents a thorough and creative way to promote peace between Pakistan and India. With an emphasis on synergistic activities in healthcare, education, and renewable energy industries, social impact finance will create opportunities for mutual interests and mutually beneficial interdependencies necessary to avoid future hostilities. If the two countries paid more attention to cross-border investments, which would encourage inclusiveness and local and sustainable development, they would jointly deal with issues that lead to violence, like poverty, inequality, and lacking access to essential services.

Nevertheless, some issues need to be dealt with to harness the full potential of social impact financing as a peace-building strategy. Overall, political and economic divisions, governance issues, and the will to ensure accountability and transparency have remained a bitter reality for quite some time. These hurdles to attaining economic cooperation between Pakistan and India include using international bodies, enhancing legal procedures, and people-to-people contacts.

The study also points out that there can be no lasting peace between Pakistan and India without political and military means – that is a wrong conclusion. There needs to be a comprehensive approach that covers economic, social, and environmental aspects. Social impact finance works precisely like that. Since it has positive societal embedment while providing financial returns, it is very suitable. Considering and investing in areas that foster joint welfare and security can bring both nations closer to achieving real peace and unity in the region.

Works Cited

- Adnan, M., & Fatima, B. (2020). The peace process with India: A challenge for Pakistan. *South Asian Studies*, 31(1).
<https://journals.pu.edu.pk/journals/index.php/IJSAS/article/view/3047>
- Ashraf, T. (2020). Economic Relations between Pakistan and India: Impressions from the Past. *Global Strategic & Security Studies Review*, 22–33.
<https://doi.org/10.31256/GSSR.2020.07.23>
- Aslam, B. (2022). Hydro-diplomacy and the prospects of environmental peacebuilding between Pakistan and India. *Journal of Humanities, Social and Management Sciences (JHSMS)*, 3(1), 204–216. <https://doi.org/10.52633/jhsms.v3i1.75>
- Azam, S., Hussain, R., & Khan, F. (2024). Trading for Peace: How Economic Ties Can Pave the Way for Pakistan-India Coexistence. *Journal of Global Peace and Security Studies (JGPSS)*, 5(2). <https://www.pakistanreview.com/index.php/JGPSS/article/view/313>
- Haider, M. W., & Azad, T. M. (2021). The role of confidence-building measures in developing relations between Pakistan and India. *World Affairs*, 184(3), 294-317.
<https://doi.org/10.1177/00438200211030222>
- Khan, A., Rani, F., & Khan, P. (2023). India and Pakistan Trade: A Vision for Peace. *Global Political Review*, VIII (8), 29-38. [https://doi.org/10.31703/gpr.2023\(VIII-I\).04](https://doi.org/10.31703/gpr.2023(VIII-I).04)
- Khan, M., Alam, R., & Yasir, M. (2024). China-Pakistan Economic Corridor (CPEC) and Regional Economic Dependency: Fostering Peace and Stability in the Region. *Jahan-e-Tahqeeq*, 7(1), 899-916. <https://doi.org/10.47391/JET.2024.07.21>
- Misson, F. A. (2022). Pakistan-India Relations. *Strategic Studies*, 42(2), 54-71.
<https://www.jstor.org/stable/48732350>
- Naseer, S. (2024). Regional Power Balance and Peace Building: India and Pakistan. *Journal of Indian Studies*, 10(1), 121–140. <https://doi.org/10.1234/jis.v10.1.2024.08>
- Pathak, M., & Mushahary, J. (2024). A Descriptive Study on Bilateral Relations Between India and Pakistan: Causes of Rifts, Peace Agreements, and Nuclear Tests Impacts. *Acta Scientiae*, 7(1), 581-596. <https://doi.org/10.5563/actasci.2024.131>
- Riaz, M., Simon, A., Syed, N., Majeed, C. N., & Abbas, S. (2024). Economic Development Rapport With Financial Inclusion Composite Index.
<https://doi.org/10.5281/zenodo.382799631>
- Rid, S. A. (2020). Aman ki Asha (a desire for peace): A case study of a people-to-people contacts peacebuilding initiative between India and Pakistan. *Contemporary South Asia*, 28(1), 113-125. <https://doi.org/10.1080/09584935.2019.1666090>
- Shukla, A. (2020). Theoretical underpinnings of India–Pakistan relations. *India Quarterly*, 76(2), 294–312. <https://doi.org/10.1177/0974928420917785>

Tesla's Use of Artificial Intelligence in Self-Driving Cars By Yitao Wang

Summary

Artificial Intelligence is a rapidly growing branch of computer science in which software is designed to use massive amounts of data to improve its functioning by learning from past experience. This means that the software automatically updates itself without the need for more coding from human programmers. Artificial Intelligence is being applied in many different areas, such as medicine, data science, robotics, computer imaging, computer vision, robotics, finance and the arts to increase efficiency and create new technologies. One of the areas in which Artificial Intelligence has been applied is in the design and production of electric vehicles; in particular, self-driving electric vehicles. Tesla has been a pioneer in the development of Artificial Intelligence used in the production of safer electric vehicles and is currently developing its own version of a Robotaxi (self-driving taxi) with an eye toward capturing a large share of what is expected to be an extremely lucrative market in the future. Another major area pursued by Tesla is Robotics, which are intended to be used in industry and also sold to consumers for household tasks. To date, Tesla appears to be falling behind some of its competitors such as Waymo (Google) and is gambling on its Robotaxi to help maintain its status as one of the most profitable companies in the world as well as a major innovator in the area of Artificial Intelligence. In order to promote Tesla's self-driving vehicles and its robots, the company held a widely-hyped event "We, Robot" on October 10, 2024. This gave Tesla an opportunity to show off the results of its intensive research into Artificial Intelligence, displaying the Robotaxi and the Optimus Robot as well as a surprising new product known as the "Robovan." Unfortunately, investors were disappointed as the times to delivery of these products appeared to be uncertain and further in the future than anticipated, with many unanswered questions remaining to be resolved.

1) Introduction

Artificial intelligence is a rapidly advancing technology that has revolutionized many fields such as pharmaceuticals, data science, gaming, computer imaging, robotics, financial planning, etc. Among its many potential applications, artificial intelligence is currently being used to design and build safer cars with more advanced capabilities that can be used in self-driving cars. Tesla is a pioneer in the use of this new technology to design and develop its new generation of electric vehicles, along with computer chips and robots. Tesla is also focusing much of its attention on developing self-driving cars, in particular a "Robotaxi" as a way of capturing a potentially lucrative market in the future. Although these cars still face many obstacles, Tesla is showing how Artificial Intelligence can be used to produce dramatic improvements in the technology required to develop self-driving cars. Tesla is facing stiff competition from companies such as Waymo (Google), Amazon (Zoox), General Motors, Mercedes Benz, etc. as this market is expected by observers to be potentially worth hundreds of billions of dollars in the future.

2) What is Artificial Intelligence?

Artificial Intelligence (AI) is a specialized type of computer technology that is specifically intended to replicate human decision-making. Unlike other types of computer applications, Artificial Intelligence uses massive amounts of data to “learn” from experience and reduce mistakes in order to make better decisions in the future. For example, Tesla’s self-driving cars are equipped with cameras, radar, sensors, etc. to gather data about a driver’s experiences so that the car can improve its driving capabilities by learning from its analysis of massive amounts of data.

Artificial Intelligence as a field dates back to the early 1950’s and is found in the work of Alan Turing, the renowned computer scientist whose codebreaking work at Bletchley Park helped to end World War II. Over the next several decades, Artificial Intelligence applications developed slowly due to constraints on data availability and computer processing speed. Recently, with rapidly increasing computer power and the ready availability of massive amounts of data, many new Artificial Intelligence applications have become available. One well-known example is ChatGPT, which was released in November 2022. ChatGPT, which is part of OpenAI, has made it possible to increase efficiency in many areas such as:

- improving online customer service including recommendations for new products
- content creation (such as newspaper articles, computer programs, art, etc.)
- automating routine tasks
- improving medical diagnoses
- creating story lines and scripts for video games
- translating from one language to another
- writing computer code

Artificial Intelligence is also being used in many types of research, such as:

- machine learning/deep learning – this enables a computer to learn from experience and update software based on previous experience without specific instructions from programmers
- neural networks - a programming approach that processes information in a way that replicates the way that humans learn from experience
- natural language processing – increases the ability of computers to gather information from interactions with people based on everyday language making it easier to obtain useful results; for example, Apple’s Siri program enables users to request information by having a dialog with Siri
- computer vision – enables computers to recognize patterns and interpret their meanings, which is particularly useful for self-driving cars and facial recognition software
- education – customizing educational materials to the needs of individual students

Artificial Intelligence has many advantages but some potential disadvantages as well. Some of its advantages are increased processing speed, a higher level of accuracy, more flexibility, etc. Some of the drawbacks include the potential for criminal activity, the possibility of deliberately spreading misinformation and the potential disruption to the job market as older technologies are phased out and replaced with Artificial Intelligence.

3) Potential Applications of Artificial Intelligence

Artificial Intelligence is currently being applied to a wide variety of applications; the number of applications that AI is applied to is likely to increase rapidly in the future. For example, in the field of medicine, AI can be used for improving the diagnosis of patients, developing more effective forms of treatments for patients, producing drugs more efficiently, etc. In the field of finance, AI can be used to generate more accurate forecasts of future economic conditions; it can be also used to improve risk management practices. It can also be useful for financial planning purposes as it can be used to process massive amounts of information very quickly. In the field of marketing, AI can be used to better determine the preferences of consumers so that marketing campaigns can be more effectively targeted to the relevant audience. In the field of entertainment, AI can be used for a wide variety of creative efforts, such as creating video games, helping in the development of movie and TV scripts, writing stories, developing art and music, etc. In the field of computer science, AI can be used to generate computer code in many different languages based on the requirements of end users.

In the automotive field, AI is used to help design new cars, increase the ability to manufacture complex parts, etc. In addition, AI is being used for improved navigation by analyzing traffic patterns, road conditions and user preferences to find the optimal path for reaching a specified destination. AI is also used to help the driver make safer decisions, such as applying brakes when other cars are extremely close, steering the car away from hazards, identifying potential obstacles on the road, etc.

Artificial Intelligence has been used in e-commerce for several years to improve the shopping experience of customers. Amazon has used Artificial Intelligence to make customer recommendations for many years. In addition, Amazon offers its own virtual assistant program called Alexa, which can be used to respond to questions, set up reminders, turn on and off lights, set alarms, etc. AI can also be used to design ads that are more appealing to customers.

4) Artificial Intelligence and Self-Driving Cars

One of the most heavily researched areas in recent years is the potential use of Artificial Intelligence to produce self-driving cars. These are cars that use information gathered from cameras and sensors to “learn” from experience how to drive a car without input from human passengers. Tesla is one of the leading researchers in this area and plans to introduce self-driving taxis (“Robotaxis”) in the fall of 2024. Waymo (a subsidiary of Alphabet) is currently producing self-driving taxis that are available in San Francisco, Los Angeles and Phoenix, with the number of weekly rides currently exceeding 100,000. In China, Pony.ai produces several different

autonomous vehicles, including trucks. Pony.ai has obtained permission to run its Robotaxis in Beijing and hopes to expand into South Korea and Saudi Arabia. Amazon's subsidiary Zoox is currently testing a Robotaxi in California.

Self-driving cars are designed to enable a car to make all decisions and carry them out without any input from the driver. The software that makes this possible is based on Artificial Intelligence methods that enable the car to obtain a continuous stream of real-time information from the car and then use this information to make more informed decisions in the future. At the moment, Tesla and General Motors have Level 2 functionality available in their cars, while Waymo is at Level 4. Audi, Honda and Mercedes-Benz offer Level 3 functionality. The six levels of self-driving cars are:

Level 0: no automation

Level 1: some basic capabilities such as automatic breaking, blind-spot warnings, etc.

Level 2: contains the features of Level 1 plus the ability to change lanes and park automatically. With levels 0, 1 and 2 the driver must continuously monitor the environment and is legally responsible for any accidents that take place. With higher levels, the car can perform more of the tasks normally done by human drivers; these are:

Level 3: can perform more advanced driving tasks but still requires human intervention

Level 4: the vehicle can perform all driving tasks but human intervention is still possible

Level 5: the vehicle can perform all driving tasks; there is no human intervention

With levels 3, 4 and 5 the vehicle is monitoring the environment with optional input from the driver. At these levels, the vehicle is legally responsible for accidents.

Self-driving cars require the use of sensors to gather large amounts of information about the driving environment in real time. Lasers are used to emit pulses and the time taken to be reflected is measured, providing a three-dimensional image of the driving environment. This process is known as Lidar, which stands for Light Detection and Ranging. Lidar is also used to develop weather forecasting models as well as surveying applications. This technology makes it possible to identify hazards on the road such as pedestrians, stalled vehicles, etc. Radar is also used to identify the location and speed of nearby objects. Radar is less sensitive to weather conditions than Lidar. Self-driving cars also make extensive use of cameras to interpret signs, traffic lights, traffic patterns, etc.

Self-driving cars require substantial amounts of data to be processed rapidly in real-time; they also require sensors that are highly accurate. Artificial intelligence can be used to ensure that data is analyzed rapidly and accurately and can also be used to ensure that sensors are properly calibrated.

In order to continuously update the monitoring of the driving environment, machine learning algorithms are used to ensure that the self-driving vehicle can recognize hazards, make accurate decisions in real time, respond to changes in the driving environment, etc. Machine

learning algorithms are designed to make it possible for software to be updated to reflect new information without explicit coding from programmers.

Artificial Intelligence can be used in several ways to make a car safer and more efficient, such as:

- o choosing the most efficient path to reach a specific goal
- o braking or steering a vehicle to avoid collisions
- o predicting the behavior of other vehicles or pedestrians on the road

There are many advantages to using Artificial Intelligence in self-driving cars, such as:

- o reduced risk of human error
- o quicker response time to hazards
- o the ability to adjust driving to local conditions
- o less traffic congestion

There are also some key drawbacks to using Artificial Intelligence in self-driving cars, such as:

- o the risk of poor decision making by the Artificial Intelligence software
- o the potential for hacking of the data gathered by a self-driving car, which could lead to the loss of personal information for the passengers in the car
- o legal/regulatory issues
- o liability and insurance issues
- o the potential for sensors to be affected by poor weather conditions
- o difficulty in integrating the artificial intelligence applications in a car with other systems, such as navigation
- o high cost; cars built with extensive Artificial Intelligence applications are likely to be very expensive

5) The Use of Artificial Intelligence in Tesla's Self-Driving Cars

Tesla has been a pioneer in the use of Artificial Intelligence algorithms in the production of automobiles and in the development of self-driving cars since 2014-2015 when it released its Autopilot software and included the necessary hardware in its electric vehicles. Tesla combines Artificial Intelligence with “Big Data” software (techniques designed to process huge amounts of information very quickly) to analyze information about car's environment to make rapid decisions about potential hazards on the road, the behavior of other vehicles, lighting conditions, information about speed limits, the optimal path to the driver's destination, etc. The information needed to make real-time decisions comes from an array of cameras and sensors. Machine learning methods are used to enable the car to “learn” from its past mistakes and adjust its programming to improve in the future. Machine learning can also be used to customize the driving experience to the individual driver.

Artificial intelligence can also enable the car to perform diagnostic tests on itself and ensure that there are no issues that need to be attended to. For example, it could inform the driver that the tires need to be changed or that the car needs a new battery.

Artificial Intelligence is also being used to customize the car's environment to the preferences of the driver. For example, it can automatically produce the ideal temperature in the car as well as play the types of music preferred by the driver.

Another application of Artificial Intelligence is to find the optimal path for a vehicle, which can potentially reduce congestion on the roads so that driving is more efficient for all vehicles.

One of the ways that Tesla is developing its Artificial Intelligence capabilities to produce its own specialized silicon computer chips that are optimized for AI applications. Specifically, they are designed to rapidly process huge amounts of information very quickly. Although this is a very expensive process, Tesla sees this as giving it more control over its applications and is expected to be more cost effective in the long run than relying on outside vendors to produce these chips, such as Intel or Nvidia.

Tesla has been developing a Robotaxi for at least five years, but there have been many delays along the way. Elon Musk has famously claimed that Tesla is now an Artificial Intelligence company rather than an auto company, which shows that it is counting heavily on the self-driving technology for its future profits. Tesla's profitability has soared over the second half of 2024; some of this is due to improved electric vehicle sales, but at least some of this is based on expectations of future profits from Robotaxis. According to the consulting firm McKinsey, by 2035 sales of autonomous vehicles could reach \$300 billion to \$400 billion or more, which explains why Tesla has dedicated so much time and money to developing its Robotaxi.

Tesla has also announced that it plans on introducing its Full Self-Driving Service (FSD) in Europe and China in the first quarter of 2025. Full Self-Driving Service is not yet capable of making vehicles completely self-driving, but it is a large step in that direction. FSD is a major component of Tesla's strategy to turn itself into an Artificial Intelligence company and away from being primarily an electric vehicle manufacturer. In the meantime, Tesla is expected to announce a low-cost electric vehicle known as the Model 2 to stimulate sales in the short run.

6) Conclusions about Tesla's Self-Driving Cars and Future Directions of Artificial Intelligence

Elon Musk's claim that Tesla is now an Artificial Intelligence and robotics company rather than an auto manufacturer is supported by several key factors. The company has not released any new types of electric cars since 2023 due to several factors, such as shrinking demand for electric vehicles, price cuts by Tesla, increased competition, etc. While the demand for electric vehicles has slowed, the demand for hybrid vehicles has accelerated due to the lack of charging stations and the high cost of electric vehicles. Tesla has refused to produce hybrid vehicles and gives no indication of doing so in the future.

As a result, the focus of Tesla is shifting toward more advanced technology in the areas of Artificial Intelligence and Robotics. Elon Musk has claimed that Tesla's Optimus robots will soon be used on its production lines to produce electric vehicles and he also expects them to be sold to individual consumers to perform routine tasks that humans would rather avoid. Tesla also has a thriving energy storage business, and it is focusing more of its attention on its self-driving cars, which it believes are the wave of the future.

Tesla expects to make its "Full Self-Driving" driver assistance software available in Europe and China in the first quarter of 2025 (assuming that Tesla receives regulatory approval.) The Full Self-Driving product is an upgrade over the existing Autopilot program. This service is an add-on that will be made available to Tesla cars. This is consistent with Elon Musk's goal of further increasing Tesla's focus on Artificial Intelligence products and developing self-driving cars rather than producing a wider variety of electric vehicles.

Tesla is using a different technological approach to building self-driving cars than its competitors. For example, Waymo depends on radar, sensors, cameras, etc. while Tesla is relying more on Artificial Intelligence software to enable a car to be self-driving. Waymo estimates that the cost of equipment in each of its cars is about \$40,000 so that the total cost per mile driven of an autonomous vehicle is about three times that of a traditional gas-powered car. Waymo has attempted to keep costs as low as possible by "refitting" existing cars to become self-driving. Amazon's Zoox takes a different approach, building cars from scratch that are intended to be self-driving. At the moment, Zoox has not yet started up its taxi service but expects to do so very soon. One of Tesla's long-range goals is to be able to produce self-driving vehicles at a lower cost than its competitors, which is one reason that it uses different technology than the competition.

Analysts have pointed out that the competition in the electric vehicle market has become more intense in recent years and that Tesla has fallen behind in the production of self-driving vehicles compared with competitors such as Waymo and Pony.ai. In addition, Tesla's unique features are becoming available from other manufacturers. Tesla has placed a great deal of its credibility on the line with the Robotaxi; skeptical investors need to see that they can make it work in order for Tesla to maintain its extremely high valuation. Tesla's market capitalization is \$800 billion as of October 2024, which is far greater than its major automotive competitors such as GM, with a market capitalization of \$51 billion and Mercedes-Benz, with a market capitalization of \$61 billion. Further, Tesla's electric vehicle sales dominate the market, accounting for over 50% of total sales. Future profitability is likely tied to Tesla's ability to continue to innovate and produce high-tech products such as Robotaxis.

Elon Musk remains convinced that the future of the company is based on Robotaxis and that they will be highly successful. He claimed at the Tesla shareholder meeting in June that the technology used to produce autonomous vehicles would increase the company's market value by at least a factor of 10. He claimed in a post on the social media site X (formerly Twitter) that the October 10 "We, Robot" event will be "one for the history books."

Stock analysts believe that the greatest profits will accrue to the companies with the best business models, not necessarily the best technology. There are several companies that are competing directly with Tesla for the self-driving car market, especially the Robotaxi market. Tesla held its “We, Robot” event on October 10, 2024 at Warner Brothers Studios. At this event, Tesla revealed its Robotaxi, which Elon Musk referred to as the Cybercab. The Robotaxi has no steering wheel or pedals but may still require human intervention from time to time. Elon Musk claimed that the Robotaxi would be available to consumers for less than \$30,000. This cost would be further reduced by the possibility that consumers could actually rent them out as Robotaxis and earn income when the cars are not being used. Elon Musk also stated that Robotaxis should be available by 2026 or 2027. Surprisingly, Elon Musk also claimed that the Robotaxis do not need to be plugged in to be charged; instead, they will use a process called “inductive charging.” Elon Musk announced that the company expects to start using autonomous Fully Self-Driving software in both Texas and California next year. The software will be used on the Model 3 and Model Y, an electric SUV. The success of this software will help determine how quickly Robotaxis can be on the road.

Elon Musk surprised the participants with the introduction of a new self-driving vehicle called the “Robovan” which is designed to either carry up to 20 human passengers or else be used to transport goods. This is clearly intended to be used by cities as a cheaper alternative to traditional gas-powered and electric buses. The delivery date for the Robovan was not announced.

The Optimus robot was also introduced; these have a human form and are intended to be used by consumers for routine tasks. They are expected to sell for \$20,000 to \$30,000, putting them within reach of a large percentage of consumers. Elon Musk stated that “The Optimus will walk among you.” to indicate that these robots can be used for any type of household activities that will free up humans for more productive activities. He added “I think this will be the biggest product ever of any kind.”

These products show that Elon Musk is right when he says that Tesla is now a technology company rather than an auto manufacturer. The focus seems to be on developing self-driving cars rather than more inexpensive electric vehicles. The goal is to sell Robotaxis to individual consumers and also sell them to cities as a replacement for traditional taxis. In addition, Elon Musk expects to sell a large number of Optimus robots to individual consumers.

In the aftermath of the We, Robot event Tesla stock fell by nearly 10% over the next two days, while shares of Uber and Lyft increased as the Robotaxi would represent powerful competition for these companies. Analysts were disappointed with many aspects of the presentation. The delivery dates of the Robotaxi, Robovan and Optimus Robots were somewhat vague and subject to revision, so that it was clear that none of these products will add to Tesla’s bottom line for a few years. Elon Musk did not provide some of the key details about the projected selling price and costs of production for the Robovan and did not provide specifics for the Robotaxi service. He did not mention where these vehicles would be produced. In addition, Elon Musk did not provide any evidence of the projected cost per mile for the Robotaxi and

Robovan, which are big selling points to customers. Investors were further disappointed that the company did not announce a low-cost Model 2, which was widely anticipated. This is in line with the company's new focus on developing Artificial Intelligence technology rather than more inexpensive electric vehicles.

In addition to the seemingly optimistic delivery dates, there are many hurdles for Tesla to overcome, such as obtaining necessary permissions from regulators, a history of accidents with Tesla's Autopilot software, technological challenges, etc. Tesla's approach to building self-driving vehicles seems to emphasize lowering costs rather than increasing safety. The company chose to focus on cameras and artificial intelligence and not on Lidar, which uses lasers to measure distances. Elon Musk has admitted to being too optimistic with delivery dates in the past.

Although stock market analysts were disappointed by the "We, Robot" event, in the long run Tesla's massive investments in Artificial Intelligence and increased focus on this technology, combined with Tesla's widespread name recognition is likely to result in successful launches of its self-driving vehicles and robots, potentially fueling massive gains in its stock price in the future.

Works Cited

- “What is Artificial Intelligence? Definition, Uses and More.” University of Cincinnati Online. <https://online.uc.edu/blog/what-is-artificial-intelligence/>
September 23, 2024. “24 Cutting Edge Artificial Intelligence Applications in 2024.”
Simplilearn.com. <https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/artificial-intelligence-applications>
- “What are AI Applications” Google Cloud. <https://cloud.google.com/discover/ai-applications>
- “From Alexa to ‘Thursday Night Football’, Amazon is Using Generative AI as a Powerful Tool to Innovate.” About Amazon.com. <https://www.aboutamazon.com/news/innovation-at-amazon/how-amazon-uses-generative-ai>
- January 23, 2024. “The Six Levels of Autonomous Driving Explained.” C & T Solution Inc. https://www.candtolution.com/news_events-detail/the-six-level-of-autonomous-driving-explained/#:~:text=Waymo%3A%20Waymo%20is%20one%20of,Angeles%20County%20and%20Austin%2C%20Texas
- Rahman, Shamim. December 26, 2023. “AI in Autonomous Vehicles: the Road to Self-Driving Cars.” Infodot.medium.com. AI in Autonomous Vehicles: The Road to Self-Driving Cars | by Shamim Rahman | Medium
- “Self-Driving Car”. Wikipedia.com. Self-driving car - Wikipedia
- March 23, 2022. “Artificial Intelligence in Self-Driving Cars”. Market-prospects.com. Artificial Intelligence in Self-driving Cars | Market Prospects
- Mahendra, Sansshep. June 14, 2023. “Autonomous Cars: How do Self-Driving Cars Actually Work?” aplusinfo.com. Autonomous Cars: How do Self-Driving Cars Actually Work? - Artificial Intelligence
- Duggal, Nikita. September 9, 2024. “Incredible Advantages of AI/Notable 23 Benefits of AI”. Simplearn.com. Incredible Advantages of AI | Notable 23 Benefits of AI (simplilearn.com)
- January 6, 2023. “Autonomous Driving’s Future: Convenient and Connected.” Mckinsey.com. <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/autonomous-drivings-future-convenient-and-connected>
- Yada, Vaishnavi. February 15, 2023. “How Tesla is Using AI and Big Data Analytics in their Self-Driving Cars.” Medium.com. How Tesla is using AI and Big Data Analytics in their Self Driving Cars? | by Vaishnavi Yada (Amira) | Dare To Be Better | Medium
- Shah, Agam. March 8, 2023. “How Tesla Uses and Improves its AI for Autonomous Driving.” AIWire.net. How Tesla Uses and Improves Its AI for Autonomous Driving (aiwire.net)
- Knight, Will. September 7, 2021. “Why Tesla is Designing Chips to Train its Self-Driving Tech” Wired.com. https://www.wired.com/story/why-tesla-designing-chips-train-self-driving-tech/?gad_sou

- rce=1&gclid=Cj0KCQjw3vO3BhCqARIsAEWblcBWwfDR-pIO5FjEFyq3KNU-SUnnr1j0XCJeqRz2y3ewFBbTPtgWBHlaAnsiEALw_wcB
- Browne, Ryan. September 5, 2024. "Tesla Says it Will Launch Full Self-Driving Product In Europe and China Early Next Year." <https://www.cnn.com/2024/09/05/tesla-to-launch-full-self-driving-in-europe-and-china-in-q1-2025.html?&qsearchterm=tesla%20ai%20self-driving%20cars>
- Castellano, Robert. August 13, 2024. "Tesla: Brilliant AI-Driven Robotaxi Model, But Not Today." Seekingalpha.com. <https://seekingalpha.com/article/4714088-tesla-brilliant-ai-driven-robotaxi-model-but-not-today>
- Root, Al. October 4, 2024. "Tesla Robotaxi Day is a Make or Break Moment for Elon Musk." Barrons.com. <https://www.barrons.com/articles/tesla-stock-robotaxi-day-elon-musk-ec7050c2>
- May 21, 2024. "Why are EV Sales Slowing?" GoldmanSachs.com. <https://www.goldmansachs.com/insights/articles/why-are-ev-sales-slowing>
- Gitlin, Jonathan. August 6, 2024. "Elon Musk Says Tesla is an AI Company Now. Here's How Plausible That Is." <https://arstechnica.com/cars/2024/08/elon-musk-says-tesla-is-an-ai-company-now-heres-how-plausible-that-is/#:~:text=%22We%20are%20an%20AI%2C%20robotics,investors%20and%20journalists%20in%20April.>
- Westbrook, Justin. September 13, 2024. "Tesla's Robotaxi: Everything we Know About the Self-Driving Car's Reveal" <https://www.motortrend.com/news/tesla-robotaxi-first-look-review/>
- Levin, Tim. September 15, 2024. "Tesla Robotaxi: Everything We Know." Insideevs.com. <https://insideevs.com/reviews/722798/tesla-robotaxi-cybercab-explainer/>
- Templeton, Brad. April 18, 2024. "Tesla, Waymo, Nuro, Zoox and Many Others Embrace New AI to Drive." <https://www.forbes.com/sites/bradtempleton/2024/04/18/tesla-waymo-nuro-zoox-and-many-others-embrace-new-ai-to-drive/>
- Gratton, Peter. August 13, 2024. "What are Tesla's Main Competitors?" Investopedia.com. <https://www.investopedia.com/ask/answers/120314/who-are-teslas-tsla-main-competitors.asp>
- Betz, Sunny. Updated by Matthew Urwin. March 11, 2024. "28 Self-Driving Car Companies You Should Know." Builtin.com. <https://builtin.com/articles/self-driving-car-companies#:~:text=Cruise%20is%20a%20driverless%20rideshare%20company.%20The,objects%20around%20them%20and%20record%20test%20drives.>
- Fujita, Akiko. October 5, 2024. "Can Robotaxis Turn a Profit? Experts Skeptical as Tesla Prepares to Unveil Plans." finance.yahoo.com.

<https://finance.yahoo.com/news/can-robotaxis-turn-a-profit-experts-skeptical-as-tesla-prepares-to-unveil-plans-123012750.html>

“Tesla’s AI Strategy: A Comprehensive Overview.” Perplexity.ai.
<https://www.perplexity.ai/page/tesla-s-ai-strategy-a-comprehe-cPPS1ydpRxKtSndmn9.w3g>

April 14, 2019. “The 6 Levels of vehicle Autonomy Explained.” Synopsys.com.
<https://www.synopsys.com/blogs/chip-design/autonomous-driving-levels.html>

September 12, 2024. “5 Top Self-Driving Car Manufacturers.”
<https://yoshalawfirm.com/blog/5-top-self-driving-car-manufacturers/#:~:text=1.,under%20the%20umbrella%20of%20Google.>

October 10, 2024. “Tesla Reveals 20 Cybercabs at We, Robot Event, Says You’ll be Able to Buy One for Less Than \$30,000. Techcrunch.com.
https://techcrunch.com/2024/10/10/tesla-reveals-20-cybercabs-at-we-robot-event/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAFz05MkLr8nwDfDQkLNLiezbbNR6lnbwmM1TbiuGjuo5t8YkI88VejI9tctj8_9KsDf2yl0TsMQDBE9657ckaOVULKbC3rrm2pEJZ69htXj9kHglllUJ_dv4R_Wc_qejCZwllfZXDv4qbHfcH7WGv8ifvj6C0T9qZStZye2j93uu

October 11, 2024. “Tesla CEO Elon Musk Unveils ‘Cybercab’ Robotaxi. Reuters.com.
<https://www.reuters.com/technology/teslas-musk-unveil-robotaxis-amid-fanfare-skepticism-2024-10-10/>

October 10, 2024. “Tesla’s Optimus Bot Makes a Scene at the Robotaxi Event.” Theverge.com.
<https://www.theverge.com/2024/10/10/24267225/tesla-robotaxi-optimus-we-robot>

October 11, 2024. “Tesla’s Robotaxi Event is ‘Stunningly Absent on Detail’, Sending Stock Lower
<https://www.morningstar.com/news/marketwatch/20241011245/teslas-robotaxi-event-is-stunningly-absent-on-detail-sending-stock-lower>

October 11, 2024. “Why Uber Stock Hit an All-Time High After Tesla’s Robotaxi Rollout.” Investopedia.com.
<https://www.investopedia.com/uber-stock-record-high-tesla-robotaxi-event-8727086>

October 11, 2024. “Elon Musk Unveils Tesla’s ‘Cybercab,’ Plans to Bring Autonomous Driving Tech to Other Models in 2025.”
<https://apnews.com/article/elon-musk-tesla-robotaxi-unveiling-a00d063f2ffc67125889a6635a0a607e>

"Advancements in Artificial Organs: Self-Sensing Actuators for Enhanced Functionality"

By Keerthi Chandran¹, Plinio Rodrigues de Oliveira Zanini²

¹Indus International School, Bangalore, India; Keerthichandran75@gmail.com; ORCID: 0009-0002-4904-3215

²4D Printing & Biomimetics (4DB) Research Group, Centro de Engenharia, Modelam e Ciências Sociais Aplicadas (CECS), Universidade Federal do ABC (UFABC), Av. Dos Estados, 5001 – Bangú, Santo André, SP 09210-580, Brazil; pliniozanini@hotmail.com; ORCID: 0000-0001-5142-2674

Abstract

The demand for organ transplants has increased over the years regardless of new medical innovations. The number of donors isn't at par with the demand for organs. The waiting list for organs can range from 4 months to 5 years. Emerging technologies such as artificial organs and self-sensing actuators can be used to create responsive artificial organs. Complex functions and biocompatibility issues are some of the difficulties with creating artificial organs. There has been recent research on membrane technology, biomaterials, integration of various functionalities in one mechanism, self-sensing, biomimetic and actuator integration, etc. Using a self-sensing approach will reduce the complexity of the device and allow for live monitoring. We aim to find ways to simplify complex issues to speed up the introduction of artificial organs. The data was collected in the form of a literature review. Search terms such as "pumping mechanisms", "piezoelectric", and "gel actuators" were used. From this secondary data, we found that flexible/elastic electrostatic actuators could be used to create self-sensing (responsive to external stimuli) pumping mechanisms. These pumps can be used in hearts, the urinary bladder, stomach, intestines and more. We propose a design of self-sensing pumps for use in dialysis. These devices can be implantable or worn externally. Through this paper, we aim to replicate complex pumping functions that require interactions with other arteries, veins or organs. To conclude, there is an immense potential for self-sensing actuators and pumping in the process of designing artificial organs.

Keywords Self sensing, artificial organs, pumping, biocompatibility, actuators

Statements and Declarations There are no financial or non-financial interests that are directly or indirectly related to the work

Fields of Research Artificial organs, Dialysis, Pumping, Biocompatibility

1 Intro

In 2022, 157,494 transplants took place worldwide. In India, more than 15,000 transplants took place [1] and the demand for organs exceeded the donation rate. We must increase the donations from 0.8 per million population to 65 per million population if we are to meet the growing needs of the nation. This, however, may not be practical as there are several ethical, cultural and economic factors that prevent individuals from comfortably surrendering their organs.

This is where artificial organs can come in to save the day. Artificial organs are also known as extracorporeal devices and they are devices or tissue that can be integrated or implanted with existing tissues to replicate vital organ functions [3, 4]. Artificial organs are highly beneficial for those with disabilities to overcome difficulties and live a close to normal life. Artificial organs come under the larger field of biomimetics, which is a field that mimics biological processes. Artificial organs will make use of sensors, pumps, artificial tissues and more to work with natural organs to perform bodily functions. They can be used to pump blood, filter waste and excess fluid, restore mobility, and more. There are 3 broad categories for artificial organs. Mechanical artificial organs are made exclusively of inanimate polymers such as plastics and metals; biomechanical organs involve both living materials such as cells and inanimate materials; biological or bioartificial organs can be made of living cells and biodegradable polymers [4, 5]. However, there are several limitations currently. Didarul Bhuiyan, a biomaterial and tissue engineering scientist said that it would take another 20-30 years for us to see this technology as a reality. Durability, biocompatibility, exact function replication, interconnectivity with natural organs and availability of materials are some of the reasons for the setback [4, 5]. The current manufacturing cost of artificial organs is high but in the long run the cost will decrease significantly. Scientists are looking into 3D printing artificial organs using synthetic muscles and using nanotechnology to improve biocompatibility and real-time monitoring of the organ function. These technologies will address the challenges mentioned above [6].

Artificial organs are typically made from biocompatible materials and are engineered to replicate natural functions such as pumping, neuron firing, hormone regulations, etc. Several organs in our body make use of pumping [3, 7]. Hearts pump blood throughout our body to ensure that all the other organs receive sufficient oxygen. Lungs facilitate the exchange of gasses while breathing and create a pumping rhythm. Digestion makes use of pumping to mix the food with digestive juices. Bladders contract when a certain volume of urine has been collected. When creating artificial organs, it is essential to replicate this pumping as it has a prominent role in the human body [8, 9]. Pumping mechanisms can often be bulky and require multiple components to coordinate functions with the natural organs [10, 11]. Self-sensing pumps will make use of biocompatible materials rather than metals and therefore will be more suitable for use in artificial organs. They are able to sense changes in pressure/ temperature and alter their pumping speeds/ quantities accordingly. Self-sensing pumps do not require external sensors and are more compact [12–14]. This paper aims to add to the existing body of knowledge and suggest potential implications of self-sensing pumps in artificial organs.

2 Methods

A literature review approach was used in the development of this paper. The reviewed papers were selected using a lenient systematic process. Papers were filtered based on their dates, titles, relevance to this study, depth of research and more.

The initial search started with minimal restrictions and broad keywords such as “artificial organs”, “past 5 years”, “actuators”, “pumping mechanisms”, “piezoelectric”, “gel actuators”, “potential uses of artificial organs”, “biomimetics” and “soft robotics”. The papers were selected based on their titles. Relevant topics with potential in this context were explored in greater detail. Papers were sorted based on the information provided in the abstract. Sources cited by these papers were also extremely important in the data collection process. Papers on proposed extensions helped understand the current scope of this field. From the broader topics, frequently mentioned concepts were explored. This led to research of the role of self-sensing in specific organs. Search terms like “heart” or “kidneys” were to understand the specific details that would need to be incorporated for each type of artificial organ. Individual organs and their primary functions were also explored. From this, the prominence of a pumping mechanism in the working of the organ was realized. The search into different types of pumping mechanisms, their actuators, size etc. began here. Based on all the gaps and technologies specific research was conducted on sensors, pumps and actuators.

The discussion section of the paper was shaped by exploring the interdependent relationship between natural and artificial organs and how this could be realized through the integration of sensors and pumps. This led to exploration in the self-sensing realm. Key words like “self-sensing” led to mixed results. Therefore, terminology like “wearable technology pumps”, “pumps with inbuilt sensors”, and “analog circuits” were used to narrow down the results.

The paper was approached in an interdisciplinary manner. Therefore, the economic and social implications of such technologies were relevant during data collection. When looking at the social implications, secondary research on the number of deaths associated due to organ donation shortages, the number of patients that need organ transplants, the number of willing donors and the success rate of such procedures was carried out. For the economic aspect, cost, market availability, current technological and financial limitations were explored.

This paper is not only a literature review. It will include a conceptual proposal for a wearable dialysis machine. The proposal will be made based on the gaps and limitations that have been identified related to both pumping mechanisms, actuators and artificial organs. The proposal aims to add to the existing body of work in the artificial organ field and help speed up their introduction to the medical world.

3 Literature Review

3.1 Artificial organs

Membrane technology plays a pivotal role in various medical applications ranging from drug delivery to artificial organs and tissue engineering. The main components in membrane technology include biocompatibility, membrane structure, and materials used for the membrane development. The biomaterials used in the medical membranes must exhibit biocompatibility in order to prevent organ rejection. In some cases, biodegradability is important as well.

Biocompatibility is the ability to interact favorably with biological systems avoiding adverse responses such as inflammation and immune rejections. Modifiers are often used to increase biocompatibility for an individual. The biocompatibility will vary based on blood type, immune strength, tissues exposed to it, etc. Therefore, artificial organs will have to be individually tailored and the material used and the ratios in which they are used will not be universal. Various categories of biocompatible polymers can be used for artificial organs. Inert, interactive viable and replant biomaterials are some examples. Blood group compatibility is one of the main concerns as compatibility will ensure necessitating the membranes with reduced coagulation and protein adsorption, which is essential for the organ to succeed in the host [7, 15].

Membrane structure significantly impacts performance in blood containing devices. Pore size, porosity, and surface properties influence factors like flux decline and selectivity. Flux decline refers to the gradual reduction in rate of fluid/ blood transport across the membrane across time. Flux decline is relevant in the context of membrane processes such as filtration/ osmosis and can occur due to fouling (accumulation of particles microorganisms on the surface or within the pores of the membrane), concentration polarization (buildup of solutes/ blood near the surface during filtration, a gradient develops across the membrane thus reducing the driving force of the flux), membrane degradation (membrane materials can decompose due to exposure to harsh environments), chemical attacks or mechanical stress. This can change the membrane's structure itself and cause flux decline [16].

Synthetic and natural materials are often employed in biomedical membranes. Non-degradable polymers like polyethylene terephthalate (PET) and polytetrafluoroethylene (PTFE) are used alongside bioresorbable polymers (polymers that provide temporary assistance and naturally degrade in the body without leaving any residue) such as polylactide and polyglycolide. Material selection is extremely important considering biocompatibility and flux decline of the organ. Using artificial tissues will enable cell growth replicating a natural organ almost perfectly. Artificial tissues also enable the organ to regenerate and reduce the impacts of degradation and therefore prevent the extent of flux decline. The use of an individual's cells will improve the biocompatibility as well [3, 17, 18].

Soft robotics technologies have a major role to play in the recent advancements in the biomedical field, particularly in terms of artificial organs. The integration of soft robotics principles facilitated the accurate replication (through biomimetics) of organ physiology, especially for organs with natural motion driven by muscle contractions. Soft robotics technologies have been pivotal in cardiovascular therapy. By embedding artificial muscles within elastomeric matrices, researchers have successfully mimicked the motion of human organs such as the left ventricle, enabling real time, dynamic interaction between natural tissues and the artificial organ. This makes the artificial organ adaptable to different situations making it more adept to provide cardiocirculatory support. However, there are still issues which persist due to the design. The longevity of the implanted devices isn't as long as we would've hoped. The need for constant surgeries to maintain the organ is not practical and can pose more risks for the individual [5, 19].

Artificial organs can be combined with soft robotics to create wearable and implantable devices. These devices aim to enhance mobility and functionality by creating a cognitive symbiosis. A HMI (Human Machine Interface) is essential to do so. Current wearable robotics feature primary rigid structures which can be combated by using artificial tissues and muscles. Using biomimetic prosthetics will create a more comfortable experience and will improve the ease with which the organs, brain and body function [19].

Artificial organs constitute a significant segment of the medical device market with substantial economic implications. The artificial organ market is projected to reach billions of dollars in the next few years. The need for organ transplants is increasing and so are all the costs associated with the procedure due to the limited number of donors worldwide. Artificial organs will initially be available for the upper classes due to the costs associated with the personalization of the organ but will soon take over the market, as real trials develop our understanding of the scope of this wearable technology. Artificial organs will be most suitable for the aging population based on current research. However, they have a fixed income and money can often be tight. As a result, government subsidies for organ replacement procedures may have to become a reality. Artificial organs will require maintenance and frequent checkups in the first few months or years. The costs associated with this again increase the financial pressure from the procedure [20].

3.2 Pumping Mechanisms

Various pumping mechanisms are used in the medical field. Syringe pumps are used to deliver precise amounts of fluids, diaphragm pumps operate by flexing the diaphragm to create suction discharge fluid, and rotary pumps use rotating elements to create fluid flow. Artificial hearts will be discussed in detail in this section but this information can be extrapolated to other artificial organs that require fluid motion [10, 11].

A pump essentially creates a pressure difference to move fluid from one place to another. Pumps typically require an external energy source and begin drawing fluid from a source (the liquid would be blood in the context of a heart). This liquid will then increase the pressure in a particular chamber (e.g. right ventricle) and this creates motion or energy. For a heart, the energy will be converted into a force that pushes out the deoxygenated blood. Consequently, hemocompatible materials will have to be used to prevent unwanted reactions and alterations in blood makeup. Hemocompatible materials should be non-thrombogenic (don't trigger blood clot formation); biocompatible; have low platelet adhesion; and low hemolysis (ability to rupture red blood cells) [21].

Current pumping mechanisms have issues such as weak pulsation and easy suction and reflux. This increases the need for maintenance on the pumps and therefore puts the patient at risk of infection. An adjusting feedback monitor can be used to respond to the input pressure. A sensor and sliding mode controller are used together to control the 2nd aortic valve's pressure. As a result, pulsatility improves. This inspired the idea of a self-sensing pump for use in artificial

organs. The mechanical pump will be created using bio materials that can be given the ability to sense changes in pressure, temperature, and more by actuators.

Flow balance between pumps or sections need to be coordinated. If one re-adjusts because of the input, the second should follow suit and maximize efficiency. In the heart, the left and right pumps need to be coordinated to prevent pulmonary congestion. Issues like this can arise between artificial organs and natural organs. This shows the importance of the ability to self-sense and respond appropriately.

The force prediction and manipulation of a pump are important. Knowing the average energy needed to sustain the pump will help create a more efficient power source that will reduce the need for frequent surgeries or device updates. The pump should have a frequency range from minimal activity (rest) to extreme activity or stress (stroke). The positioning and alignment of the pump will impact the blood pressure and rate of blood exchange. The load bearing capacity will differ for individuals. Therefore, the pumps will have to be tailored for different weight, age and lifestyles. A hydrodynamic bearing can be used to levitate the piston and increase reliability and hemocompatible operation.

The power consumption can be predicted using the axial forces and torque created by the piston (diaphragm pump). Electromagnetic drives will be present for 1 pump cycle, resistance and amplitude of force currents will have to be accounted for as well [8, 22–24].

3.3 Self- sensing actuators

Actuators are devices that convert various forms of energy into mechanical motion or force to initiate a task or function. Pumps can be considered as specific types of actuators that focus on fluid movement. Actuators (especially in the context of artificial organs) often use pumps to provide sufficient fluid pressure for the functions. Actuators enable controlled motion or force application in systems. This when paired with or made with self-sensing materials will create a self-sensing actuator [25].

There are several types of actuators. Electrical actuators [26] (EA) use electrical energy as an input and can be used for controlling pumps for drug delivery, as actuating valves or power prosthetic limbs. Hydraulic actuators [27] (HA) use pressure to create fluid motion. HA can be used in kidneys, bladders, hearts, etc. Pneumatic actuators [28, 29] (PA) use compressed gas. PA will be difficult to implement for internal organs but can be used for prosthetics due to its ability to provide lightweight and responsive movements. PA are difficult to implement in internal organs because it is difficult to make pressurized fluids biocompatible. They may cause infections and may react with tissues. The stable power source required will not be implantable as it is a safety hazard. Thermal actuators [30, 31] (TA) use heat and temperature. This can be used to trigger movement at different speeds based on the heat of other organs such as the heart. For example, when exercising, the body temperature increases. A TA in a valve can use this heat energy to synchronize its pumping. This could backfire as well. A fever can cause average body temperature to increase. A faster heart beat and pumping can create issues with an individual's immunity. It could cause hypertension and exhaustion as well. Mechanical actuators [32, 33]

(MA) rely on external mechanical components. They tend to be part of larger machinery and don't have many uses in implantable devices.

Self-sensing materials are materials that are capable of detecting changes in their own properties. The properties can be strain, stress, temperature or other physical parameters. They can detect these changes without the use of sensors making them ideal for use in artificial organs. These materials can be used with other biomimetic materials to create an artificial organ that is responsive to changes in the body's physical conditions without the need for external sensors and manual inputs [12, 34].

Capacitance [35, 36] is the measure of a system's ability to store electrical charge per unit voltage. Capacitance based self-sensing measures the deformation or displacement of a material or structure. The capacitance of the material or structure is dependent on its geometry or position. As the material deforms, its capacity changes accordingly. This method is employed in dielectric elastomers. This type of sensing offers simplicity and non-intrusiveness which is essential for artificial organs.

Resistance [37–39] is a measure of the opposition to the flow of electric current in a circuit. Resistance based self-sensing is a technique that measures the change in resistance of a specific material or structure. This can be used to infer its deformation or displacement. Changes in electrical resistance or general resistance of a particular organ can help doctors identify organ rejections or issues. This will help the patient get immediate attention and will make the artificial organ implant procedure smoother and safer.

Dielectric elastomers [12, 40, 41] exhibit large deformation in response to an electric field. They are made of thin elastomeric membranes between electrodes. They have fast response time giving them good potential for use in organs. The working of the electrodes in an internal system is out of the scope of this paper. These elastomers can sense capacitance, voltage and impedance.

Memory shape polymers [42–44] possess the ability to change shape in response to external stimuli and return to the original shape afterwards. The stimuli could be temperature or pressure. They can be used in stents, valves, tissue scaffolds, drug delivery systems, artificial muscles and more. These materials can be alloys, synthetic polymers, hydrogels, ceramics and elastomers. These polymers can self-sense thermal and electrical stimuli.

Multi-responsive hydrogels [30, 45, 46] exhibit sensitivity to external stimuli such as temperature, pH, light, electric fields and even some types of ions. These types of hydrogels can undergo changes in structure, volume and mechanical properties. An artificial organ made with these hydrogels will be extremely adaptive and versatile. They can be used for controlled release systems where a particular change will trigger the release of a certain output. These hydrogels can self-sense mechanical, volume and chemical stimuli.

Self-sensing actuators will have fewer moving parts. Therefore, they will be safer and more predictable. They have the potential to be cheaper in the long run and as a result, artificial organs will become accessible to a larger section of society. These devices will be more responsive to external stimuli and therefore less invasive. Development of a position self-sensing

method for motion actuators utilized a dielectric elastomer membrane. It used voltage and current to estimate resistance and capacitance. The high frequency and low amplitude voltage component served as a good actuation signal for parameter estimation. Dielectric elastomers enable controllable deformation with high energy density, lower power consumption and fast dynamic response. They act as both actuators and sensors. The electrical measurements during actuation allow for the reconstruction of material deformation enabling its use in feedback control algorithms [12, 45]. Extrapolating this, it is possible to connect the feedback control to considering the inputs from various organs and altering outputs to keep the body and its internal organs working as a team. This technology will eliminate the need for large databases to store frequent movements or positions. Voltage and data are typically available from the same circuit used for actuation.

Similar technology was used in human collaborative robots. Safety measures were emphasized through tactile sensors. It was proposed that these sensors would be replaced by self-capacitance sensors and shock absorbing structures. Data indicates that the increase in shock ability came with an increase in the thickness of the self-sensing material regardless of the decrease in contact detection sensitivity. This technology can be used for more life-like prosthetics [35].

3.3 Anti-thrombogenicity

Artificial organs mostly will involve the passage of blood through their structures. To ensure that these devices work optimally, thrombogenicity has to be addressed. Thrombogenicity refers to the tendency of a material in contact with the blood to produce thrombus or clots. It is a significant challenge that can compromise the success of blood-containing medical devices like artificial organs and the dialysis machine proposed in the discussion section [54].

Device-initiated thrombus formation has local and systemic consequences. Some of them are: device failure requiring clinical reintervention, thrombi detaching and traveling through the vasculature (potentially causing life threatening obstructions in critical tissues and organs) [55].

A bioinspired strategy was proposed to construct an artificial vascular endothelium (AVEC) to block thrombo inflammation. The design aims to mimic the physiological structure, mechanical character, and biophysical function of natural endothelium. The cobblestone-like surface exhibited superhydrophilicity and superlubricity. These properties enabled a blood-unperturbed bioinert feature characterized by total anti-adhesion of blood components, avoidance of complement system and platelet activation, classical polarization of macrophages, inhibition of thrombo inflammatory reactions. The properties were tested on extracorporeal circulation lines in hemodialysis which is why this technology could complement the design proposed in the discussion [54].

Another technology development, Tethered-liquid Perfluorocarbon (TLP) coatings are a class of lubricant-induced surface coatings aimed at reducing adverse reactions in implanted medical devices, such as blood clot formation (thrombosis) [56]. TLP coatings are created using a vapor phase silanization reaction and the self-assembling properties of a fluorinated silane to

form tethered perfluorocarbon (TP) layers. This vapor phase method offers advantages over the liquid phase deposition method (LPD) as it does not require control of humidity conditions. TLP coatings infused with more viscous lubricant effectively reduce fibrin adhesion from human whole blood [56].

Poly(2-methoxyethylacrylate) (PMEA) is a new coating material that aims to reduce protein and platelet adsorption. The study concluded that PMEA-coated circuits maintain anti-thrombogenicity during use. PMEA-coated and heparin-coated circuits show similar hemocompatibility as well [57].

Overall, there are several upcoming and existing biomaterials with varying topographical properties and surfaces that incorporate anti-thrombogenic substances that have shown potential in reducing thrombosis [55]. However, clinical results are needed to fully demonstrate long-term benefits of these solutions [55].

4 Discussion

Dialysis is a medical procedure that assists in removing waste products and excess fluid from the blood when there are issues with the kidneys. Chronic kidney disease, kidney injuries, end-stage renal disease, and even diabetes can lead to the need for dialysis. This process involves special machines or solutions to filter the blood, thus mimicking the function of the kidney. Blood is pumped out of the body through a dialysis machine where it is filtered before being returned or the body's natural filtering membrane in the abdomen is used with no external pumping. Maintaining the proper balance of minerals and fluids in the body is essential to staying healthy.

There are several issues with the current dialysis process, however. The strict treatment requires changes in diet and the consumption of medications regularly. Dialysis can be done at home or in-center depending on the patient and severity of the case. In-center usually happens 3 times a week for 3-5 hours while at home dialysis happens daily for 2 hours [47, 48]. Either way, it is a huge responsibility and can cause stress. A dialysis machine typically weighs around 70.5 lbs. (32 kg). Most patients find it difficult to spend so much time helping their body do something that they weren't even aware of. This can cause mental health issues, sleep disorders and overall discomfort. A portable or implantable dialysis machine will enable patients to live their lives and undergo dialysis while doing whatever is meaningful to them.

Based on the research above, we propose an electrostatic actuator pump-based dialysis mechanism that can be externally worn or implanted based on the severity of the kidney issues. There are some existing electrostatic pumps (dielectric elastomer based) which replicate a pumping mechanism. The Peano-HASEL [49] (hydraulically amplified self-healing electrostatic) pump is a novel pump that uses electrostatic forces to generate fluid movement. It is a muscle mimetic electro-hydraulic transducer that linearly contracts on activation. This can be used to replicate pumping mechanisms for dialysis or even the filtering process. Dielectric elastomer actuators (DEAs) offer muscle like actuations but require rigid frames and complex configurations for effective performance [49–51]. The Peano-HASEL doesn't require complex

configuration as it combines fluidic and electrostatic principles to linearly contract. These actuators demonstrate controllable linear contraction up to 10%. This means that it can perform precise pumping patterns and change the pumping pressure easily. It can take a strain of 900% per second which will prevent failure in extreme cases. The actuation frequency is 50 Hz but this part is out of the scope of this paper. The elastomer can lift more than 200 times its weight and hence can be miniaturized easily for use in dialysis. The zipping pump or electro-pneumatic pump can be used to obtain high pressure and to improve fluid flow rate. These pumps are lightweight and hence are suitable for wearable technology. This pump uses dielectric-liquid amplified zipping mechanisms for air volume and pressure control. It is a low-power soft pump and can feature different insulated electrodes. Therefore, we can manipulate the power capabilities for implantable and externally worn devices. The shape of the zipping region significantly impacts the pumping performance and must be kept in mind while designing the dialysis pump [52, 53].

Flexible electrostatic pumps can be created using fluids, as exemplified by the examples mentioned above. This versatility makes it possible to adapt to challenges faced when replicating the dialysis process. It is important to consider the energy density and other metrics that contribute to power generated by the pump.

These pumps would also allow self-sensing from resistance and capacitance [37]. Nowadays, it is common for there to be issues with the high voltage to power these devices, but it is out of the scope of this conceptual design proposal. If a filtering mechanism can be miniaturized and automated to coordinate with the natural organs, an artificial kidney can be proposed. Even if the self-sensing flexible electrostatic pumps cannot perfectly replace traditional dialysis, they could reduce the frequency of hospital trips required thus immediately improving the patient's quality of life.

5 Conclusion

In this paper, we reviewed technologies such as artificial organs, actuators, dielectric elastomers, hydrogels, etc. Concepts such as biocompatibility, capacitance, resistance, pumping, and functionalities of artificial organs were explored as well. The economic considerations were considered briefly as well.

We initially looked into the relevance of artificial organs in society today. Organ donations and the ethical and cultural factors associated with it were explored in brief. From that, the paper split into two broad categories: functionality and compatibility of artificial organs.

Membrane technologies, bio-materials, organ rejections, and more were explored under compatibility. Functionality of the artificial organ consisted of various existing technologies and concepts. For example, actuators, self-sensing and pumping. This was extrapolated in the discussion section where we proposed a conceptual design based on the technologies reviewed for dialysis. The technology proposed can be both an external and implantable device that will improve the immediate quality of life for various individuals.

To conclude, this paper has delved into artificial organs and their current progress. It has highlighted the importance of self-sensing in the context of artificial organs and finally proposed a possible implication of self-sensing pumps for use in dialysis. By fostering continued innovation and collaboration, we can develop the biomedical engineering sector majorly.

Conflict of interest

The authors did not receive support from any organization for the submitted work. The authors have no relevant financial or non-financial interests to declare.

Contributions

Conceptualization: [Keerthi Chandran and Plinio Zanini]; Methodology: [Keerthi Chandran]; Formal Analysis and investigation: [Keerthi Chandran]; Writing- original draft preparation: [Keerthi Chandran]; Writing review and editing: [Keerthi Chandran and Plinio Zanini]; Supervision: [Plinio Zanini]

Works Cited

1. "Organ Shortage Continues to Cost Lives." *The Hindu*, 2024, <https://www.thehindu.com/sci-tech/health/indias-poor-organ-donation-record-continues-to-cost-lives/article67161978.ece>. Accessed 16 Feb. 2024.
2. "Nearly 5 Lakh People Die a Year Due to Organ Scarcity." *India Today*, 2024, <https://www.indiatoday.in/india/story/nearly-5-lakh-people-die-a-year-due-to-organ-scarcity-288499-2015-08-15>. Accessed 26 Apr. 2024.
3. Zrinscak, Debora, et al. "Soft Robotics for Physical Simulators, Artificial Organs, and Implantable Assistive Devices." *Progress in Biomedical Engineering*, vol. 5, 2023, IOP Publishing, 012002, <https://doi.org/10.1088/2516-1091/acb57a>.
4. "Artificial Organ - an Overview." *ScienceDirect*, 2024, <https://www.sciencedirect.com/topics/materials-science/artificial-organ>. Accessed 16 Feb. 2024.
5. Zrinscak, Debora, et al. "Soft Robotics for Physical Simulators, Artificial Organs, and Implantable Assistive Devices." *IOPscience*, 2023.
6. Cascalho, Marilia, and Jeffrey L. Platt. "The Future of Organ Replacement—Needs, Potential Applications, and Obstacles to Application." *Transplantation Proceedings*, vol. 38, 2006, pp. 362–364, <https://doi.org/10.1016/j.transproceed.2005.12.055>.
7. Stamatialis, Dimitrios F., et al. "Medical Applications of Membranes: Drug Delivery, Artificial Organs, and Tissue Engineering." *Journal of Membrane Science*, vol. 308, 2008, pp. 1–34, <https://doi.org/10.1016/j.memsci.2007.09.059>.
8. Bierewirtz, Tim, et al. "A Novel Pumping Principle for a Total Artificial Heart." *IEEE Transactions on Biomedical Engineering*, vol. 71, 2024, pp. 446–455, <https://doi.org/10.1109/TBME.2023.3306888>.
9. Liu, Xin, et al. "A Physiological Control Method Based on SMC and GAPSO for Artificial Heart Pumps to Maintain Pulsatility and Avoid Regurgitation and Suction." *Journal of Medical and Biological Engineering*, vol. 43, 2023, pp. 42–52, <https://doi.org/10.1007/s40846-022-00767-8>.
10. Bhandari, D. S., and Dharmendra Tripathi. "Alteration in Membrane-Based Pumping Flow with Rheological Behaviour: A Mathematical Model." *Computer Methods and Programs in Biomedicine*, vol. 229, 2023, 107325, <https://doi.org/10.1016/j.cmpb.2022.107325>.
11. Moorthi, P. V. P., et al. "Mechanistic Origins of Concrete Pumping: A Comprehensive Outlook and Way Forward." *Magazine of Concrete Research*, vol. 75, 2023, ICE Publishing, pp. 353–366, <https://doi.org/10.1680/jmacr.21.00278>.
12. Rizzello, Gianluca, et al. "Towards Sensorless Soft Robotics: Self-Sensing Stiffness Control of Dielectric Elastomer Actuators." *IEEE Transactions on Robotics*, vol. 36, 2020, pp. 174–188, <https://doi.org/10.1109/TRO.2019.2944592>.
13. Rizzello, Gianluca, et al. "A Self-Sensing Approach for Dielectric Elastomer Actuators Based on Online Estimation Algorithms." *IEEE/ASME Transactions on Mechatronics*, vol. 22, 2017, pp. 728–738, <https://doi.org/10.1109/TMECH.2016.2638638>.

14. "A Self-Sensing Approach for Dielectric Elastomer Actuators Based on Online Estimation Algorithms." *IEEE Journals & Magazine | IEEE Xplore*, 2024, <https://ieeexplore.ieee.org/abstract/document/7781641>. Accessed 28 Feb. 2024.
15. "Improving Biocompatibility for Next Generation of Metallic Implants." *ScienceDirect*, 2024, <https://www.sciencedirect.com/science/article/pii/S0079642522001347>. Accessed 26 Apr. 2024.
16. "Membranes | Free Full-Text | The Roles of Membrane Technology in Artificial Organs: Current Challenges and Perspectives." *MDPI*, 2024, <https://www.mdpi.com/2077-0375/11/4/239>. Accessed 13 May 2024.
17. "Advanced Soft Materials, Sensor Integrations, and Applications of Wearable Flexible Hybrid Electronics in Healthcare, Energy, and Environment." *Wiley Online Library*, 2024, <https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.201901924>. Accessed 13 May 2024.
18. Shi, Zhengya, et al. "Morphological Engineering of Sensing Materials for Flexible Pressure Sensors and Artificial Intelligence Applications." *Nano-Micro Letters*, vol. 14, 2022, p. 141, <https://doi.org/10.1007/s40820-022-00874-w>.
19. Paternò, Linda, and Lucrezia Lorenzon. "Soft Robotics in Wearable and Implantable Medical Applications: Translational Challenges and Future Outlooks." *Frontiers in Robotics and AI*, vol. 10, 2023.
20. "The Economics of Artificial Organs." *Artificial Organs*, 2024, <https://onlinelibrary.wiley.com/doi/abs/10.1111/aor.14592>. Accessed 28 Feb. 2024.
21. "Hemocompatibility - an Overview." *ScienceDirect*, 2024, <https://www.sciencedirect.com/topics/engineering/hemocompatibility>. Accessed 13 May 2024.
22. Zhou, Jianshu, et al. "Antagonistic Pump with Multiple Pumping Modes for On-Demand Soft Robot Actuation and Control." *IEEE/ASME Transactions on Mechatronics*, 2023, pp. 1–13, <https://doi.org/10.1109/TMECH.2023.3339814>.
23. "Bio-SHARPE: Bioinspired Soft and High Aspect Ratio Pumping Element for Robotic and Medical Applications." *Soft Robotics*, 2024, <https://www.liebertpub.com/doi/abs/10.1089/soro.2021.0154>. Accessed 28 Feb. 2024.
24. Tse, Yu Alexander, et al. "Novel Design of a Soft Pump Driven by Super-Coiled Polymer Artificial Muscles." *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 8789–8794, <https://doi.org/10.1109/IROS45743.2020.9341366>.
25. Becky. "A Brief Guide to Electric Actuators." *JHFOSTER*, 25 May 2022.
26. "Electric Actuator - an Overview." *ScienceDirect*, 2024, <https://www.sciencedirect.com/topics/engineering/electric-actuator>. Accessed 13 May 2024.
27. "Hydraulic Actuators." *Encyclopedia*, 2024, <https://encyclopedia.pub/entry/8168>. Accessed 13 May 2024.
28. Diteesawat, Richard Suphapol, et al. "Design Exploration of Electro-Pneumatic Pumps (EPPs) to Obtain High Pressure and Air Flow Rate Improvement." *Electroactive Polymer*

- Actuators and Devices (EAPAD) XXIV*, edited by John D. Madden, et al., SPIE, 2022, p. 20, <https://doi.org/10.1117/12.2611744>.
29. C1_Admin. "How Does Pneumatic Actuator Work?" *JHFOSTER*, 1 Dec. 2020.
 30. Lo, Chiao-Yueh, et al. "Highly Stretchable Self-Sensing Actuator Based on Conductive Photothermally-Responsive Hydrogel." *Materials Today*, vol. 50, 2021, pp. 35–43, <https://doi.org/10.1016/j.mattod.2021.05.008>.
 31. Wu, Shuang, et al. "Fast Thermal Actuators for Soft Robotics." *Soft Robotics*, vol. 9, 2022, pp. 1031–1039, <https://doi.org/10.1089/soro.2021.0080>.
 32. Chandorkar, Yashoda, et al. "Cells Feel the Beat – Temporal Effect of Cyclic Mechanical Actuation on Muscle Cells." *Applied Materials Today*, vol. 27, 2022, 101492, <https://doi.org/10.1016/j.apmt.2022.101492>.
 33. "Chronic Kidney Disease in the United States." *CDC*, 2023, <https://www.cdc.gov/kidneydisease/publications-resources/ckd-national-facts.html>.
 34. Liu, Tingting, et al. "Soft, Stretchable, and Skin-Mountable Drug Delivery Devices for Transdermal and Topical Applications." *Advanced Functional Materials*, vol. 30, no. 12, 2020, <https://doi.org/10.1002/adfm.201907199>.
 35. "Polymeric Actuators Market Analysis." *Precedence Research*, 2022.
 36. Wang, Zhihong, et al. "Vortex Flow Control for Robotic Blood Pump with a Long-Term Implantable Feature." *Science Robotics*, vol. 7, no. 69, 2022, eabo1990, <https://doi.org/10.1126/scirobotics.abo1990>.
 37. Su, Hao, et al. "Challenges and Frontiers of Bio-Inspired Soft Robotics." *Acta Biomaterialia*, vol. 132, 2021, pp. 102–115, <https://doi.org/10.1016/j.actbio.2021.04.047>.
 38. "Wearable Robots: The Role of Transient Thermal Actuators in Biomedical Devices." *Medical Devices and Sensors*, vol. 5, no. 2, 2022, e10281, <https://doi.org/10.1002/mds3.10281>.
 39. "Advances in Materials for Soft Robotics and Actuation Systems." *ScienceDirect*, 2022, <https://www.sciencedirect.com/science/article/pii/S2542454822001525>.
 40. Vasquez, Trevor. "Soft Robotics and Its Application in Bio-Medical Devices." *SciTechDaily*, 2023.
 41. Okamura, Allison M., et al. "Soft Actuators for Surgical Robots." *Science Robotics*, vol. 7, no. 69, 2022, eabo1991, <https://doi.org/10.1126/scirobotics.abo1991>.
 42. "Polymer-Based Soft Robotics for Biomedical Applications." *Advances in Colloid and Interface Science*, vol. 306, 2022, 102714, <https://doi.org/10.1016/j.cis.2022.102714>.
 43. "Robotic Artificial Organs: Concepts and Future Directions." *ScienceDirect*, 2022, <https://www.sciencedirect.com/science/article/pii/S1745272821001342>.
 44. Bonilla, Alicia, et al. "Wearable and Implantable Sensors for Human Motion Sensing and Body Monitoring." *Frontiers in Robotics and AI*, vol. 7, 2023.
 45. Yi, Min, et al. "A High-Performance Graphene-Based Soft Actuator Powered by Electrostatic Repulsion for Biomedical Application." *Nanoscale*, vol. 14, 2022, pp. 8324–8330, <https://doi.org/10.1039/D1NR08716C>.

46. Wang, Hao, et al. "Soft Robotics and Artificial Muscle Technologies: The Next Frontier in Biomedical Engineering." *Nature Communications*, vol. 14, no. 1357, 2023, <https://doi.org/10.1038/s41467-023-37615-4>.
47. Diteesawat, Richard Suphapol, et al. "Self-Sensing Smart Artificial Muscles Using Piezoelectric-Based Strain Sensors." *IEEE Transactions on Mechatronics*, vol. 26, no. 5, 2021, pp. 2391–2401, <https://doi.org/10.1109/TMECH.2021.3059279>.
48. Wu, Xiaojun, et al. "Bio-Inspired and Smart Soft Robotic Devices for Medical and Healthcare Applications." *Advanced Intelligent Systems*, vol. 3, no. 8, 2021, 2000284, <https://doi.org/10.1002/aisy.202000284>.
49. "Revolutionizing Healthcare with Bio-Inspired Robotics." *The Robotics Institute*, 2022.
50. Doyle, Rachael A., et al. "Flexible Bio-Mimetic Soft Robotics: A Step Toward Human-Like Devices for Healthcare." *IEEE Transactions on Robotics*, vol. 7, no. 1, 2023, pp. 59–69, <https://doi.org/10.1109/TRO.2023.3246598>.
51. Zhang, Yiwen, et al. "A Soft Pneumatic Actuator for Wearable Hand Rehabilitation Devices." *Journal of Neuroengineering and Rehabilitation*, vol. 20, 2023, <https://doi.org/10.1186/s12984-022-01138-w>.
52. Paternò, Linda, et al. "Soft Robotics in Medical Applications: Current Advances and Future Perspectives." *Advanced Intelligent Systems*, vol. 4, no. 3, 2022, 2100305, <https://doi.org/10.1002/aisy.202100305>.
53. Cheng, George, et al. "The Future of Robotic Surgical Systems." *Nature Biomedical Engineering*, vol. 6, 2022, pp. 531–543, <https://doi.org/10.1038/s41551-022-00879-6>.
54. Li, Shuangyang, et al. "Bio-Inspired Robust, Superhydrophilic and Superlubric Artificial Vascular Endothelium Coating for Anti-Thromboinflammation on Blood-Contacting Devices." *Composites Part B: Engineering*, vol. 257, 2023, p. 110670, <https://doi.org/10.1016/j.compositesb.2023.110670>.
55. Lavery, Karen S., et al. "Anti-thrombotic Technologies for Medical Devices." *Advanced Drug Delivery Reviews*, vol. 112, 2017, pp. 2–11 <https://doi.org/10.1016/j.addr.2016.07.008>.
56. Hong, Jun Ki. "Liquid-Infused Surfaces for Anti-Thrombogenic Cardiovascular Medical Devices." 2023.
57. Yoshizaki, T., et al. "PMEA Polymer-Coated PVC Tubing Maintains Anti-Thrombogenic Properties during In Vitro Whole Blood Circulation." *The International Journal of Artificial Organs*, vol. 28, no. 8, 2005, <https://journals.sagepub.com/doi/abs/10.1177/039139880502800809>. Accessed 14 July 2024.

I've Been Freed: Understanding How Womanism Informs Black Women's Identity Development By Zoe' Bell

Abstract

This study investigated how womanist theory contributes to the solidification of Black women's identities. Through focus groups conducted via Zoom, participants engaged in discussions about their experiences with womanism and identity. Each session included three participants and lasted 50-60 minutes, guided by questions formulated from existing research on Black feminist theory and intersectionality. The research aimed to uncover nuanced insights into the lives of Black women and femmes, highlighting the lifestyle of womanists and the influence of womanism on their self-identity. The findings revealed four primary themes that illustrate the impact of womanist theory on self-identity, encompassing personal growth, empowerment, and the interplay of individual and collective experiences. Overall, the study underscores womanism as a transformative force, bridging the gap between theoretical understanding and practical implications, thereby enhancing the lives of Black women and femmes.

Introduction

Womanist theory, rooted in the experiences of Black women and femmes, is centered around the intersectional oppression of race, gender, class, and other social identities (Phillips). Research has explored how womanist and black feminist theory informs social activism for Black women and femmes in the past and present (Taylor 26). While womanism and Black feminism share similarities, there are important distinctions between the theories. While there has been some research on the impact of Black feminism on identity development, there remains a gap in understanding how womanism informs Black women's and femme's self-identity. This study aimed to showcase the fundamentals of womanism and its ongoing influence on Black women and femmes.

Womanism and Black Feminist Theory

While womanism offers a distinct perspective on the experiences and struggles of Black women and femmes, the research data is relatively scarce. So, the discourse on intersectionality, gender, race, and class centers on the black feminist framework. The connection between womanism and black feminism is intricate and close-knit. Focusing on the key concepts created by both womanist and Black feminists can help better analyze the experiences and challenges of Black women and femmes (Collins 15). Both movements aim to dismantle systems of oppression while appreciating Black women's voices.

Womanism and Black Feminism are interconnected movements centering on the experiences and empowerment of Black women and femmes. In "Making Waves: The Theory and Practice of Black Feminism" Ula A. Taylor (26) states that "Black feminism and womanism would seem to be nearly interchangeable empowerment theories" meaning that both ideologies share fundamental principles and goals. Taylor further elaborates by saying "Both theories

encourage Black women to value and love self...both recognize Black women's serious, responsible commitment to creating a whole community void of dominance” (Taylor 26) Taylor emphasizes the fact that womanism and black feminism both champion the dismantling of oppressive systems that cause Black women and femmes marginalization—additionally advocating for the creation of inclusive and equitable communities that are free from said dominance.

Black Feminist Theory

Black feminist theory evolved during the second wave of feminism as a response to mainstream feminist movements that did not address racism and serious concerns from Black women, femmes, and other women of color (Taylor 239). During the second wave of feminism, there was a surge of social, political, and cultural movements created to tackle gender inequality. Women campaigned for equal pay, workplace equality, and greater political representation. However, the unique challenges faced by Black women were not accounted for in these efforts.

One example of Black women’s unique circumstances being ignored can be seen in the Equal Pay Act of 1963. While this legislation was a crucial step in addressing gender-based pay discrimination, its implementation mostly benefited white women. The wage gap experienced by Black women and femmes has historically been wider than that of white women due to the intersections of racism and sexism (Dozier 280-282). The mainstream feminist movement focused on closing the wage gap through legal reforms without acknowledging the unique challenges Black women and femmes faced in the workforce. Issues such as segregation, racial discrimination in hiring and promotion, and lack of access to education were often overlooked in discussions about equal pay. The strategies that were advocated for by some white feminists, such as negotiating for higher salaries, failed to account for the systemic barriers Black women and femmes faced. For example, Black women may encounter backlash when attempting to assert their rights in the workplace (Hightower 26-27). In failing to take an intersectional approach to the issue of equal pay, the feminist movement at times perpetuated the marginalization of Black women and femmes which hindered efforts to achieve genuine gender equality in the workforce. Thus Black feminism emerged to specifically focus on the unique position of Black women and femmes in society, highlighting the intersecting oppressions of race, gender, and class. Black feminism seeks to amplify the voices of Black women by advocating for their liberation in all aspects of society (Hamer and Neville 438-441).

Scholars have reflected on the depiction of Black women and femmes in literature and other media through a womanist/black feminist lens. For example, Reynolds (591-606) studied the connection between literature and Black feminist standpoint theory. Reynolds critically examined the intersection of literary and pop cultural forums in centering Black women and femmes' voices. She highlighted the commodification of Black women and femmes experiences which reinforces existing stereotypes. While acknowledging the political importance of Black feminist standpoint theory, Reynolds argued for a more grounded approach within feminist discourse. Reynolds then further advocated for a framework that truly captures Black women and

femmes outside of societal marginalization and oppression. She argued that making oppression the center of Black feminist standpoint theory can lead to a narrow portrayal of Black womanhood that fails to capture its full range of experiences. Moving beyond essentialized portrayals acknowledges the agency, resilience, and multifaceted identities of Black women and femmes. Thus, moving beyond a singular focus on victimhood and oppression. Overall, Reynolds argued that a more evolved black feminist standpoint theory has the ability to effectively relay Black women's lived experiences accurately. Meaning, this lens has the ability to better portray Black women as fleshed-out and multifaceted individuals.

Scholars have also theorized regarding the importance of womanist/black feminist theories on resistance against white supremacist oppression. Collins's work "Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment," reflected on the ideas of oppression, resistance, and empowerment. Collins outlined five features of black feminist thought emphasizing the intersection of gender, race, class, and sexuality within the matrix of domination that shapes the resistance of Black women and femmes. She then highlighted the collective identity of Black womanhood shaped by oppression, while still acknowledging the differences of Black women and femmes since they are not a monolith. She stated that this community knowledge of experience is essential for mobilizing resistance due to the shared history. Collins also highlights the importance of advocating for other diverse groups in the struggle against white supremacy. She reiterates that possessing a critical knowledge of one's position in society can lead to better resistance against power structures. Overall, Collins's framework offers a nuanced understanding of the intersecting forms of Black women and femmes oppression. She addresses these power dynamics through a critical lens while promoting social justice and resistance.

Additionally, there is research that builds upon existing knowledge regarding Black women's gendered racial identity development which aims to fill gaps in understanding key experiences associated with identity specifically with Black women (Williams and Lewis) The researchers acknowledge previous studies that have identified contextual factors influencing Black women's gendered racial identity, while also highlighting the need for a deeper exploration of the developmental process and the ideologies Black women ascribe to their identities. The researchers conducted semi-structured individual interviews with 19 Black women at a predominantly White university in the Southeast. The findings revealed a dynamic and changing developmental process consisting of four phases: hyperawareness, reflection, rejection, and navigation. These phases are shaped by various factors unique to Black women's intersectional experiences. The Hyperawareness phase involves becoming acutely aware of one's identity as a Black woman within society. It may include heightened sensitivity to discrimination, stereotypes, and social expectations based on race and gender. Subsequently in the reflection phase, individuals engage in introspection and critical examination of their experiences and identity. They may reflect on how societal perceptions and expectations influence their sense of self and their place within their communities. During rejection, individuals may reject societal norms, stereotypes, and expectations that do not align with their authentic selves. They may resist

pressures to conform and assert their agency in defining their identity and values. Finally, in the navigation phase, individuals develop strategies to navigate the complexities of their intersecting identities. This may involve finding ways to reconcile conflicting societal expectations, advocating for themselves and their communities, and creating spaces where they feel empowered and affirmed.

Additionally, the study identified six gendered racial ideologies—assimilation, humanist, defiance, strength, pride, and empowerment—that represent Black women's values, beliefs, and attitudes towards their gendered racial identity. Assimilation for instance involves conforming to dominant cultural norms and values to gain acceptance. While the humanist ideology emphasizes equality and inclusivity of all individuals, regardless of race and gender. Defiance involves challenging systemic oppression whereas strength celebrates the ability of Black women to overcome barriers despite discrimination. Pride encourages positive affirmation and celebration of Black womanhood and empowerment focuses on fostering autonomy and agency among Black women. Importantly, the study highlighted that the developmental process is not linear, with participants often cycling through different phases and believing in multiple ideologies at different points in their lives. This nuanced understanding emphasizes the complexity of Black women's gendered racial identity development and provides insights for researchers, educators, and policymakers to increase critical awareness and support for Black women in various contexts.

While both Black feminist theory and womanist theory offer valuable perspectives on the experiences and struggles of Black women, it is essential to acknowledge the limitations of studying Black feminist theory alone. Black feminist theory often focuses on issues specific to gender within the Black community. However, it may not fully encompass the broader socio-cultural, historical, and spiritual dimensions that womanist theory addresses. Womanist theory, coined by Alice Walker, considers the intersections of race, gender, class, and community. It offers a more holistic understanding of the experiences of Black women and femmes. Therefore, while Black feminist theory provides critical insights, studying womanist theory alongside it can offer a more comprehensive understanding of the complexities of Black women's and femmes lives.

Womanist Theory

Womanist theory first emerged as a framework within the feminist discussion as a perspective that centers on Black women and femmes' lived experiences. Rooted in the works of several scholars such as Alice Walker, Chikwenye Okonjo Ogunyemi, Audre Lorde, Patricia Hill Collins, Paula Giddings, and bell hooks, womanist theory was created to address the intersections of race, gender, class, and other social identities (Phillips), which further acknowledges the unique struggles Black women and femmes face under white supremacist systems. Womanist theory can be traced back to the writings of Black feminist scholars, including Alice Walker. Walker coined the term womanist in her 1983 essay “In Search of Our Mothers’ Gardens.” (Walker). In this essay, Walker explores the creativity and resilience of

Black women and femmes. The title itself, “In Search of Our Mothers’ Gardens,” suggests a journey of rediscovery. Walker seeks to uncover the hidden treasures of creativity from her foremothers. (e.g. Walker's research and discovery of the late Zora Neale Hurston.) She discusses the importance of recognizing the artistic contributions of Black women and femmes throughout history. Especially those who were denied opportunities to express themselves fully due to societal constraints such as slavery, racism, and sexism. Walker argues that by celebrating this creativity, society can gain a deeper understanding and appreciation of Black women and femmes throughout history. Womanism provides a broad appreciation of the cultural, spiritual, and socio-political factors of Black women and femme's lives.

According to Walker, the womanist framework is multifaceted. Walker offers several definitions. A womanist is a person who identifies as a Black feminist or feminist of color who loves other women, sexually and/or non-sexually, and appreciates and prefers women’s culture, women’s emotional flexibility (values tears as a natural counterbalance of laughter), and women’s strength. Additionally, a womanist loves music, loves dance, loves the moon, and loves the Spirit. A womanist loves love and food and roundness. Furthermore, a womanist loves struggle, loves the Folk, and loves herself... regardless” (Walker 1-3) Finally, the fourth facet of the theory of womanism is the infamous expression "Womanist is to feminist as purple is to lavender” (Walker xii). This analogy emphasizes the idea that womanism and feminism are related but distinct movements, with womanism encompassing a broader perspective on gender, race, and social justice. Just as purple contains shades like lavender but also extends beyond it, womanism builds upon feminist principles. At the same time, offering a more comprehensive framework for addressing the intersecting oppressions faced by Black women and femmes.

Unfortunately, there is limited research focusing on womanist theory. In one study, Rodgers utilizes a womanist framework to examine the issue of domestic violence within Black women's experience. Rodgers highlighted the importance of centering Black women in discussion of domestic violence. Through a womanist lens, the study addressed the immediate issues of domestic violence but also within a broader historical and cultural context. Rodgers stated that domestic violence prevalence can be attributed to the messages Black women receive from community, family, and institutions. She elaborated saying “The messages encourage them to uphold the sanctity of marriage, to show “racial loyalty” to protect Black men (and the concomitant fear that the police and the criminal justice system will harm or display bias against their partner) and to treat domestic violence as a private or personal matter not to be shared in public settings (Bent-Goodley and Rodgers,).” Rodgers emphasized that the belief of keeping domestic violence private reinforces the idea that Black women should not seek external support for their safety. Meaning these societal messages contribute to the silencing of Black women’s experience, making them feel obligated to prioritize the needs of others over their autonomy. Through the use of a womanist framework, Rodgers asserts that womanist ideology is essential in eradicating domestic violence in Black women. By highlighting the womanist principles of community and empowerment, the study offers insights for addressing domestic violence within the Black community.

Current Study

The rationale for focusing this study on womanism heavily rests on the second facet of womanism. This facet of womanism recognizes the worth of Black women and femmes in a society that often marginalizes or disregards their experiences and contributions. Womanism, while advocating for the liberation of Black women and femmes, also celebrates and values their creations deeply. This appreciation has the power to profoundly impact Black women and femmes' self-identity. The love and recognition that womanism presents Black women and femmes can inspire them to cultivate self-love and acceptance outside of traditional feminist movements. By honoring Black women and femmes' contributions and inherent worth, womanism fosters an environment where Black women can see themselves through a lens of empowerment. Thus promoting a positive sense of self-identity and agency.

While womanism has made strides in addressing intersectional oppression, there is a notable gap in understanding the role it plays in individual identity. Despite the emphasis on empowerment and liberation, there's no research on womanist theories' influence on self-perception in Black women and femmes.

Thus, this study aimed to investigate how womanist theory helps solidify Black women's and femme's identities. By centering the voices of Black womanists I aim to uncover the ways in which womanist theory shaped their self-identity. This study aims to fill existing gaps by exploring real-world implications of womanist theory on self-identity through focus groups which will provide insights into if the womanist journey can be transformative.

Methods

The Institutional Review Board approved this study. Participants were recruited using personal contacts. Participants who were interested in participating emailed the researcher for additional information about the study. After the participants emailed the researcher, the researcher responded to the participants' inquiries with a Google Forms link that contained inclusion criteria questions, a consent form, and a demographics questionnaire. If participants did not meet the inclusion criteria, they received a message thanking them for their interest, but declining their participation. If participants did meet the inclusion criteria, they were shown the consent form. Once the participant indicated consent, they were taken to the demographics questionnaire and provided a link to the scheduling platform When2Meet to indicate their availability for a focus group.

The focus groups were conducted using the Zoom platform. Once 3 participants indicated that they were available for a focus group at the same time, a link to the Zoom meeting was created and sent to each available participant. Focus groups contained 3 participants and each session lasted 50-60 minutes. During the focus groups, participants were asked a series of questions regarding Womanism and identity. The researcher drew inspiration from existing research focused on Black feminist theory and intersectionality to formulate questions. The existing research helped the researcher formulate open-ended questions along with deciding to adhere to a semi-structured interview protocol. Probing questions were asked when necessary to

encourage participants to elaborate on their responses. Throughout the focus groups, the comfort of the interviewees was the central focus. The researcher reviewed informed consent, discussed how participant identity would be protected, and answered any initial questions the participants had before continuing with the interview. At the conclusion of the focus groups, participants received a thorough explanation of the study's purpose and the researcher addressed any remaining questions the participants had. After the study, participants were given the researcher's contact information should they have questions arise after the conclusion of the focus group.

After the conclusion of the focus groups, interviews were transcribed using Microsoft Word. Once interviews were transcribed, thematic analysis was conducted to determine common themes that were present across focus groups.

Overview of thematic analysis

Thematic analysis is a method used to identify, analyze, and report patterns (themes) within data. It involves systematically reviewing qualitative data to find recurring themes that represent the core meanings within the dataset.

To complete the thematic analysis, focus group recordings were first transcribed. Next, the transcripts were read multiple times by the researcher until the content became familiar. Then, coding the data began with the highlighting of significant phrases, sentences, or sections that related to the research questions. These codes were then grouped into broader themes, which were reviewed to ensure they accurately represented the data. Finally, the themes were synthesized to draw meaningful insights and conclusions.

Participants

The sample consisted of five Black cisgender females between the ages of 18 to 40. All participants identified as heterosexual and heteroromantic. The majority were undergraduate students, with a range of familiarity with Womanist theory—from those who were very familiar to others who were not familiar at all. None of the participants reported having a disability. This group provided a focused and relevant perspective, especially concerning the intersections of race, gender, and social activism.

The Researcher

As a Black woman and high school student, my interest in exploring the impact of womanism on Black women's self identity stems from my personal journey. I entered this research with the belief that womanist theory would demonstrate a positive influence on self-identity, as it has profoundly shaped my own. My academic position for womanist readings has fueled my commitment to this study. While I share both race and gender identity with participants, I approached the research as an outsider in terms of personal connection, ensuring that my analysis was both respectful and reflective of the diverse experiences shared with me.

Results and Discussion

Four primary themes emerged that highlight the impact of womanist theory on self-identity. These themes are: Advocacy and Rejection of Stereotypes, Identity Formation and Self-Awareness, Personal Growth and Transformation, and Empowerment Through Community. Each theme provides insight into different aspects of how womanist theory supports and enriches the lives of Black women and femmes, addressing both individual and collective experiences.

In the following sections, we will delve into each theme, presenting key findings that illustrate the significance of womanist theory in shaping and reinforcing Black women's and Black femmes identities.

Theme 1 Advocacy and Rejection of Stereotypes

Womanist theory provides a comprehensive framework for understanding and addressing the intersectional struggles faced by Black women and femmes. Central to this theory is the emphasis on liberation, not just for individuals, but for the entire community. This inclusive approach encourages the dismantling of systemic barriers that restrict the autonomy and potential of Black women and femmes. Participants in the study frequently articulated how womanism's emphasis on collective liberation resonated deeply with their experiences, guiding them to challenge and resist the restrictive societal norms imposed upon them.

"I feel like just it's empowering in the sense that it's validating all of like the stances and the words that black women hold and express" - S

By embracing womanist principles, participants found a sense of empowerment that transcended their circumstances and connected them to a broader struggle for justice. The study reveals that this collective mindset fosters a sense of responsibility among Black women to liberate themselves and contribute to the liberation of others within their community. This aspect of womanism is particularly significant in its rejection of societal stereotypes. Participants described how womanism empowered them to reject reductive labels and assert their identities in ways that were authentic and self-determined.

"But then after learning about womanism and how historically embedded it is [that] we're angry black women, I just let that go. Like, if I come off as angry I don't care. I'm expressing myself." - RB

The theory's insistence on the interconnectedness of all struggles against oppression provided participants with a framework to critique and resist how dominant cultural narratives perpetuate stereotypes. The liberatory aspect of womanism, therefore, becomes a tool for participants to redefine their identities and reject external attempts to impose limiting or harmful stereotypes. As Robyn, a participant in this study, noted, "... it's given me a foundation for understanding the world, understanding liberation, and doing something about it."

Furthermore, the study highlighted the role of womanism in challenging internalized stereotypes and encouraging self-advocacy. Participants noted that womanist theory helped them to recognize and reject the negative stereotypes that they had internalized about themselves and other Black women. By doing so, they were able to develop a stronger sense of self-worth and become more effective advocates for their own needs and the needs of others. Womanism's emphasis on self-love and self-acceptance thus becomes a powerful tool for advocacy, enabling Black women to assert their rights and demand respect in both personal and public spheres.

A significant finding of the study is how womanist theory actively encourages Black women and femmes to engage in resistance against systemic oppression. Womanism is not merely a theoretical construct but a call to action, urging individuals to confront and dismantle the societal structures that perpetuate inequality. Participants expressed how womanism inspired them to become more socially aware and to take concrete steps in combating various forms of oppression in their personal lives, within their communities, and on a larger societal scale.

“Yeah, I would have to agree just from the knowledge I've taken from class and through my research I'm just like, more aware of the Injustices that occur in our society and how, like I'm sort of vulnerable to experiencing those injustices as well, being aware of like my workplace and in school” - S

This commitment to action is deeply embedded in the womanist tradition, which views the struggle against oppression as both a personal and a collective responsibility. The participants' narratives highlighted how womanism empowered them to recognize and address the multiple layers of discrimination they face, including racism, sexism, and classism. Womanism's focus on intersectionality allowed them to see how these different forms of oppression are interconnected and how they reinforce one another. As a result, participants were more equipped to develop strategies for resisting these oppressive forces in their everyday lives.

Moreover, the study found that womanism's emphasis on social awareness extended beyond personal empowerment to include a broader commitment to community activism. Participants discussed how their understanding of womanist principles led them to advocate not just for themselves, but for other marginalized individuals as well. This often took the form of advocating for others in their professional spaces.

“So I feel like especially like in terms of advocating for my patients, I think I definitely use womanism.” - Butterfly

Womanism, therefore, serves as both a lens for understanding oppression and a blueprint for resistance. Womanist theory's strong focus on advocacy emerged as a critical theme in the study, particularly concerning the support and uplift of Black women. Participants emphasized that womanism provided them with the tools and language necessary to advocate for the rights and well-being of other Black women. This advocacy is not limited to formal political or social

movements but extends into everyday interactions and relationships. By embracing womanist principles, participants felt more empowered to challenge injustices directed at themselves or at others.

The study also revealed that womanist theory fosters a sense of solidarity among Black women, creating a supportive network that is crucial for effective advocacy. Participants described how womanism helped them to see the importance of standing together with other Black women, recognizing that their struggles are interconnected. This sense of solidarity not only strengthened their resolve to advocate for each other but also deepened their understanding of the systemic nature of the challenges they face.

“It's the community like we uplift each other. I uplift others and knowing that I can make a difference on other black women's lives and knowing that I have a say and that my voice is only amplified by many others helps empower me” - Grace Green

The final key finding in this theme is the support that womanist theory generates for the broader womanist movement. Participants in the study both identified with womanist principles on a personal level and expressed a commitment to supporting and advancing the womanist movement as a whole.

The study found that this collective advocacy effort is a crucial aspect of womanist theory, as it reinforces the idea that the struggle for liberation and justice is a shared responsibility. Participants discussed how their involvement in the womanist movement provided them with a sense of purpose and belonging, as well as a platform to amplify their voices and those of others. By supporting the womanist movement, participants contributed to a larger, more powerful force for change, one that seeks to dismantle the oppressive structures that affect not only Black women and femmes but all marginalized groups.

The theme of Advocacy and Rejection of Stereotypes highlights the multifaceted ways in which womanist theory empowers Black women and femmes to challenge societal norms, advocate for themselves and others, and contribute to the broader womanist movement. This theme underscores the importance of womanism as both a personal and collective framework for liberation, resistance, and empowerment.

Theme 2 Identity Formation and Self-Awareness

The theme of Identity Formation and Self-Awareness delves into how womanist theory helps Black women and femmes navigate and resist the historical tropes and beauty standards imposed by a dominant culture that often devalues their identities. Womanism offers a critical lens through which participants can interrogate these expectations and understand their origins, impact, and the ways they perpetuate marginalization. Participants in the study frequently referenced how womanist theory allowed them to critique mainstream beauty standards that privilege Eurocentric features and marginalize Black beauty.

“I think of Angela Davis's fro and like, you see black women in their natural full authentic state. And so hair is a big thing I think about a lot. And I would agree like, my hair journey definitely changed. I stopped straightening my hair for sure. I started looking at myself differently” - Robyn

Through womanist principles, participants found validation in their natural features and cultural expressions, which led to a stronger sense of self-worth and identity. Before engaging with womanist theory, many participants in the study reported feeling isolated and, at times, questioning their own perceptions of reality when confronting experiences of discrimination. These feelings of self-doubt and confusion often stemmed from society's pervasive invalidation of their lived experiences by frequently downplaying or outright dismissing the unique challenges Black women and femmes face. Without a supportive framework to contextualize these experiences, participants described feeling "crazy" or overly sensitive, unsure if their perceptions were valid or if they were somehow overreacting. However, womanism provided a critical shift in perspective by affirming the legitimacy of their experiences and offering a theoretical foundation that validated their experiences of discrimination and marginalization.

"I feel like, just it's empowering in the sense that it's validating all of the stances and the words that black women hold and express.” - S

This validation was important, as it not only confirmed that their experiences were real and significant but also connected them to a broader community of individuals who shared similar struggles. As one participant stated,

”So I think in a way, womanism can validate some of your experiences of discrimination and help remind you that it's not like you're crazy and all of these other things.” - RB

Through this newfound understanding, participants moved past feelings of self-doubt and embraced their reality with confidence and clarity, recognizing that they were not alone in their experiences. A critical aspect of identity formation within womanist theory is the journey toward self-acceptance and personal growth. Womanism encourages a process of self-reflection and revelation, in which individuals come to understand and accept themselves on a deeper level. Participants in the study frequently discussed how womanism facilitated a process of self-discovery that was transformative in their lives. This journey often began with a rejection of externally imposed identities and progressed towards an embrace of their true selves, informed by a deeper understanding of their cultural and personal histories.

The study revealed that this journey is often marked by significant moments of revelation—instances when participants gained profound insights into their own identities and the societal forces that shaped them. These revelations were often accompanied by a growing sense of self-acceptance. Participants began to appreciate the fullness of their identities,

including aspects that had previously been sources of insecurity or shame. Womanist theory provided the language and concepts necessary to articulate these experiences and integrate them into a coherent sense of self.

RB stated, “How you should be as a woman, or what you feel comfortable with as a woman, like, solidified my part in the womanist movement because I was able to abandon the ideas of having straight hair, looking presentable in front of people, and learning to love my natural hair. Love my fiery side. Yeah. So that's how I came to self-acceptance with womanism”

The study also found that this hope for self was closely linked to a greater sense of agency and autonomy. Womanism’s emphasis on self-determination and empowerment resonated with participants, encouraging them to take control of their lives and make decisions that were in alignment with their values and aspirations.

“I shouldn't have to put myself in a little box and not express myself to the fullest extent because it doesn't make you happy” - RB

Furthermore, this process of self-acceptance was not static but rather a continual journey of growth. Participants noted that womanism taught them to view their identity as dynamic, something that evolves over time as they gain new insights and experiences. This perspective was particularly liberating, as it allowed them to see personal growth as a lifelong process rather than a fixed endpoint. Womanism, therefore, becomes a tool for ongoing self-improvement and empowerment, encouraging individuals to continually refine their understanding of themselves and their place in the world.

“So I would say like, that's a parallel between my womanist journey and my self identity journey that you're just continuously learning” - Robyn

The theme of Identity Formation and Self-Awareness demonstrates how womanist theory plays a crucial role in helping Black women and femmes navigate the complexities of their identities. Through a combination of historical critique, self-reflection, and practical application, womanism provides a comprehensive framework for identity formation that is both empowering and transformative. This theme underscores the importance of womanism as a tool for self-discovery, self-acceptance, and ongoing personal growth, all within an inclusive and supportive framework.

Theme 3 Personal Growth and Transformation

The theme of Personal Growth and Transformation examines how engagement with womanist theory facilitates profound personal development and life-changing experiences for Black women. Central to this theme is the concept of reflection, in which individuals use

womanist principles to critically examine their lives, relationships, and societal roles. Participants in the study often described their engagement with womanism as a journey, one that involved continual self-reflection and reassessment of their values, beliefs, and behaviors. This reflective process was seen as essential for their personal growth, allowing them to identify and overcome internalized oppression and other barriers to their empowerment.

“I think sometimes people don't do enough self-reflection or enough figuring out, like where they're at with different constructs or different theories, and I think womanism is one of them, where it takes a lot of critical self-awareness and reflection to really figure out where you're at in the theory, how you want to apply it to your life.” - Robyn

Reflection within the womanist journey was frequently linked to moments of realization when participants came to understand the deep-seated impact of societal norms on their self-perception and behavior. For many, womanism provided the tools to deconstruct these norms and replace them with values that were more aligned with their true selves. This process of reflection and realignment was transformative, leading to significant changes in how participants viewed themselves and their place in the world.

“I started looking at myself differently, especially like RB you brought up earlier, the hyper-sexualization, I think I started interrogating a lot of messaging I was getting. I started asking myself like, ‘Why are you wearing these clothes? Is it because you want to wear these clothes or because you feel like you need to?’”- Robyn

The benefits derived from engaging with womanist theory were a central finding in the theme of Personal Growth and Transformation. Participants overwhelmingly reported that womanism offered them a sense of empowerment and a clearer understanding of their identity and purpose. This empowerment was not just theoretical but had practical implications, influencing how they navigated their personal lives.

“And so I think it gets you to a place of evaluation and it helps you kind of evaluate like where you're at in life, but also like is this authentic to me? Is this like congruent to who I want to be and all of these things?” - Robyn

Womanism provided participants with a framework to critically assess their circumstances, identify areas for improvement, and take actionable steps towards achieving their goals. One of the key benefits of womanism highlighted by the study was the development of a stronger sense of self-worth and confidence. Participants described how womanist principles helped them reject societal narratives that devalued their identities and contributions. Instead, they embraced a narrative of self-empowerment, in which their experiences and perspectives were valuable and deserving of respect. This mindset shift profoundly impacted their personal

growth, leading to greater self-assurance and a willingness to pursue opportunities that they might have previously shied away from.

“The short answer just [is it] empowers me to be myself. I'm just RB and I don't care how you perceive me. That's basically how I see it.”- RB

Additionally, the study found that womanism provided participants with a sense of hope and resilience. The theory's emphasis on collective struggle and liberation resonated deeply with participants, giving them the strength to persevere in the face of challenges. Womanism's focus on the interconnectedness of personal and communal well-being also encouraged participants to draw strength from their communities, further bolstering their resilience and capacity for personal growth.

“Just being a listening ear for those who have experienced or for black women who have experienced sexism, like intersectional sexism and racism together and just being a listening ear for them and relating to their issues and letting them know that they're not alone.” - Grace Green

This combination of self-empowerment, hope, and resilience was identified as a key driver of the transformative experiences reported by participants. The concept of hope emerged as a significant finding within the theme of Personal Growth and Transformation, particularly in relation to how womanism instills a sense of hope for one's future. Participants frequently spoke about how womanism gave them a renewed sense of possibility. It allows them to envision a future in which they could fully realize their potential without being constrained by societal expectations or limitations.

“I would say with knowledge comes power, right? So when you use that knowledge, it's freeing. And I think what's been cool is I've been freed.” - Robyn

This sense of hope was seen as a powerful motivator for personal growth, driving participants to take risks, pursue new opportunities, and strive for personal fulfillment. Womanism's holistic approach to liberation, which encompasses both individual and collective well-being, was particularly influential in fostering this sense of hope. Participants noted that womanism encouraged them to see their personal growth as part of a larger process of social change, where their successes and achievements contributed to the advancement of their communities. This perspective provided a sense of purpose and direction, reinforcing the belief that their personal development was not only possible but also necessary for the broader struggle for justice and equality.

“Understanding Womanism and knowing that you can feel empowered from the trauma and things that you go through in society can make a better impact on your life and your mental health, so that's one example for me.”- RB

Another key aspect of the theme of Personal Growth and Transformation is how deeply participants integrate womanist theory into their everyday lives. Womanism was not just an abstract or academic concept for participants. Rather, it was a lived philosophy that influenced how they approached every aspect of their lives, from their relationships and careers to their personal goals and self-care practices. The study found that this integration of womanist principles into daily life was a critical factor in the participants’ personal growth and transformation.

“The way I center black women in my life, like even in day-to-day things or in my thinking about things...I would say that I also center womanist theory, but also like black women, the work of black women and all of those things and how I portray myself, who I give credit to is a big thing that can kind of be tied into womanist theory” - Robyn

Participants discussed how womanism informed their decision-making processes, helping them to align their actions with their values and long-term aspirations. For example, when faced with difficult choices or challenges, participants often turned to womanist principles for guidance, using them as a framework to evaluate their options and determine the best course of action. This alignment between values and actions was seen as crucial for maintaining a sense of integrity and authenticity, both of which were identified as important components of personal growth.

“I just feel like by going into spaces and believing that you have every right to be there and like I'm a stem major, so I feel like in a lot of the spaces that I'm in, I'm not in the majority like I'm in the minority, especially as a black woman and that hasn't deterred me from continuing on so in that sense it mirrors the womanist movement.”- Grace Green

The final key finding in the theme of Personal Growth and Transformation is the sense of alignment that participants felt between their personal values and the principles of womanism. They often described this alignment as a process of coming home to themselves, in which they were finally living in a way that was true to their identities and aspirations. Womanism provided a framework that validated their experiences and offered a roadmap for living a life that was in harmony with their values and beliefs.

The study revealed that this sense of alignment was deeply empowering for participants, as it reinforced their sense of self-worth and confidence. By living in accordance with womanist principles, participants felt that they were not only being true to themselves but also contributing to a larger movement.

“Just like the whole encapsulation of it, I think that I have changed because I finally have found something that I can align myself with and advocate for.” - Butterfly

This alignment between personal and collective goals was seen as a key factor in their personal growth, as it provided both the motivation and the framework for continuous self-improvement and development. The theme of Personal Growth and Transformation highlights the profound impact that womanist theory has on the lives of Black women and femmes. Through a process of reflection, self-discovery, and alignment with womanist principles, participants experienced significant personal growth and transformation. This theme underscores the importance of womanism as a tool for navigating life’s challenges, fostering hope and resilience, and achieving personal and collective liberation. Womanism, as revealed in this study, is not just a theoretical framework but a lived philosophy that empowers individuals to grow, transform, and thrive in all aspects of their lives.

Theme 4 Empowerment Through Community

The theme of Empowerment Through Community emphasizes the crucial role that collective support plays in the lives of Black women and femmes engaged with womanist theory. Participants consistently highlighted the importance of community as a source of strength, resilience, and empowerment. Within these communities, participants found emotional support and practical guidance and solidarity in their struggles.

“And just knowing that there's other people that are potentially experiencing the same experiences as me or something very similar and I can like you know, talk to someone and they can actually hear me and, listen to my experiences and actually relate to my struggles and whatnot is just so amazing.” - S

The shared experiences and mutual understanding among community members fostered a sense of belonging and validation, which was essential for navigating the challenges they faced. This support network provided a safe space where participants could express their fears, frustrations, and aspirations without fear of judgment or dismissal. The sense of mutual understanding and support was described as deeply empowering, allowing participants to confront challenges with renewed confidence and determination.

“So I feel like knowing what womanism is and supporting the movement myself has helped me, especially like with self-confidence and just knowing that wherever I go there's someone there to help.” - Grace Green

The study also revealed that community support played a vital role in helping participants resist internalized oppression and reclaim their identities. By engaging with others who shared their struggles and aspirations, participants challenged negative stereotypes and affirmed their

value. The community acted as a buffer against the isolation and alienation that often accompany experiences of discrimination, providing participants with a collective voice that amplified their individual experiences.

“And so I feel like womanist theory in a way allows women to come together and just share their struggles to one another, and it's a safe space because they're not being judged because, you know, they've experienced something very similar, if not like the exact same.” - S

This collective empowerment was a key factor in participants' ability to resist oppression and assert their rights and dignity in various aspects of their lives. Another significant finding within the theme of Empowerment Through Community is the role of community encouragement in fostering personal and collective growth. Participants frequently spoke about encouraging others in their community to pursue their goals, take risks, and challenge societal norms. This encouragement was often rooted in the belief that each member of their community had unique strengths and contributions to offer and that their success was intertwined with the success of the group as a whole.

“I'm a therapist. So how I practice therapy, the people I work with. I specialize in working with black women, so it impacts a lot of what I do. And like who I see, how I interact with them, and things like that.” - Robyn

The study found that participants' empowerment was closely linked to their ability to effect change within their communities. By standing up for others, participants not only contributed to the well-being of their communities but also reinforced their own sense of agency and power. This reciprocal relationship between individual and collective empowerment was a key finding of the study, highlighting the importance of community in fostering both personal growth and social change.

Finally, the theme of Empowerment Through Community underscores the comfort that participants found in being part of a like-minded and supportive group. For many participants, the community provided a sense of safety and reassurance that was otherwise lacking in their interactions with the broader society. This comfort was not just emotional but also deeply connected to their sense of identity and belonging. Within their communities, participants felt free to express themselves fully, without the need to conform to external expectations or suppress aspects of their identities.

“I feel like it's always easier to accept yourself when you know that you're not alone and you know that people are going through the same things as you and those people are there to support what you're going through and to uplift you and to advocate for you.” - Grace Green

The act of advocating for others strengthened participants' resolve and deepened their commitment to the principles of womanism, further solidifying their identities as empowered individuals within a supportive and resilient community.

The sense of comfort and security that came from being part of a community where they were accepted and valued for who they were was described as profoundly empowering. This comfort also facilitated participants' engagement with womanist theory and its application in their lives. The study found that participants were more likely to explore and embrace womanist principles when they felt supported and understood by their communities. This collective exploration of womanism reinforced the bonds within the community, creating a shared language and set of values that further strengthened participants' sense of belonging and empowerment. As Grace Green stated, "It's the community where we uplift each other. I uplift others and knowing that I can make a difference in other black women's lives and knowing that I have a say and that my voice is only amplified by many others helps empower me" In this way, the community served as both a source of comfort and a foundation for the participants' personal and collective empowerment, enabling them to navigate the complexities of their identities and experiences with confidence and resilience.

Conclusion

The study found that womanist theory significantly impacts the identity formation, personal growth, and empowerment of Black women. Womanism validated participants' experiences of discrimination, reducing self-doubt and isolation. It also fostered personal growth, increasing self-awareness and confidence. Community played a key role, providing support, encouragement, and a sense of belonging, which empowered participants to advocate for themselves and others. Overall Womanism proved to be a transformative force. The theory helped participants affirm their identities and find fulfillment, even in a society that often works against them.

Works Cited

- Collins, Patricia Hill. *Black Feminist Thought : Knowledge, Consciousness, and the Politics of Empowerment*. Routledge, 2002.
- Collins, Patricia Hill. "What's in a name?" *The Black Scholar*, vol. 26, no. 1, Dec. 1996, p. 15, <https://doi.org/10.1080/00064246.1996.11430765>.
- Dozier, Raine. "Accumulating disadvantage: The growth in the black–white wage gap among women." *Journal of African American Studies*, vol. 14, no. 3, 21 Mar. 2010, pp. 280–282.
- Hightower, Chelsea D., "Exploring the Role of Gender and Race in Salary Negotiations" (2019). LSU Master's Theses. 4943. 26-27
https://digitalcommons.lsu.edu/gradschool_theses/4943<https://doi.org/10.1007/s12111-010-9122-5>.
- Neville, Helen A., and Jennifer Hamer. "'We Make Freedom.'" *Journal of Black Studies*, vol. 31, no. 4, Mar. 2001, pp. 438–441, <https://doi.org/10.1177/002193470103100404>.
- Phillips, Layli. *The Womanist Reader: The First Quarter Century of Womanist Thought*. Taylor and Francis, 2006.
- Reynolds, Tracey. "Re-Thinking a Black Feminist Standpoint." *Ethnic and Racial Studies*, vol. 25, no. 4, Jan. 2002, pp. 591–606, <https://doi.org/10.1080/01419870220136709>.
- Rodgers, Selena T. "Womanism and Domestic Violence." *Encyclopedia of Social Work*, May 2020, <https://doi.org/10.1093/acrefore/9780199975839.013.1073>.
- Taylor, Ula. "The historical evolution of Black Feminist Theory and praxis." *Journal of Black Studies*, vol. 29, no. 2, Nov. 1998, p. 239, <https://doi.org/10.1177/002193479802900206>.
- Taylor, Ula Y. "Making waves: The theory and practice of Black Feminism." *The Black Scholar*, vol. 28, no. 2, June 1998, p. 26, <https://doi.org/10.1080/00064246.1998.11430912>.
- Walker, Alice. *In Search of Our Mothers' Gardens: Womanist Prose*. Harcourt Brace Jovanovich, 1983.
- Williams, Marlene G., and Jioni A. Lewis. "Developing a Conceptual Framework of Black Women's Gendered Racial Identity Development." *Psychology of Women Quarterly*, vol. 45, no. 2, Feb. 2021, p. 036168432098860, <https://doi.org/10.1177/0361684320988602>.

A Comparison of LSTM and ARIMA For Stock Price Prediction

By Muhammad Ayaan Ameen

Abstract

Artificial intelligence (AI) has dramatically affected how business and finance are conducted. In financial sectors, AI is utilized for stock market analysis, particularly for predicting stock prices using large data sets and time series data to predict future trends without bias compared to prior analytical tools. This paper is a literature review that explores machine learning algorithms used in stock price prediction. The primary datasets in the identified literature focus on the historical, opening, closing, high, and low stock prices for various stocks. After reviewing eight papers that evaluated the AI models Long Short Term Memory and Autoregressive Integrated Moving Average, it was found that ARIMA is a better model for forecasting long-term datasets and LSTM is better for short-term forecasting with short-term datasets. This research helps investors better understand how to optimize machine learning models for stock price forecasting. ARIMA's strength in long-term forecasting makes it viable for long-term trading. LSTM's short-term strength makes it a better option for initial public offerings and other financial assets with few data points.

Introduction

Artificial Intelligence can be classified as machines that simulate human intelligence. Since its introduction to finance in the 1980s, there has been considerable discussion regarding the application of artificial intelligence in the financial sector. AI technologies can analyze vast amounts of data to identify trends and correlations; they also eliminate human emotions and biases from financial outcomes, leading to purely data-driven decisions. They are different from other methods of financial computing as they can simulate human-like reasoning. Countless studies have been done on how AI can be trained and optimized for stock price prediction. Stock price prediction is the act of trying to guess the future value of a stock or other financial instrument using patterns, trends, and other information. Stock price prediction is a fundamental aspect of trading. Traders buy and sell stocks, hoping to predict what the price will move to. Stock traders use news sentiment, financial data, and other forms of information to influence their actions. AI can be trained in a variety of ways to predict stock prices. AI uses algorithms and models to carry out its processes (genetic algorithms, artificial neural networks, support vector machines, etc.) These algorithms/models use data to see patterns and trends and use that to predict future values. Such AI Models have proven to be very accurate for financial trading. Datasets concerning stock price prediction can be historical stock prices, minute-by-minute stock prices, trading volume, social media input, etc.

How is an algorithm chosen?

The machine learning models considered in this review can analyze time series data. Time series data is a sequence of measurements that occur over time. Time series data include

snowfall per year, annual sales, and stock prices. Time series analysis utilizes historical data to establish a foundational understanding of trends. Most of the studies examined will use historical stock prices, opening prices, or closing prices as data for their algorithms.

Time series data is used in time series analysis. This method looks at the historical context of data, offering a foundation to assess how past trends and events can affect future price movements. By evaluating historical data, investors gain valuable insights into potential price behaviors, enabling them to make informed decisions about their investment strategies. This approach involves identifying patterns and trends in sequential data points over time, which aids in forecasting future market movements.

Data sets most commonly used in the studies in this review are historic open/close/high/low stock prices. These are forms of market data that give relevant information about a stock's performance.

What types of algorithms are used for prediction

Long short-term memory (LSTM) is a recurrent neural network that is good at sequence prediction. It has three gates controlling the information flow, solving the vanishing gradient problem in recurrent neural networks.

The forget gate is shown here.

$$ft = (Wf * [ht - 1, xt] + bf)$$

The forget gate determines whether information from the preceding cell state (C_{t-1}) should be retained or disregarded. It uses the input (x_t) and the prior hidden state (h_{t-1}), producing a value between 0 and 1 for each element within the cell state.

The input gate is shown here.

$$it = (Wi * [ht - 1, xt] + bi)$$

$$\Theta_t = \tanh(WC * [ht1, xt] + bC)$$

The input gate decides what information should be added to the cell state. The input gate layer, a sigmoid layer, and the tanh layer generate a vector of new candidate values.

$$C_t = ft * C_{t-1} + it * \Theta_t$$

The cell state is then updated, combining the information retained from the previous cell state (determined by the forget gate) and the new candidate values (determined by the input gate).

The output gate is shown here.

$$ot = (Wo * [ht - 1, xt] + bo)$$
$$ht = ot * \tanh(Ct)$$

Finally, the output gate decides what information should be output based on the updated cell state. It's determined by the current input (x_t) and the previous hidden state ($ht-1$)

Auto-Regressive Integrated Model (ARIMA) is a time series model that predicts future values using past values. ARIMA makes use of lagged moving averages in forecasting.

Results

Research has already shown AI's promising future role in stock price prediction. Research in the Journal "Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction" [1] used Dell stock data of open, low, high, and close prices. This data was inputted into an Artificial Neural Network (ANN) and an ARIMA model. Both models managed to predict the Dell stock price with minimal forecast error. The results concluded that both models accurately forecasted the Future Dell stock price, although the ANN model was slightly more accurate.

ARIMA may encounter challenges with long-term forecasting, particularly in predicting sudden changes such as market crashes. ARIMA models assume that the impact of past shocks diminishes exponentially as you move back in time. This assumption may not be favorable for long-term forecasting, where the effects of shocks can have permanent effects in the future. Long-term stock price forecasting often involves dealing with data that changes, such as shifts in economic policies, technological advancements, or changes in consumer behavior, all of which will impact stock prices and public opinion. These structural shifts might result in long-lasting impacts that stray from the exponential decay assumed by ARIMA models. ARIMA models also use stationary data and stock prices; most economic and financial data is nonstationary. To combat this, ARIMA models need to differ the data to make it stationary. Errors in differencing can result in over-differenced or under-differenced data, which may yield inaccurate parameters and forecasts.

ARIMA models also require lots of data; the Box Jenkin method requires a minimum of 50 data points, which results in ARIMA models tending to work better with longer datasets. The study "The Impact of Lengths of Time Series on the Accuracy of the ARIMA Forecasting" [2] used an ARIMA model to predict undergraduate enrollment using data from 10 Historically Black Colleges/Universities. They used enrollment data from the past 5, 10, 20, and 30 years.

The results found a significant increase in accuracy between 5 years and 10+ years. The conclusion was that 20 years of data was the most accurate. This conclusion did not support the idea that longer datasets result in more accurate forecasts. Still, this disparity could be the introduction of extra time-series data, which results in increased autocorrelation and residual error within the ARIMA model, resulting in increased bias in the machine learning forecasts. Besides that, the study still supported the idea that more years of data leads to higher forecasting accuracy. ARIMA's dependence on long-term datasets hints that their accuracy in forecasting stock prices with minimal data like initial public offerings prices is reduced.

The paper "A Comparative Study of ARIMA and LSTM in Forecasting Time Series Data" [3] addressed the issue of data points and found that the LSTM model is more accurate than the ARIMA model when given less than 36 data points. This is understandable, as the moving average concept in ARIMA would be affected by irregular data. This paper also concluded that ARIMA has a rapid run-time performance. Regardless of how many data points there are, the ARIMA model constantly performed in the range of 0.1 seconds of modeling time. LSTM had a linear relationship with its run time performance and the size of the dataset it used; more data points required more processing time.

A qualm with this paper is that both algorithms' parameters were not optimized and instead kept at standard parameters. The ARIMA p,d,q values were (1,1,1), and the LSTM was kept at 25, unit at 25, gradient descent as an optimizer, and Tanh as an activator. More fine-tuned parameters for both algorithms would be more accurate and could have differing results. The paper notes that LSTM slightly increases in accuracy as the amount of data increases compared to ARIMA. A study by Gosh and others called "Stock Price Prediction Using LSTM on Indian Share Market" [4] used the LSTM model to predict future stock prices of different sectors and banks on the Bombay Stock Exchange. They found that the accuracy of LSTM drastically increases when it forecasts longer, similar to the Arima model, which refutes the findings of [3]. This trait would make LSTM less reliable for intraday trading, where the model might have to forecast in hours, and more reliable for long-term trading. The study "ARIMA vs. LSTM on NASDAQ stock exchange data" [5] compares the length of forecasting on accuracy between the two models and finds that ARIMA is more accurate when doing longer forecasts (1 month, three months, nine months), while LSTM was better when the forecast was for one day. [5] and [3] show that ARIMA is superior with longer datasets, and LSTM models are more accurate than ARIMA models with short-length datasets. [5] also tuned the parameters for the LSTM model, which was not done in [3], yet the conclusions were similar. A future scope of [5] was how adding features to the LSTM could change its forecasting results. The study "LSTM Neural Network Model with Feature Selection for Financial Time Series Prediction" [6] addressed the effect of adding features to the LSTM model. The authors used the open, high, low, and close prices (OHLC) as one of the features. They then also used the OHLC with other features and technical indicators. Lastly, they used the OHLC plus the most essential technical indicators generated by a feature Importance Algorithm. The results stated that only using the OHLC was the most accurate, using the OHLC plus the most critical technical indicators was the second,

and using the OHLC with other features was the least. This shows that putting the required features that strongly correlate to your target variable onto the LSTM model is vital. Putting on too many features that are not important will reduce the accuracy of your forecasts. Another future scope [5] talked about was the effect of using sentiment analysis as a variable. The paper, “Sentiment Analysis as a Factor Included in the Forecasts of Price Changes in the Stock Exchange” [7], used sentiment analysis in the form of information from discussion forums as input for a machine learning model to predict stock prices of video game companies. The paper found that sentiment analysis helps predict stock behavior. This can be used in LSTM models as they are multivariate and can use sentiment analysis as input alongside other features; this is not the same for ARIMA because it is univariate, meaning it can only use one input type. LSTM and ARIMA are powerful models that can be used for stock price prediction; in a scenario with little historical data, LSTM will be more accurate. ARIMA models will be more favorable when you have longer datasets spanning months to years.

A combination of LSTM and ARIMA is valuable to note as well; the study “Prediction of COVID-19 Data Using an ARIMA-LSTM Hybrid Forecast Model” [8] did this; researchers created four different models, an ARIMA, Support Vector Machine (SVR), and 2 ARIMA-LSTM hybrid models and used them to predict COVID-19 cases using given COVID-19 data. The article proposes combining ARIMA and LSTM models to leverage their respective strengths in capturing linear and nonlinear trends in data. Two types of combination models are mentioned; the first is in series, which means the LSTM is used to predict the error of the ARIMA model, which is then added to the ARIMA predictions. The other hybrid model is parallel; ARIMA and LSTM predictions are given weights and combined, typically using equal weight average, a weighted average of error variance, or the reciprocal process of relative error. The paper introduces a novel approach to the parallel combination model, using regression coefficients as model weights. A regression equation is proposed, where the predicted value is a linear combination of LSTM and ARIMA predictions weighted by regression coefficients. The regression coefficient weighting method of the parallel ARIMA-LSTM model made it more accurate than the ARIMA model, the ARIMA-LSTM model in series, and the SVR model. This validates using a hybrid model of LSTM and ARIMA in effective forecasting, especially for volatile non-linear data.

Discussion

The results show that ARIMA is more suitable for long-term stock price forecasting, especially when using large datasets spanning months or years. ARIMA’s reliance on historical data and its ability to model linear relationships make it effective for predicting long-term trends. As a recurrent neural network, LSTM is particularly well-suited for short-term forecasting, especially when handling smaller datasets. LSTM’s ability to model nonlinear relationships allows it to capture intricate patterns in smaller datasets more effectively than ARIMA. Incorporating sentiment analysis (e.g. market sentiment from social media or forums) into LSTM models can further improve stock price prediction accuracy, making LSTM a more flexible

option for handling complex data inputs. ARIMA's limitations are apparent in scenarios with sudden market changes, such as crashes or shifts in economic policies. This is due to its assumption that the effects of past events diminish over time, which may not hold for events with lasting impacts. LSTM takes longer to forecast than ARIMA. Whereas dataset size does not change ARIMA's forecasting time, LSTM's forecast length increases with the size of the dataset.

Combining ARIMA and LSTM models can yield superior forecasting results compared to using either model independently. Using ARIMA to capture linear trends and LSTM to model nonlinear patterns can enhance prediction accuracy.

Relevance

LSTM models could be applied to predict the short-term behavior of newly listed companies, where there is limited historical data. Traditional models like ARIMA are less accurate in this scenario. LSTM's strength in handling short datasets makes it a valuable tool for investors interested in initial public offerings. By integrating sentiment analysis into LSTM models, real-time data from news sources, social media, and discussion forums can be used to predict stock price movements. This can be useful in industries where public sentiment heavily influences stock prices, such as the technology and entertainment sectors.

ARIMA can be leveraged to forecast stock prices for long-term returns. ARIMA's forecasting capabilities can give insight into forecasts, such as how new tax regulations or monetary policy adjustments might affect the market over a longer period.

Hybrid models of LSTM and ARIMA can be used in algorithmic trading systems where decisions are made without human intervention. The integration of hybrid models into these algorithms allows the algorithms to utilize their short and long-term capabilities, allowing for more dynamic and adaptive trading strategies.

Strengths, Weaknesses, and Future Scope

This paper provides an in-depth review of key machine learning algorithms, specifically ARIMA and LSTM, highlighting their distinct strengths and weaknesses. The paper's findings have highlighted real-world applications as they relate models to stock market behavior, trading strategies, and portfolio management. This research is relevant for financial institutions and traders looking to implement artificial intelligence into their practices.

Although this paper emphasizes ARIMA and LSTM, it overlooks other machine-learning models relevant to stock price prediction, such as Support Vector Machines, Random Forests, and Genetic Algorithms. A broader review could provide a more comprehensive overview of AI's role in stock prediction. For deeper insights, explore the multivariate capabilities of LSTM by experimenting with various types of input data. While this paper briefly addresses that, it would be valuable for more research into this feature. Another area of further study is the hybrid models of LSTM and ARIMA, it would be of interest to consider other machine learning models being incorporated into hybrid models.

Works Cited

- [1] Adebisi, Ayodele Ariyo, Aderemi Oluyinka Adewumi, and Charles Korede Ayo. "Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction." *Journal of Applied Mathematics*, vol. 2014, no. 1, 2014, p. 614342, doi:10.1155/2014/614342.
- [2] Qin, Lu, Kyle Shanks, Glenn Phillips, and Daphne Bernard. "The Impact of Lengths of Time Series on the Accuracy of the ARIMA Forecasting." *International Research in Higher Education*, vol. 4, no. 3, 2019, p. 58, doi:10.5430/irhe.v4n3p58.
- [3] Taslim, D. G., and I. M. Murwantara. "A Comparative Study of ARIMA and LSTM in Forecasting Time Series Data." *2022 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, Semarang, Indonesia, IEEE, 2022, pp. 231-235, doi:10.1109/ICITACEE55701.2022.9924148.
- [4] Ghosh, Achyut, Soumik Bose, Giridhar Maji, Narayan Debnath, and Soumya Sen. "Stock Price Prediction Using LSTM on Indian Share Market." *2019 International Conference Proceedings*, doi:10.29007/qgcz.
- [5] Kobiela, Dariusz, Dawid Krefta, Weronika Król, and Paweł Weichbroth. "ARIMA vs LSTM on NASDAQ Stock Exchange Data." *Procedia Computer Science*, vol. 207, 2022, pp. 3836-3845, doi:10.1016/j.procs.2022.09.445.
- [6] Pai, N., and V. Ilango. "LSTM Neural Network Model with Feature Selection for Financial Time Series Prediction." *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud) (I-SMAC)*, Palladam, India, IEEE, 2020, pp. 672-677, doi:10.1109/I-SMAC49090.2020.9243376.
- [7] Wojarnik, Grzegorz. "Sentiment Analysis as a Factor Included in the Forecasts of Price Changes in the Stock Exchange." *Procedia Computer Science*, vol. 192, 2021, pp. 3176-3183, doi:10.1016/j.procs.2021.09.090.
- [8] Jin, Y., et al. "Prediction of COVID-19 Data Using an ARIMA-LSTM Hybrid Forecast Model." *Mathematics*, vol. 10, no. 21, 2022, p. 4001, doi:10.3390/math10214001.

Is There Such A Thing as Too Much Democracy? By Michael Wu

Abstract

The upheavals in today's democracies aren't a mystery. The real mystery is why democracy is so widely celebrated in this world. A man like American Romantic Poet James Russell Lowell would certainly argue that it's because "Democracy gives every man the right to be his own oppressor", but when the United States utterly neglects its people's will on the matter of the Israeli Palestinian conflict, democracy showed it can have some use. This unattainability seems to be indicating that there isn't a thing as too much democracy, but a democratic government can certainly want democracy to be unattainable. To make a decision like this, the government has to firmly believe that it is the right decision for the country to maximize its interest while knowing that ethical condemnation is unavoidable. Hence, when true democracy seems attainable by letting everyone including "non-philosophers" make a difference, it is too much democracy for a country like America to consolidate its position as the world leader. But first, it is important to understand what is a "philosopher". It is a concept suggested by Greek Philosopher Socrates through the dialogue of Plato. A "philosopher" satisfies the 8 virtues-wisdom, courage, temperance, justice, piety, integrity, humility, and love of truth. On the contrary, people who don't own all 8 virtues are "non philosophers", which are also the ones not qualified to handle important affairs. For this reason, having a lot of non-philosophers is a double-edged sword for a government, because they are either easy to make mistakes or easy to manipulate. In the case of Brexit, the UK's group of non-philosophers are rather easy to make mistakes.

Why "Philosopher"?

It is too much democracy when too much power is being distributed to the non-philosophers, which is often the case in direct democracy. In direct democracy, all people eligible, rather than a body of elected representatives, are able to decide on the implementation of policies and laws. Being eligible doesn't usually require having wisdom, so the destiny of certain national affairs would inevitably, to a certain extent, fall into the hands of "non-philosophers"

The UK held a referendum on its EU membership on June 23, 2016. 51.9% of respondents were in favor of leaving, while 48.1% were in favor of staying. Advocates of Brexit argue that the EU undermines the UK's sovereignty, because they think they don't have full control over its laws, regulations, and borders. Autonomy and independence blinds the eyes of the populace because people weren't, to a great extent, aware of the consequences and what they were fighting for. In the post-brexit period, many citizens and political figures realized that it was a mistake. Brexit brought trade barriers, chain issues, and increased costs for businesses. Some individuals who voted for Brexit have expressed regret, citing unforeseen consequences and the complexity of the transition. Surveys by the British Chambers of Commerce (BCC) have consistently shown that a significant proportion of businesses are facing increased costs, delays,

and paperwork due to Brexit. In a 2021 survey, 60% of exporters reported difficulties adapting to the new rules for trading goods. This indicates that Brexit was regretful to many people because there were too many “non-philosophers” in the referendum, either due to lack of education or lack of time for people to think deeply. Putting aside the question of right or wrong, it is supported that there is such a thing as too much democracy when non-philosophers make up a larger proportion of the country’s voter pool as it brought the country to an uncertain orientation.

However, when looking at the 2016 Colombian peace referendum, a revised peace agreement was reached and approved by the Colombian congress to fix the conflict between the Colombian government and the Revolutionary Armed Force of Columbia. Despite the initial rejection of the agreement in the referendum by a narrow margin, a revised peace deal was later ratified. This suggests that the mistake of a democratic process can be fixed by another democratic process. Hence, democracy is always adjusting and correcting itself, so there is no such thing as too much democracy.

In response to that: success is subtle to be measured, so the easy reversal of a previous vote is contemptuous of the public will. Even if the way to do so is to re-examine the public will, the level of seriousness and significance would be greatly diminished, because people might assume that there will be many chances to revisit their decision. When democracy is so frequent that it affects its solemnity, there is such a thing as too much democracy.

Deliberation Day

Frequency of democracy can be an abstract concept, but it comes into shape when it’s being referred to as “deliberation day”. Deliberation day is the idea of calling registered voters together to discuss central issues raised by the campaign before election day. Supporters of deliberation day argue that the implementation of such a process will increase voters’ understanding of their candidates and the rationality of the election. An extra opportunity to discuss politics is certainly beneficial to the final decision making process, but endless deliberation day occupying people’s time requires extremely high opportunity cost, including the productivity that people could bring to work, the joy people could gain from leisure. The economy would collapse if people had no time to work, decreasing the real output of an economy. Excessive attention on current political issues will overly harm the country’s efficiency, prolonging political, business, and cultural activity. All of this was evidenced by Ackerman and Fishkin’s description of how a deliberation day could be possibly operated. If a deliberation day is being held, then it would be seen as a public holiday, which requires pausing economic activities. In addition, endless deliberation days also require a high economic cost. According to Ackerman and Fishkin’s paper, “Successful operation of deliberation day would require the payment of \$150 stipend per person, 15 billion annual cost”. Such tremendous expense on democratic process is beyond the necessary level. If the intervals between deliberation days are too short, there will not be much difference between the topics explored and changes in current events, resulting in these extra deliberation days becoming ineffective

exchanges. Therefore, the economic and opportunity costs of endless deliberation days are vast, even unnecessary.

It can be imagined that supporters of numerous deliberation days would say that increased deliberation day increases civil engagement which makes implementation of policies more efficient. If voters were more well informed with the policies that might be proposed, they wouldn't need to revise the policy again and waste time on parliamentary debates. This covers the cost of time spent on the deliberation day. Thus, spending extra time on deliberation day greatly increases efficiency after the ruling party is determined. With longer time of consideration, there will be less swing voters stagnating the progress of the democracy. However, though high expense before voting can save time for those who are responsible for implementing things, it is not always the case in direct democracy. When a candidate stands out in a vote and represents the country to take action, it is representative democracy, and its efficiency can be increased by resolving matters in advance. However, if the country is voting on a specific matter, for instance, the banning of drugs, endless deliberation day is not going to let the implementation of a policy come faster, because what direct democracy voted for is the direct consequence, not anything that is still negotiable. Therefore, it is too much democracy when deliberation day is endless before direct democracy.

In addition, a question should be asked: "When politics is the most important part of your life, wouldn't it be easier to be angered by others who have a different viewpoint from yours?" Certainly! Arguments would be easily triggered if people's political stances are still different after endless sharing of ideas through deliberation day because we must wonder why the opposition is still not convinced after we have justified our "truth" for endless time. It's similar to the case when it's the 67th time you've reminded your grandma on how to turn on the television. Thus, when politics occupies most of our life, we are more bound to the "truth" we believe in and less trust in our opponent's ability to understand, which might lead to further dispute. Even if humanity can all unite in one political stance through endless deliberation, the existence of democracy itself would be the definition of too much democracy, because there is no need for democracy due the lack of difference between people's viewpoints. Democracy and its complex process would simply become an obstacle that decreases the efficiency for a united country to act.

Stagnation from Polarization

Apart from the controversies inherent in representative democracy, it is also quite capable of creating controversies. In almost all democracies we see in today's world, partisan rivalries are particularly pronounced. Jack Wright, a professor emeritus at Ohio State University who specializes in American politics said that "congress has also become increasingly polarized, hampering lawmakers' ability to work together across party lines to get things done". Indeed, it can be argued that the time it takes for a law or policy to be passed in America is considerable, and it's increasingly harmful to the country's efficiency. On the contrary, another significant portion of the population enjoys the recreational aspects of representative democracy, and they

view those controversies as the embodiment of freedom and equality. However, for a country that aims to maximize its benefit and the wellbeing of its people, the recreational aspect of democracy shouldn't be overly appreciated. Instead, minimizing the time for the country to progress should be the ultimate goal for a government. Therefore, inefficiency arising from polarization between political parties is too much democracy.

According to this article "Protest Erupt in Taiwan as China-leaning lawmaker Aim to Weaken New Presidents". Lawmakers from the Kuomintang actively promoted a bill that requires the new president Lai Ching Te, democratic progressive party politician, to give speeches regularly. Along with that, part of the Kuomintang's demand is to put dishonesty in court as a path to imprisonment. In other words, the Kuomintang want to be able to weaken Lai Ching Te by putting enormous pressure on him. Wen-ti Sung, a Taipei-based nonresident fellow with the Atlantic Council's Global China Hub, said that "It appears to be an attempt to weaken the government's ability to govern effectively". Indeed, with the Kuomintang holding more power on passing legislations, Lai Ching Te and his party will find themselves lack decisive power even though they're the ruling party. And this is just a snapshot of the many controversial bills. Since the Kuomintang won the majority of seats in the legislative body, they have gained advantages in many votes from their numerical advantage. As the newly inaugurated president, Lai may often find his policies for the country's aspiration stagnating.

It is without controversy that one of the most celebrated benefits of democracy is moderation. Having two major parties means that there is greater moderation, so parties can monitor and restrict each other's power. Theoretically, no extreme or absurd proposal can be passed.

However, proposals in favor of the common interest of mankind should not require oversight and moderation. Specifically, peace and absence of violence are the most important for the wellbeing of mankind. After the mass shooting in Uvalde, Texas, a survey from CNN indicated that 64% of US citizens favor stricter gun control, that's around two-thirds of the nation's people. In fact, CNN's polling has discovered this consistent majority since 2016. From the original 60%, support rate of stricter gun laws has increased gradually to 64% today. However, even though the number of mass shootings and the anger of many democrats continued to rise, the republicans won't back down. The second amendment of the US constitution left gun restriction almost impossible, it is a part of America's founding father's legacy, which is to never abandon your freedom for security. The republicans, who seem to have a great extent of tolerance with firearms, utilizes the second amendment as a barrier towards gun control. The reality is that there are people dying every day from gun violence, and mass shootings also seem infinite. As an epitome of the general controversy between liberal and conservative political parties, the second amendment to some extent made peace unreachable, this is too much democracy.

It is true that bipartisan policies might also not be passed. The rural act is a bill that protects electric cooperatives from the risk of losing their tax-exempt status, it is supported by more than two thirds of the house and nearly half of the senate. Seemingly, we can see a bright

future for the rural act. The truth is that about 10,000 bills are introduced on average in a session of Congress, and only about 3% to 4% are passed. Therefore, it can be argued that polarization between parties shouldn't be responsible for the stagnation of proposals.

However, if we look at information outside of the bill itself, then we may be able to find an answer. "We are just coming into that with the 2020 election," said Jack Wright, a professor emeritus at Ohio State University who specializes in American politics. The rural act was brought up within an election year- 2020. The opposition party may oppose bills that they would ordinarily support if it meant giving the Democrats something they could claim credit for. Indeed, the Republicans know that the success of the rural act would increase the democrat's chance in winning the upcoming presidential election, so they rather postpone it and prioritize the Republican's benefit. Therefore, the nature of the stagnation of this bipartisan act is still polarization between opposing parties, it is still based on the principle that their opponent can't look good.

Democracy has its reason for being the most celebrated political system and theory in history. The respect and protection it has for freedom and rights inspire people. However, as civilizations evolve, humanities become more complex and uncontrollable. Therefore, we can't guarantee how leaders use and to what extent they use democracy. The involvement of non-philosopher as voters disturbs a country from making the right decision. When the decision-making process is more dependent on the less intellectual community, there is too much democracy; The idea of a deliberation day is helpful in perfecting the decision-making process, but when deliberation day is our life-long, daily company, there is too much democracy; In democracy, debate and disagreement are fun, but when the debate is too intense and intentions are distorted, we barely reach the decision-making process, thus there is too much democracy.

Works Cited

- ACKERMAN, B., & FISHKIN, J. S. Deliberation Day. *The Journal of Political Philosophy*, no. 10 Accessed 2002, pp, 129-152.
- Schwartzberg, M. Democracy. *The Encyclopedia of Political Thoughts*. Accessed 2015
- Smith, D. A. Direct Democracy. *The Encyclopedia of Political Thoughts*. Accessed 2015
- Plato Republic (C. D. C. R. Reeve, Trans.). Hackett Publishing Company. Accessed 2004
- A quote by James Russell Lowell. (n.d).
<https://www.goodreads.com/quotes/12238-democracy-gives-every-man-the-right-to-be-his-own>
- Statista. Share of people who think Brexit was the right or wrong decision 2020-2024. Accessed July 04 2024 <https://www.statista.com/statistics/987347/brexit-opinion-poll/>
- British Chambers of Commerce. Brexit - British Chambers of Commerce. Accessed April 03 2024 <https://www.britishchambers.org.uk/news/category/brexit/>
- BBC News. Colombia referendum: Voters reject Farc peace deal. Accessed October 03 2016 <https://www.bbc.com/news/world-latin-america-37537252>
- RURAL Act: Why Popular, Bipartisan Legislation Can Face Hurdles in Congress. Accessed November 25 2019
[Cooperative.com.https://www.cooperative.com/news/Pages/rural-act-why-popular-bipartisan-legislation-faces-hurdles-in-congress.aspx](https://www.cooperative.com/news/Pages/rural-act-why-popular-bipartisan-legislation-faces-hurdles-in-congress.aspx)
- Bloomberg, C. W. a. D. W. /. Protests Erupt in Taiwan as China-Leaning Lawmakers Aim to Weaken New President. *TIME*. Accessed May 21 2024
<https://time.com/6980495/protest-taiwan-parliament-reform-bill-lai-dpp-kmt/>

Case Study: Improving the efficiency and output of solar power systems of KCI Solar Array Project By Seonbeom (Joshua) Kim, Andrew Lin



Image from <https://flatlandkc.org/news-issues/every-to-lead-group-building-massive-kci-solar-farm/>
The solar array systems currently under construction. This is how the solar array systems will look when it is completed.

Abstract

The growing dependence on electricity and awareness for sustainability have driven a global shift towards renewable and eco-friendly energy systems. The Kansas City International Airport(KCI) is one of the major facilities participating in this global movement, as they turned to solar energy systems in order to meet a portion of their energy needs. However, the implementation of renewable energy systems brings on challenges in optimizing efficiency and energy production, which is influenced by factors such as placements, maintenance procedures, and system design.

This research paper explores various methods to increase the efficiency of solar panels in the context of the large scale solar array systems in KCI Solar Array Project by analyzing the strengths and weaknesses of each method. The study aims to not pick out the best combination of methods to improve the efficiency of the overall system, but provide insight into each method's individual impact on the large scale solar array system. Thus, this study demonstrates how individual methods play their part in order to improve the overall system and help meet the airport's energy needs.

This paper also serves a broader purpose, as to learn more about the field of engineering in the advancements of clean energy, develop critical thinking about how to capitalize on available resources and technology, and to research ways to provide a more sustainable earth for future generations.

Prologue: History of Harnessing Solar Energy

Among the most revolutionary moments of human history is the discovery of electricity, transforming the way humans harnessed energy as well as daily life. Across the 17th to 19th

centuries, a series of experiments were conducted by Benjamin Franklin, who discovered a connection between energy and lightning with the famous kite experiment; Alessandro Volta, who developed the first electric battery; Thomas Edison, who invented the all-famous incandescent lightbulb; and Nikola Tesla, who developed numerous essential electrical gadgets¹. These scientists have catalyzed the advancement of electricity and its prominence in society.

Stimulated by the industrial revolution in the 18th century, electricity quickly rose as a main medium of energy in human society. During this era, energy was mainly harnessed by fossil fuels for their energy density, ease of accessibility, and high efficiency of electrical production. Fossil fuels surpassed other energy sources by a high margin. However, as the population increased, the energy demand also increased exponentially. To keep up with the skyrocketing demand, fossil fuels have become the main source of energy, resulting in the diminished use of less accessible sources.

Although the method of harnessing solar energy was invented in 1883 by Charles Fritzsche, its efficiency was lower than 1%^{2,3}. Thus, solar was quickly phased out as fossil fuels dominated the energy market. However, the use of fossil fuels was soon met with rising environmental concerns, such as the increase in greenhouse gas emissions and worsening air quality. As a result, more sustainable sources of energy production such as hydropower resurfaced, as well as new technologies to increase efficiency of solar energy production. In the recent years, advancements in the accessibility of sustainable energy sources opened the possibility of reducing fossil fuel consumption. Solar energy quickly became more available, becoming one of the prominent sources of sustainable energy production. Solar energy is non-depletable, unlike conventional fossil fuels. In order to receive that non-depletable energy and convert it to usable energy, the conventional solar panels common-place are developed in 1939, drastically increasing in use in the beginning of the 21st century. It is not an exaggeration to say that solar panels are the icon of sustainability.



Map 1: Installation plan of KCMO Solar Array project³

Basics of Solar PV and Efficiency

Solar panels, or Solar Photovoltaic systems, are made of photovoltaic cells. When the sunlight shines down onto the panels, some of that sunlight energy from the sun gets absorbed in the form of photons by the panels consisting of multiple photovoltaic (PV) cells. Then, the

energy absorbed creates electrical changes that move in response to an internal electric field in the cell, causing electricity to flow⁵. However, not all the solar energy input is converted to usable electric energy by the solar panels due to **numerous limitations**.

For solar panels, PV efficiency is used to gauge the conversion of solar energy to electricity. PV efficiency is governed by the percentage of solar power shining on the PV cells converted to usable electricity. Despite the environmental benefits, harnessing solar energy has a lower energy production efficiency than other sources of energy, which is its fundamental problem that hinders higher electricity production.

Solar cell Efficiency (Maximum):-

$$\eta_{max} = \frac{P_{max}}{E * A_c} \times 100 \%$$

P_{max} = Maximum Power Output (in W)
 E = incident radiation flux (in W/m²)
 A_c = Area of Collector (in m²)

Equation 1:

Equation for measuring efficiency of solar panel

The efficiency of the solar panels is relatively lower than most other sources of energy, however. Solar energy has an average efficiency of 15-22% unlike fossil fuels which range from 33-40% and natural gasses which generally range from 50-60%. Although solar energy can be one of the least favorable energy sources to hardness looking only by its efficiency, the location and climate where the system is located can increase the amount of energy that it can produce. The reason why solar panels are preferably located in clear open space is that this geography allows for increased flexibility in placement and rotation options to collect the most amount of sunlight. The more surface area the sunlight reaches, the more energy the panels can absorb.

The placement of the solar panel system allows for maximized electricity production. Since the electricity production of solar panels is also affected by their proximity to the recipient as resistance increases with the distance traveled by current, they are placed in viable reach of infrastructure and residential areas. Kansas City International Airport (KCI) is a good example of this approach, where the Solar Array Project is currently being initiated and plans are being started to construct photovoltaic(PV) panels around the airport in order to provide power supply to the airport and the greater Kansas City. Because the solar panels are right next to the airport, they will provide substantial power to the airport with minimal loss of efficiency. Although the airport is built away from the city, the construction of solar panels there would eliminate the risk of urban skylines blocking the sunlight for energy production. The 3100 acres around KCI makes it a suitable location to place solar panels.

According to the feasibility study of the Solar Array Project, the panels would produce approximately 500 MW and would power over 70,000 residences. The vast amounts of energy that the panels will produce may mitigate the energy demanded by KCI and the Kansas City residents on a daily basis. The possible benefit of this study is the contribution to meeting the

electricity demands of residences and the airport, while working around potentially unfavorable weather conditions^{3,4}.

Methods to Increase Solar Panel Efficiency

Increasing the efficiency of solar panels involves various methods, ranging from condition maintenance to the placement of the solar panels themselves. Such methods will make solar panels more viable and allow for upscaled energy production. According to the feasibility study for the Solar Array Project KCI⁶, the solar panels placed similarly to the Solar Array Project plans Sacramento International Airport (SMF), Indianapolis International Airport (IND), Denver International Airport (DEN) have experienced other benefits as well:

- Increased revenue from land lease to solar facilities
- Reduced energy costs for airport facilities
- Reduced greenhouse gas emissions
- Increased the Airport's public image by promoting sustainability

The Solar Array Project calls for as many methods as possible in order to maximize electricity production, further reaching sustainable goals for the airport and city. The various steps to improve and maintain the efficiency of solar panels starts with maintaining the condition of the solar panels which includes both manual and automatic methods. Other methods of improving solar panel efficiency include strategic positioning, solar pathfinding, and chemical composition of the solar panel.

Solar Panel Maintenance

One of the most simple ways to improve solar panel efficiency is the periodic maintenance of the solar panels. Cleaning solar panels periodically is crucial, since accumulating matter on solar panels will overtime decrease the solar panels' efficiency. Due to settling particles such as dust, pollen, raindrops, and organic matter, the output of the solar panels can be reduced by 7% in just one month ⁷. This would result in a reduction of 35 MW of energy which is equivalent to the amount of energy consumed by 4,900 residences. To mitigate these losses, system maintenance is needed.

Simple scrubbing



Fig 1: Example of simple cleaning process utilizing manual labor ⁷

Residential solar panels often adopt this simple maintenance technique. Though used in small-scale solar systems, this method can be applied to the environment around KCI, as there are water sources drawing from the KCI water supply to clean the system. To avoid safety hazards such as electric shocks, the solar panels must be disabled. Then, a mild soap solution is sprayed, washing all the accumulated matter on the surface without harming the solar panels themselves. Lastly, the solar panels are dried to remove any water on the surface that may cover the solar panels. The panels are dried with dry cloth or a squeegee that can clear large surface areas with a single sweep.

This cleaning process must be performed during specific conditions when the weather is cooler. Spraying water onto the solar panels when they are heated up can damage the solar panels. When the water is sprayed onto the solar panel's heated glass, the sudden change in temperature shrinks the glass material rapidly. The internal stress that is exerted by the change damages the glass. Additionally, applying corrosive chemicals will also damage the solar panels and reduce their efficiency.⁷

This method is common practice for small-scale residential buildings. However, cleaning the thousands of acres of solar panels near KCI would result in high manual labor costs. Hence, automated cleaning solutions would be a more viable method to increase the efficiency of large-scale solar energy projects.

Automatic cleaning with a robot



Fig 2: Image of SolarCleanO B1 at work in a Solar Array. From <https://www.youtube.com/watch?v=omZRKmPLQcQ>

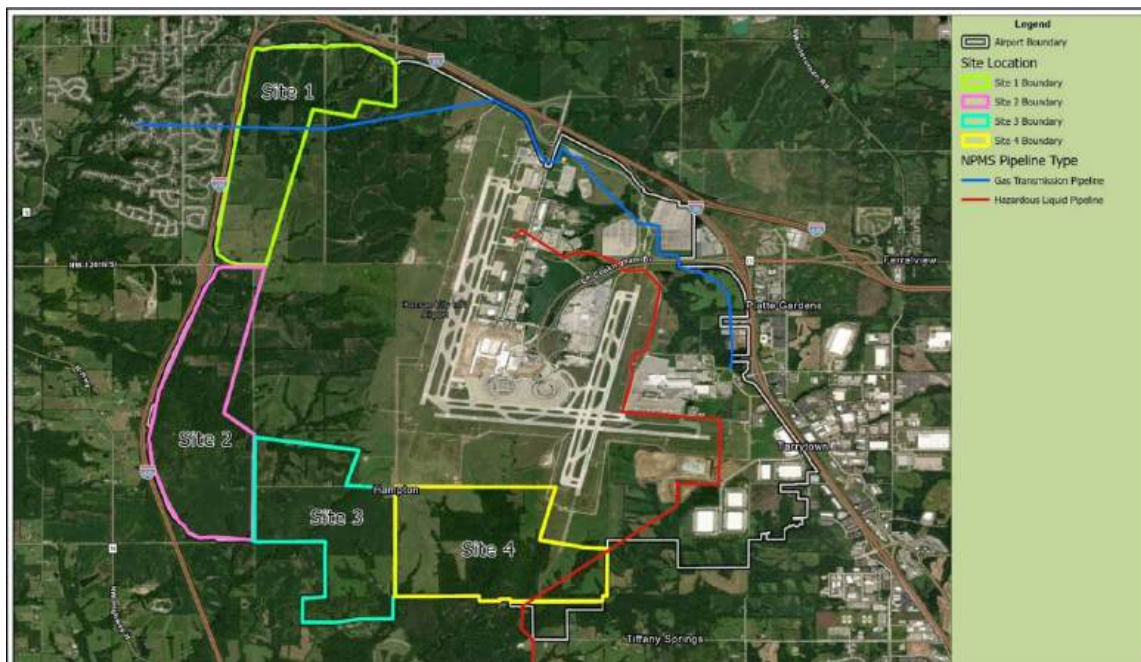
Cleaning robots are ideal for large scale systems. The Solar Arrays Project will have 3,100 acres of solar panels, which makes manual labor an inefficient method. One example can be SolarCleanO B1, a remote-controlled robot that is designed to clean both wet and dry solar panels placed on the ground. It is manufactured by SolarCleanO. The robot's capability to operate under adverse weather conditions enables consistency in electricity production and maintenance of the solar panels. Furthermore, the self-adjusting mechanisms will allow for

flexibility with the solar panel's physical position by self-adapting to the height and inclination of the solar panels. Its rotating brushes clean all sizes of particulate and liquid matter, including bird droppings. This robot is not fixed to the solar arrays themselves and can be stored away at any time.

These autonomous robots can be deployed across large areas of solar panels and increase flexibility for solar panel positioning, reducing maintenance costs and increasing energy efficiency. For large-scale systems, autonomous cleaning is preferred over manual scrubbing, as little to no manual labor will be required for maintenance. While periodic maintenance and cleaning gives a small boost in efficiency, facilitating good placement, positioning, and structural design, will have greater effect on the solar panel efficiency and determine the upper limit of the solar panels' total energy output.

Placement, Positioning, and Structural Design

The placement, positioning, and structural design of the solar array system are the most crucial factors that affect the efficiency of solar panels. They will vastly determine how much sunlight the solar panels will receive and how much maintenance is required, thus affecting how efficient they will be. Solar panels are placed near the airport which reduces energy loss while the current is being transported to the recipient. Their orientation on a vast open terrain enables simple placement patterns (reducing construction costs), allows for increased exposure to the sunlight and light energy collection on a large scale. Also, because they are placed on the ground, it provides long-term stability of supporting structures, lowering their technical difficulty.



Map 2: Annotated View of Kansas City International airport in the planning stage of building the array system⁶



Map 3: Google Earth View of KCI. Shows that the level terrain is suitable for the construction of solar array panels.



Map 4: Google Earth View of KCI. One can see that the land has been cleared for the construction of the solar array system.

Map 2,3,4: Google Earth Capture of areas around KCI. It can be seen that the land is cleared out for the construction of PV panels for KCMO Solar Array Project. The geographic feature around is flat and open, allowing for the construction of low-profile PV panels that maximize safety and energy production.

Fixed Array System

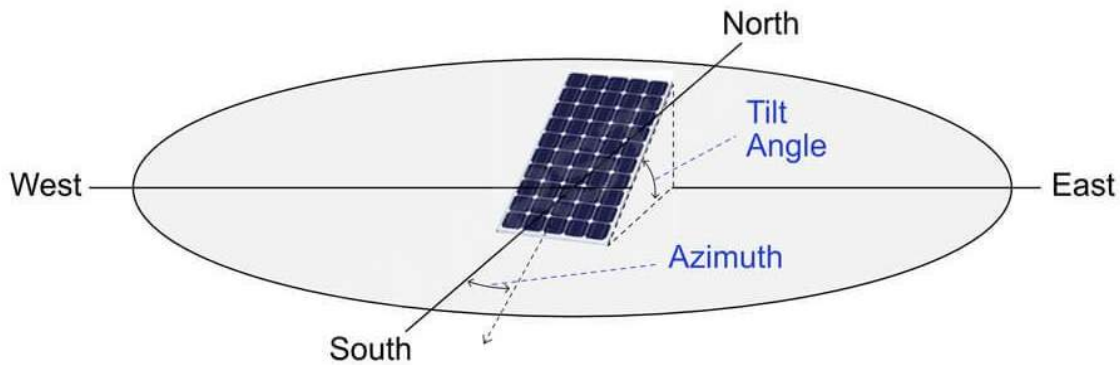


Fig 3 Image of the fixed tilt solar panels' orientation in adherence to the location's azimuth and latitude
 From <https://solardesignguide.com/solar-panel-tilt-and-azimuth/>

As the name suggests, a fixed array system has a fixed angle where maximum energy output is achieved between just a few hours during the day. When sunlight is hitting almost directly at the angle of the solar panels, the most sunlight is captured and converted to electrical energy. While simple, this option is unable to maximize the conversion of sunlight during the daylight hours since it can only remain in one position. However, one way to still maximize the exposure of the panels to the sunlight in order to increase electricity production would be to orient the angle of the fixed tilt arrays relative to the site's latitudinal coordinate and the orientation of the sun. Since Kansas City is situated on the northern hemisphere of the earth, the solar panels should face true south, where the sun would relatively appear.

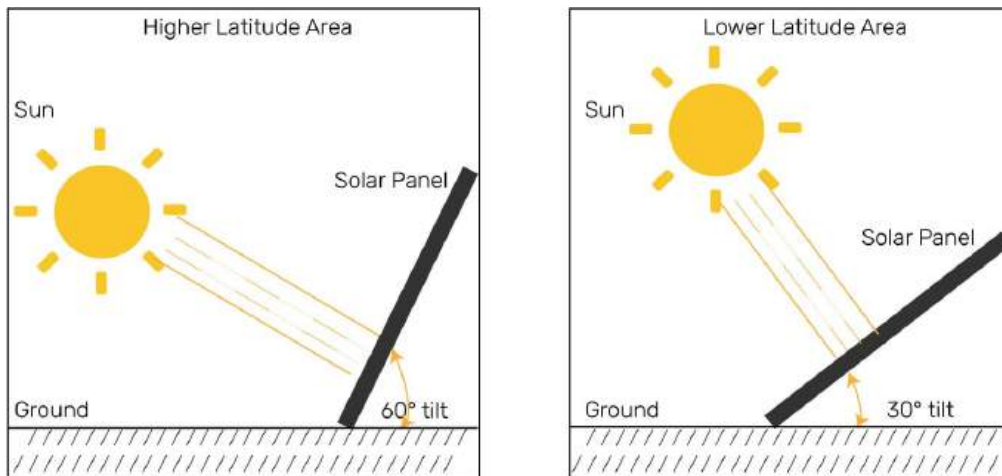


Fig 4: The higher the latitude, lower the angle of the sun.

Image from <https://www.renogy.com/learning-center/how-to-find-the-best-orientation-and-angle-of-solar-panels>

The angle of the solar panels from the horizontal surface is determined by the location's latitude, and specifically can be calculated by adding 15° to the latitude during winter, and subtracting 15° during summer.

Equation 2:

$$\text{Tilt Angle} = \text{Latitude} \pm 15^\circ$$

Equation from Renogy.com

For Kansas City with the latitude of 39.3°N , the optimal angle in the summer is 24.3°N and in the winter is 54.3°N , but this range of angle can be applied by the tracking array system. For the fixed array system, the permanent angle for maximized efficiency can be the average of summer and winter angle, in which for Kansas City, it is 39.3°N , under the assumption that the fixed array system remains fixed year round.

The azimuth also plays a factor in increasing the efficiency of solar panels. Defined as the angular distance measured in degrees clockwise from either true south or true north around the observer's horizon (the solar arrays in this case) and the sun. For the solar arrays to have maximized efficiency, the azimuth of the solar arrays must be near 0° from the true south or 180° from the true north.

The simulation that has been run via the program Photovoltaic Geographical Information System from the European Commission yields a similar result to 39.3°N . It is known that the solar panels use crystalline silicon material in order to increase efficiency; the installed peak PV power in (KWp) is 285 KWp under the assumption that no energy loss has occurred; the mounting position of the solar array system is free standing, as they are installed on the fields around the KCI. When the input of 39° slope and 2.5° azimuth relative to the south is applied, the simulation has produced a maximized result in power output and solar ray exposure. With these inputs, the yearly PV energy production is predicted to be 476,290 MWh, and the yearly in-plane radiation, or as known as exposure to solar radiation, would be $1,828 \text{ MWh/m}^2$. Particularly, the yearly PV energy production would increase by 2.97% compared to 462,550 MWh previously predicted by the feasibility study of Solar Array Project in KCI. This increase in efficiency would power up to 1,924 more residences, signifying a significant impact of azimuth orientation and slope of the solar arrays in increasing their efficiency and power production.

Another simulation using the PVWatts calculator from National Renewable Energy Laboratory (.gov) yields a much higher result than the first simulation. It is known that the Under the assumption of a level surface, no energy is lost, and module type is standard, the inputs of around 39° slope and 182.5° azimuth relative to north are inserted. The simulation calculated that the annual production is predicted to be 764,662.9 MWh, 65.31% more than 462,546 MWh, thus being able to power more than 42,297 residences.

In conclusion, the azimuthal orientation and the angular orientation of fixed tilt arrays have a significant impact on the efficiency of solar array systems. Although calculations may vary, these factors can increase the production from 3% to 65%.

The feasibility study acknowledges the fact that PV system modeling simulation software may display different energy outputs. Thus, simulating with different programs may provide a general range of the increase in production that slope and azimuth may allow for the solar array system.

Tracking Array System

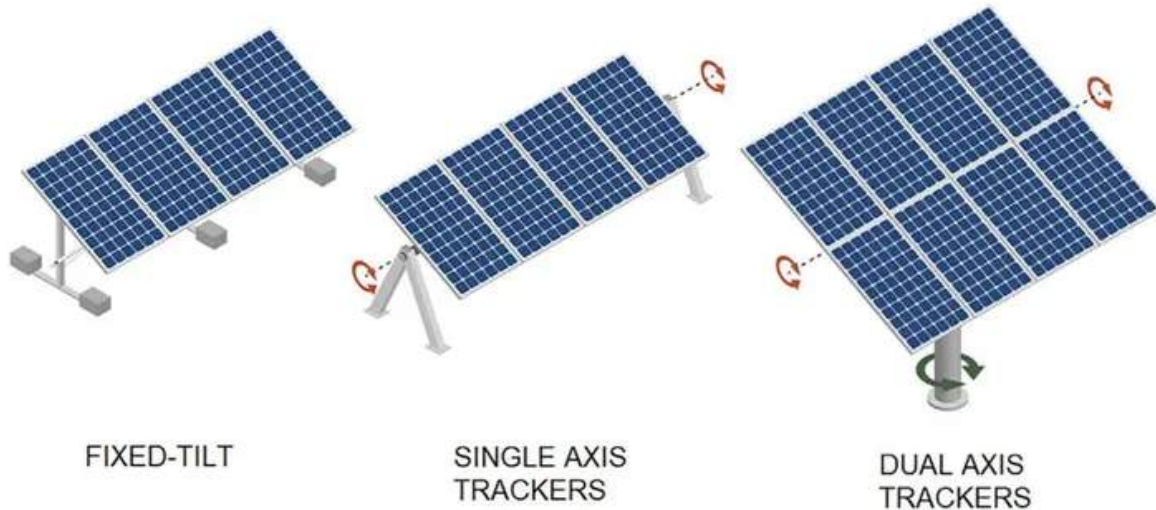


Fig 5: Fixed tilt(left) is stationary. Single axis tracking arrays (middle) rotates vertically. Dual axis rotates both vertically and horizontally.

Image from Renewsys, <https://www.renewsysworld.com/post/solar-trackers>

The tracking array system is a more efficient way of managing the solar panel's structural design. This allows for angular flexibility and more exposure to the sun's rays. When the sun rises from the east, the sun's incident angle changes. That angle of incidence differs throughout the day, and throughout the seasons. In summer time close to winter solstice, one can see the sun's zenith at a higher angle than the winter times near winter solstice. And throughout the day, the sun reaches its zenith mid-day. In order to mitigate the risk that occurs due to the frequently changing angles of the sun, tracking systems are often implemented on both large and small scale applications using different operating systems.

The most simple tracker uses heat from sunlight to shift chemicals around within the solar panels and uses gravity to move the solar panels in the proper direction. These passive trackers move based on the lighting of the sun, causing the center of gravity of the solar panels to shift in the direction of the sun's rays.

More complex tracking systems move the solar panels by hydraulic or/and motor driven mechanisms so that the panels are continuously perpendicular to the sun⁸. The tracking system uses different operating systems in order to track the sun throughout the day and possibly throughout the season. These so-called active trackers use motors and sensors to detect the position of the sun and properly angle the solar panels in the direction of sunlight.

The solar panels change their tilt in either single axis or dual axis tracking systems. The single axis system allows the solar panels to tilt in a horizontal direction to track the sun moving throughout the day. The dual trackers move in both horizontal and vertical direction, tracking the sun's orientation throughout the day and throughout the season. The dual-axis trackers are more effective at capturing sunlight due to more flexibility, thus allowing for maximized electricity

production. On the other hand, the single axis trackers are less flexible, but applicable on a larger scale. Because the single axis trackers have fewer moving parts and are less complex, they have a longer lifespan, and are cost effective. In terms of cost, complexity, and efficiency, the single-axis tracking system is a balance between dual-axis tracking systems and fixed systems, making it a more viable option in the context of airports like KCI.

Overall, the tracking system is an option to improve the efficiency of the solar panels, with single-axis trackers increasing their electricity production by 10-20% and dual-axis trackers by 20-30%⁹, making the output of 500MW possible, in which the feasibility study for the Solar Array Project has taken into consideration the partial implementation of tracking array system. However, the implementation of tracking array system comes with a tradeoff, sacrificing stability and cost effectiveness. If the system ever breaks down or malfunctions, the downtime will result in loss of production until the system is fixed.

Chemical composition

Semiconductor material

Various semiconductive materials are used to create photovoltaic cells for solar panels. Crystalline silicon material is most commonly used for industrial-grade solar panels, having around 27.6% single cell energy conversion efficiency¹⁰. As researchers begin to discover new materials that can potentially increase this efficiency, the perovskite mixed with silicon is emerging as a viable option which yields 36.1% efficiency¹⁰ by a single cell itself, this could be the next best material for large-scale applications.

There are also multi-junction versions of solar panels, one of which is Gallium Arsenide (GaAs). With 47.6% cell efficiency¹⁰, such solar panels could be seen being used industrially in the future. These panels capture different wavelengths of light by letting certain wavelengths pass through different layers and capturing all different types of sunlight and converting it to electrical energy.

Anti-reflective Coating

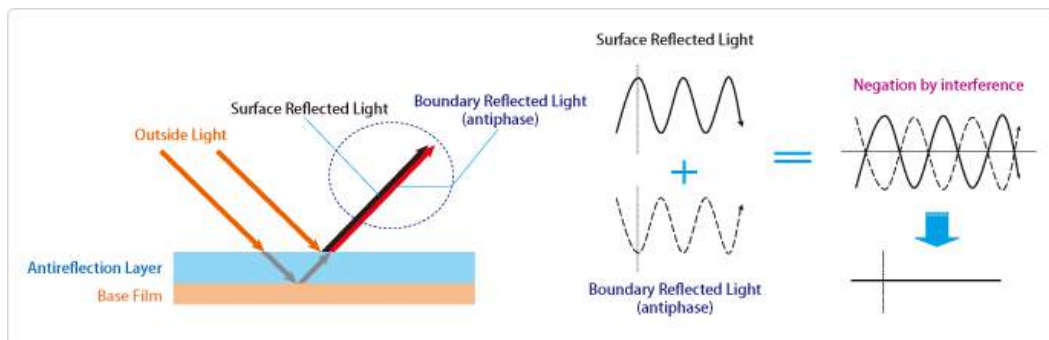


Fig 6: The light reflected by base film experiences interference with light reflected by the anti-reflection layer.

Image from

<https://physics.stackexchange.com/questions/663769/how-do-anti-reflection-coatings-in-solar-cells-make-light-stay-inside-a-solar-cell>

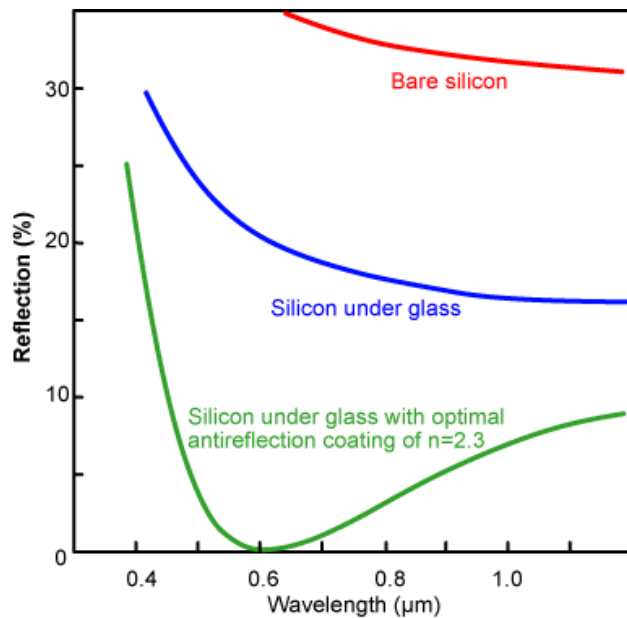


Fig 7: Graph representing the percentage of light reflected off of different material compounds. It can be seen that silicon with anti-reflective coating reflects the least amount of light overall.

Image from <https://www.pveducation.org/pvcdrom/design-of-silicon-cells/anti-reflection-coatings>

A special coating can be applied to the surface of the solar panels in order to maximize the conversion of sunlight to electrical energy. With anti-reflective coating, the reflection of the sun's rays into the atmosphere is minimized and the absorption of sunlight by the solar panels is increased. This results in higher energy production. According to Nazar, "more than 35 percent of incident light is reflected away from the panel's surface before it can be converted to usable energy"¹¹. Anti-reflective coating is made of dielectric material and works as a phase cancellation layer, resulting in reduced reflections. 75% more sunlight is absorbed with such a coating, resulting in 3% higher energy production¹¹. More absorption of sunlight results in reduced glint from the solar panels. Glint coming from the solar panels often act as a safety hazard to both incoming aircrafts and taking off, in which reduced glint improves safety as well.

Conclusion

Improving the efficiency of solar arrays requires the careful consideration of numerous factors. Ranging from the simple process of scrubbing to the intricacies of solar panel composition, this research explored various mechanical, chemical, and positional factors that can increase solar panel efficiency. For one, the day-to-day efforts of maintaining the solar panels reduce declines in efficiency and keep the solar panels durable. Moreover, the placement of these units enables more exposure to sunlight and increased electricity production. Lastly, The chemical composition of solar panels ensures more sunlight is converted to electrical energy, serving as one of the largest factors in improving solar panel efficiency.

It also cannot be ignored that some methods must be chosen over others. Fixed array systems versus tracking array systems serve as an example. While fixed arrays cannot move in conjunction with the moving sun throughout each day and seasonal cycle, they require less maintenance and have increased durability, providing more energy stability and ensuring a consistent supply of electricity to the residences and the airport. On the other hand, the tracking array system moves according to the sun's position in the sky and captures more sunlight, but sacrifices mechanical stability and consistency. Implementing both systems is necessary to maintain a stable flow of energy production and also keeping efficiency as high as possible. As experiments continue to be performed and more discoveries are made, researchers are constantly assessing the benefits and drawbacks of each method and determining advantageous combinations of these methods.

The purpose of this paper is not just to perform a case study regarding a real-life scenario, but also to learn more about the field of engineering in the advancements of clean energy, develop critical thinking about how to capitalize on available resources and technology, and to research ways to provide a more sustainable earth for future generations. As young, aspiring scientists and engineers strive to make the world a cleaner place to live, this endeavor of increasing solar energy efficiency and meeting sustainable development goals opens up paths for better technologies and quality of life.

A social purpose of this study is to proliferate this knowledge and research to younger audiences who aspire to pursue engineering with sustainability in mind. As the solar panels near Kansas City International Airport continue to be built, it is imperative to raise awareness of the efforts of pursuing the sustainable development goals authored by the UN, and encourage the policymakers, educators, and the general public to take initiatives in reaching these goals. This research aims to serve as a catalyst for the next generation of efficient and sustainable energy production.

Acknowledgements

Seonbeom Kim

Research Lead

Author of Case Study

Andrew Lin
Project Supervisor
Technical Advisor

Shane Oh
Engineering SME
Editor

Special thanks to the following Authors for their simulation software and their contributions to research data:

The European Commission - Photovoltaic Geographical Information System

The National Renewable Energy Laboratory (.gov) - PWWatts Calculator

Many thanks to the following experts, corporations, and organizations for supporting information, data, and figures in this research:

The Just Energy Group, Inc., a natural gas/electricity retailer

Re-Volv, a Solar Energy company based in San Francisco, CA

City of Kansas City Missouri

Colin Thompson, <https://medium.com/@colin.thompson.solar>

Energy.gov: US Department of Energy

Tony Carrick, Roxanne Downer and Allaire Conte

Federal Aviation Administration

Summer Xia, Co-founder & Marketing Director at Shenzhen Solarstone Technology Co., Ltd.

Rupali Nazar: Mtech. Student, Department of Electrical and Electronics Eng.,
Lingaya's University, Faridabad, (India)

Univ of Michigan Center for Sustainable Systems

Michael Tinnesand from American Chemical Society

Works Cited

- ^[1]Electricity explained. (2024, March 26). U.S. Energy Information and Administration. Retrieved June 8, 2024, from <https://www.eia.gov/energyexplained/electricity/>
- ^[2]The history of solar. (n.d.). Re-volv. Retrieved June 10, 2024, from <https://re-volv.org/get-involved/education/historysolar/>
- ^[3]Solar array project at KCI. (n.d.). City of Kansas City. Retrieved June 10, 2024, from <https://www.kcmo.gov/city-hall/departments/city-manager-s-office/solar-farm>
- ^[4]Thompson, C. (2024, April 29). Thriving solar installations in Kansas City: Mastering Missouri's sustainability despite cloudy hurdles. Medium. <https://medium.com/@colin.thompson.solar/thriving-solar-installations-in-kansas-city-mastering-missouris-sustainability-despite-cloudy-a5f6b13c6857>
- ^[5]How does solar work? (n.d.). Department of Energy. Retrieved June 15, 2024, from <https://www.energy.gov/eere/solar/how-does-solar-work>
- ^[6]Westside airport solar siting and feasibility study [PDF]. (2022). City of Kansas City. <https://www.kcmo.gov/home/showpublisheddocument/8600/637907921580570000>
- ^[7]Carrick, T., Downer, R., & Conte, A. (2024, May 21). How to clean solar panels: 8 steps to improve efficiency. CNN. Retrieved June 20, 2024, from <https://www.cnn.com/cnn-underscored/home/how-to-clean-solar-panels>
- ^[8]Plante, J. A., Barrett, S. B., DeVita, P. M., & Miller, R. L. (2018, April). Technical guidance for evaluating selected solar technologies on airports. In M. Lawrence & P. W. Magnotta (Eds.), Technical guidance for evaluating selected solar technologies on airports. Federal Aviation Administration. Retrieved July 12, 2024, from <https://www.faa.gov/sites/faa.gov/files/airports/environmental/FAA-Airport-Solar-Guide-2018.pdf>
- ^[9]Solar PV energy factsheet [Fact sheet]. (n.d.). Center for Sustainable Systems, University of Michigan. Retrieved July 15, 2024, from <https://css.umich.edu/publications/factsheets/energy/solar-pv-energy-factsheet>
- ^[10]Best research cell efficiencies [Illustration]. (n.d.). National Renewable Energy Laboratory. <https://www.nrel.gov/pv/cell-efficiency.html>
- ^[11]Nazar, R. (2015). Improvement of efficiencies of solar panels using different methods. *International Journal of Electrical and Electronics Engineers*, 07(01), 12-16. <http://www.arresearchpublication.com/images/shortpdf/117.pdf>
- ^[12]Evergy to lead group building massive KCI solar farm [Image]. (n.d.). FlatlandKC. <https://flatlandkc.org/news-issues/evergy-to-lead-group-building-massive-kci-solar-farm/>
- ^[13]Solar cell efficiency equation or formula [Image]. (n.d.). RF Wireless World. <https://www.rfwireless-world.com/calculators/Solar-Cell-Efficiency-Calculator.html>
- ^[14]SolarCleano B1 [Video]. (2021, October 6). Youtube. <https://www.youtube.com/watch?v=omZRKmPLQcQ>
- ^[15]Which direction should your solar panels face? [Illustration]. (n.d.). Solar Design Guide. <https://solardesignguide.com/solar-panel-tilt-and-azimuth/>

^[16]Comparison of surface reflection from a silicon solar cell, with and without a typical anti-reflection coating. [Illustration]. (n.d.). PVEducation.
<https://www.pveducation.org/pvcdrom/design-of-silicon-cells/anti-reflection-coatings>

^[17]Solar trackers [Illustration]. (2021, June 18). RenewSys.
<https://www.renewsysworld.com/post/solar-trackers>

^[18]How do anti-reflection coatings in solar cells make light stay inside a solar cell? [Illustration]. (n.d.). Physics Stack Exchange.
<https://physics.stackexchange.com/questions/663769/how-do-anti-reflection-coatings-in-solar-cells-make-light-stay-inside-a-solar-ce>

Navigating Vietnam’s Legal Landscape Amidst AI and Digital Economy Growth

By Dieu-Linh Do

Abstract

Vietnam is at a crossroads in its journey toward modernization and international integration, driven by the rapid expansion of the digital economy and advancements in artificial intelligence (AI). With the digital economy projected to contribute 30% to the national GDP by 2030, Vietnam is emerging as a significant player in AI research, development, and application within the ASEAN region and globally. This paper examines the transformative effects of AI and the digital economy on Vietnam's legal system, the associated challenges and opportunities, and the essential legal frameworks needed to harness these technologies effectively while safeguarding fundamental rights.

Introduction

Natural Language Processing (NLP), a subfield of Artificial Intelligence, is machine-learning-driven and can effectively extract information from big data, contextualize it, and improve itself based on feedback and user input. While AI has existed since the 1940s, it was not until 2015 that the field received significant investment from technology corporations. The increasing use of neural network models, combined with the availability of large datasets and computational power, has accelerated the pace of NLP development.

Starting in 2016, with the introduction of the “Google Translate” service, Google shifted from statistical machine translation to neural machine translation. Older models for machine translation relied on analyzing large parallel corpora of translated texts and heavily depended on human-curated linguistic rules and resources. In contrast, the newly developed neural models can learn to translate directly from the source to the target language without explicit linguistic rules, capturing complex semantic and contextual relationships. This has resulted in more fluent and natural-sounding translations.

In 2017, a team of researchers from Google Brain and the Google AI Language team presented the Transformer, a breakthrough in language model technology. The Transformer processes natural language input in parallel rather than sequentially, unlike its predecessor (Recurrent Neural Networks, or RNNs). This reduces both the training time and computational requirements, enabling the development and deployment of Large Language Models (LLMs). A critical component of these LLMs is the use of ‘synthetic data.’ In many domains, the lack of high-quality, diverse, real-world data—due to sensitivity and confidentiality concerns—hinders the training and optimization of data-hungry machine learning models. Synthetic data, which is artificially generated and replicates the statistical properties of real data, offers a solution. It allows researchers and practitioners to capture a wider range of scenarios, edge cases, and diverse perspectives, thereby increasing the size and robustness of databases used for machine training. Moreover, synthetic data is particularly valuable when working with confidential or

sensitive information, as it enables continued training without violating privacy or data protection regulations.

In this era of accelerating technological advancements, countries like Vietnam are poised to harness the transformative potential of AI to drive significant changes in their digital economies. The rapid development of AI-powered technologies is set to enhance human experiences by personalizing interactions, improving service delivery, and streamlining processes across various sectors. As AI becomes more integrated into daily life, consumers will benefit from more efficient and tailored experiences. However, this transformation also brings challenges, including ethical concerns and the need for robust regulatory frameworks to ensure the responsible use of AI. Understanding how Vietnam can leverage AI for digital transformation will illuminate both the potential benefits and challenges that lie ahead for its society and economy. In this rapidly evolving landscape, how can Vietnam ensure that AI integration not only drives economic growth but also enhances the quality of life for its citizens?

Background on Digital Economy and AI in Vietnam

In recent decades, Vietnam has made significant progress in industrialization and modernization, leading to substantial economic growth and increased global integration. In 2022, Vietnam experienced the highest growth in the digital economy in Southeast Asia, with a 28% increase from \$18 billion in 2021 to \$23 billion. The country's digital economy is expected to grow at twice the pace of GDP through 2030 (19% vs 9%). The digital economy and AI are central to this transformation, with the potential to revolutionize sectors such as healthcare, education, and urban management. In 2021, Vietnam ranked 59th globally in government AI readiness. Moreover, according to the Vietnam-Australia Intelligence Network (Vietnam-Australia AI), AI could be incorporated into state management, enhancing investment policies, city planning, national digital transformation, and e-government initiatives. Currently, Vietnam's digital economy is expanding rapidly, fueled by government initiatives, rising internet penetration, and a tech-savvy population. According to the Ministry of Planning and Investment, the digital economy is expected to account for 30% of national GDP by 2030, up from 10% in 2020. Investments in digital infrastructure, e-commerce, fintech, and AI technologies are driving this growth.

Vietnam has made substantial progress in internet penetration and connectivity. By 2023, over 70% of the population had internet access, compared to just 34% in 2010. The expansion of high-speed broadband and mobile internet services has supported the growth of the digital economy. The anticipated rollout of 5G technology by 2025 will further enhance connectivity and enable new applications of AI and Internet of Things (IoT) technologies. The size of Vietnam's e-commerce market is estimated to have reached \$20.5 billion in 2023, representing a 25% growth compared to 2022, according to the Vietnam E-commerce and Digital Economy Agency. Prominent players such as Shopee, Lazada, and Tiki have captured significant market shares, while local businesses are increasingly adopting digital platforms to engage with consumers. The incorporation of AI in these platforms enhances the shopping experience by

using algorithms to provide personalized recommendations, making it easier for consumers to find products that match their preferences. AI-powered chatbots, now common on most e-commerce apps, offer real-time customer support by answering inquiries and resolving issues instantly, which greatly increases customer satisfaction. Moreover, AI algorithms assist companies in analyzing customer behavior, enabling them to optimize inventory management, tailor marketing strategies, and ensure consumers have access to the most relevant products. The government aims for 55% of the population to shop online by 2025.

Additionally, the fintech sector is experiencing rapid growth, particularly in digital payments. Mobile wallets such as MoMo, ZaloPay, and VNPAY have seen a surge in adoption, especially during the COVID-19 pandemic. This trend is accelerating the shift toward cashless transactions. The State Bank of Vietnam reported a 34% increase in non-cash payments in 2021. Fintech companies are also expanding into digital banking, peer-to-peer lending, and blockchain technologies.

By 2030, Vietnam's government aims to develop predictive systems in science and technology, with the goal of creating roadmaps for technological innovation in priority sectors. These efforts will specifically target areas such as health technology, artificial intelligence, and energy storage.

AI Opportunities in Vietnam

The Vietnamese government has shown strong support for the development of AI technologies by providing corporate income tax exemptions and import tax exemptions to foster the growth of the AI industry. Vietnam has approved a National Digital Transformation Program to be completed by 2025, with the goal of promoting digital transformation in governance, the economy, and society more broadly, as well as establishing Vietnamese technology firms as global players. The program outlines several key objectives for 2025, including placing Vietnam among the top 50 countries in information technology, the top 70 on the E-Government Development Index (EGDI), and the top 35 in the Global Innovation Index (GII). As part of this effort, AI startups are granted a four-year corporate income tax exemption, followed by a 50% tax reduction for the next nine years. Additionally, these startups benefit from import tax exemptions on equipment, machinery, and raw materials used for AI research and development.⁴ Several organizations and institutions have been established to assist startups in their development, such as the National Innovation Center (NIC). Founded under Decision 1269/QĐ-TTg, the NIC is a central component of Vietnam's strategy to develop its domestic startup sector. The NIC offers training and support for young enterprises and collaborates with international technology companies like Amazon and Google. Another example of the government's commitment to innovation is the National Technology Innovation Fund (NATIF), a government agency under the Ministry of Science and Technology. NATIF provides grants and preferential loans for research and development, innovation, and technology transfer.

⁴ The Government in Vietnam - Decree 13/2019/ND-CP

Corporations have leveraged this robust government support to position themselves at the forefront of AI innovation in Vietnam, including companies like FPT Corporation and VNG Corporation. FPT.AI, a product of FPT, was established in 2016 and focuses on developing technologies in natural language processing (NLP), computer vision, and machine learning. One of its key innovations is ‘FPT.AI Reader - FaceMatch,’ a state-of-the-art deep learning algorithm that compares portraits on personal identification documents (ID cards, driver’s licenses, passports, etc.) with real-time facial images for verification purposes. Meanwhile, VNG Corporation focuses on improving user experiences in gaming, social networking, and e-commerce. A notable innovation from VNG is Zalo AI’s Kiki, the first Vietnamese voice assistant that supports multiple regional accents. Despite being officially launched in December 2020, Kiki quickly reached 100,000 installations in cars by August 2022, with that number increasing to 500,000 by September 2023, and surpassing 600,000 installations just three months later. While these advancements in AI hold great potential for driving economic growth in Vietnam, they also raise concerns about data security and regulatory compliance.

Challenges and Risks for the Development of AI in Vietnam

As Vietnam advances in digital transformation, a robust legal framework is essential for governing AI and the digital economy. AI machine learning leverages Big Data to enhance its responses and functionality. However, AI’s reliance on machine learning algorithms, which draw insights from vast datasets, raises concerns about the persistence of human biases embedded in these data. Datasets used to train many AI systems, including large language models (LLMs), often lack diversity and reflect the perspectives and demographics of their data collectors and annotators, who are predominantly white and male.⁵ This lack of representativeness can potentially lead to the reinforcement of unfair biases and stereotypes in AI outputs. Moreover, AI systems tend to perform more effectively in languages with larger amounts of available training data, such as English, Chinese, and Spanish. This creates a significant gap between AI systems’ effectiveness across different languages. The lack of linguistic diversity in AI training data can lead to systems that reflect and amplify the biases of dominant languages and cultures, marginalizing or excluding underrepresented languages, cultures, and communities.⁶

Vietnam’s Law on Cybersecurity, enacted in 2018, provides a foundation for data protection, but further measures are needed to address the complexities of AI and Big Data. Strengthening data protection mechanisms and establishing comprehensive guidelines for data usage and storage are necessary to safeguard citizens’ privacy. Since AI and digital technologies rely heavily on data, concerns about data security and privacy are growing. Robust data protection laws and cybersecurity measures are essential to prevent data breaches and misuse.

⁵ U.S. Government Accountability Office. (2021). Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities. GAO-21-519SP.

⁶ Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (Technology) is Power: A Critical Survey of "Bias" in NLP. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 5454-5476.

AI-generated content also introduces new challenges for intellectual property (IP) laws. Determining ownership of AI-generated inventions and creative works requires rethinking traditional IP frameworks. The Vietnamese government needs to establish clear guidelines on IP rights for AI-generated content to foster innovation while protecting creators' rights.

Global Regulatory Approaches to the Legal Liabilities of AI-Generated Works

The European Union has proposed a comprehensive regulatory framework for AI systems through the draft AI Act. Article 28 of the proposed legislation requires developers of high-risk generative AI models, such as text-to-image generators, to provide a detailed account of the training data used in developing their systems. This transparency requirement aims to enhance accountability and mitigate potential harms from biased or low-quality training data. As stated in the European Commission's Proposal for a Regulation on AI, "the description of the training, validation, and testing data sets used for the AI system shall be sufficiently detailed to allow the assessment of the compliance of the training, validation, and testing procedures with the requirements of [the] Regulation." Additionally, the draft AI Act includes specific provisions for moderating AI-generated content, such as deep fakes, mandating that "appropriate measures" be taken to "prevent or minimize the risks" posed by such synthetic media. These regulations aim to balance the advantages of AI innovations while ensuring safeguards for European citizens.

In contrast, Japan has taken a more permissive stance on the legal liabilities of AI-generated works. According to Japan's Copyright Act, the use of copyrighted material for the training of AI systems is allowed, even if the original source was obtained illegally. Article 30-4 of the Act states, "the reproduction of a work for the purpose of training an AI system" is permitted "without obtaining the authorization of the copyright holder." Furthermore, the Act specifies that the commercial exploitation of AI-generated works is also permissible, provided that the AI system was lawfully trained. This policy reflects Japan's strategic focus on cultivating a thriving AI ecosystem and encouraging innovation, even at the expense of certain copyright protections.

The United States, on the other hand, maintains a human-centric approach to copyright law. The U.S. Copyright Act defines eligible works as "original works of authorship fixed in any tangible medium of expression." This requirement for human authorship was demonstrated in the *Naruto v. Slater* case, where the U.S. Court of Appeals for the Ninth Circuit ruled that a copyright claim brought by a monkey over selfies it had taken was dismissed because "the Copyright Act does not permit a monkey (or any other animal) to claim copyright." This strict interpretation of authorship as a human attribute remains the prevailing legal position in the United States.

Canada, which shares a similar stance with the United States, leaves the interpretation of "author" to the courts, as the term is not clearly defined in Canada's Copyright Act. However, a recent decision by the Canadian Intellectual Property Office (CIPO) granted copyright protection to a painting titled *Suryast*, which was co-authored by an AI system. CIPO's decision stated that "the work satisfies the requirements for copyright protection as an 'original' work under the Copyright Act, notwithstanding the involvement of AI in the creative process." This landmark ruling suggests a potential shift in Canada's legal recognition of AI-generated creative works.

In contrast, Australia's patent law explicitly states that a patent may only be granted to a "person" who is the inventor or derives title from the inventor. However, in the DABUS case, the Australian Federal Court ruled that an AI system could be considered an "inventor" for the purposes of patent law. This decision represents a significant step toward the legal personhood of AI, with the court stating, "the inventor ... may be an artificial intelligence system or device." The Indonesian government has embraced the use of AI systems to digitize and streamline its public services, helping to overcome bureaucratic obstacles that have hindered efficient service delivery. Specifically, the Indonesian Government Regulation on the Implementation of the Electronic System and Transaction Law allows for AI-powered automation to replace civil servants in performing routine administrative tasks. This represents a concerted effort by the Indonesian government to leverage AI technology to improve the accessibility and responsiveness of public services for citizens. However, Indonesia's Copyright Act still maintains a requirement for human originality for a work to be eligible for copyright protection, which may pose challenges for the legal recognition and treatment of AI-generated creative outputs. In contrast to many other countries, South Africa has taken an even more progressive approach, becoming the first nation to grant a patent directly to an AI system named DABUS. According to the country's Patents Act, an "inventor" is defined as "the actual deviser of the invention." Based on this interpretation, South Africa's Commissioner of Patents determined that DABUS, an AI system, could be recognized as the inventor. This decision aligns with South Africa's broader strategy of leveraging AI to address socioeconomic challenges, as outlined in the country's Intellectual Property Policy.

Potential Regulatory Approach for Vietnam

Vietnam could adopt a more holistic approach to addressing the legal challenges posed by the rapid development of artificial intelligence (AI). This approach should focus on three key objectives: promoting continued innovation and growth in the AI industry, ensuring the responsible use and development of AI systems, and safeguarding Vietnamese citizens from potential harm. Economically, there are risks such as AI-driven job displacement and income insecurity if automation is not properly managed. Socially, biased or opaque AI decision-making could perpetuate unfairness and discrimination. In education, overreliance on AI assistants could erode critical thinking skills. Furthermore, privacy and security vulnerabilities in AI systems present risks of data breaches or misuse.

A comprehensive legal and regulatory framework must carefully balance these competing priorities—encouraging AI development while mitigating social, economic, and educational risks. This can be achieved through collaboration among policymakers, industry leaders, and civil society stakeholders. Vietnam's regulatory framework could draw inspiration from the European Union's proposed Artificial Intelligence Act, which aims to create harmonized rules to ensure the trustworthiness of AI systems. Key provisions in the Vietnamese framework should include requirements for transparency, safety, and accountability, particularly for high-risk AI applications.

⁷The framework should also provide clear legal guidelines on the status and ownership of AI-generated creations, including both creative works and inventions. To encourage innovation, the law could recognize AI systems as potential co-authors of creative works. This would require amending the Copyright Law to explicitly address AI-generated content and its eligibility for protection. For example, in 2019, Canada amended its Copyright Act to acknowledge AI-generated content, stating that "a work in which copyright subsists shall not be ineligible for copyright protection solely because a computer program was used in the creation of that work." Vietnam could take a similar approach, ensuring that AI-generated creations are adequately protected under copyright law, thereby fostering further innovation in the AI industry⁸. This precedent has set the stage for a more nuanced approach to addressing the legal status of AI-generated creations. The Vietnamese government could follow suit by amending its Copyright Law to define "AI-generated content" and establish a framework for AI co-authorship. This would involve introducing clear criteria to determine the level of AI involvement in the creative process and guidelines for allocating copyright ownership and royalties between AI systems and human contributors⁹.

Encouraging robust public-private collaboration and multi-stakeholder engagement will be crucial for the success of these legal and regulatory efforts¹⁰. The Vietnamese government should engage with the local AI research community, technology companies, and creative industries to understand the unique challenges and opportunities they face. Establishing a multi-stakeholder advisory board to provide input on the development of AI-related policies and regulations, as well as fostering partnerships between the government, academia, and the private sector, can help to ensure that these legal frameworks are informed by diverse perspectives and tailored to the needs of the Vietnamese context. For example, the advisory board could include representatives from organizations like the Vietnam Artificial Intelligence Association, leading tech companies operating in Vietnam, and prominent creators and copyright holders in the local creative industries. By incorporating diverse stakeholder input, the government can develop more robust and responsive policies that balance the interests of all parties involved.

Furthermore, the Vietnamese government should invest in AI education and skill development to build a strong foundation of AI-related competencies among government officials, policymakers, and the broader workforce.¹¹ As the world's future developments are likely to surround the innovation of AI, getting the general public to have a grasp understanding and know how to use AI responsibly and effectively would be a huge advantage for the nation.

⁷ European Commission. (2021). Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts.

⁸ Copyright Act, R.S.C., 1985, c. C-42 (Canada).

⁹ Bridy, A. (2016). The evolution of authorship: Work made by code. *Columbia Journal of Law & the Arts*, 39(3), 395-401.

¹⁰ OECD. (2019). Recommendation of the Council on Artificial Intelligence. <https://www.oecd.org/science/recommendation-on-artificial-intelligence.htm>

¹¹ World Economic Forum. (2018). The Future of Jobs Report 2018. <https://www.weforum.org/reports/the-future-of-jobs-report-2018>

To increase the knowledge of citizens about AI, it involves introducing AI-related courses and programs in universities and vocational schools, as well as providing targeted training and upskilling opportunities. Collaboration with international organizations and experts can help the government stay informed on global best practices in AI governance and education. By investing in AI literacy and expertise across various sectors, the government can better equip itself and the broader public to navigate the evolving legal and regulatory landscape surrounding AI-generated content.

Conclusion

Vietnam stands at a pivotal moment to establish a proactive and comprehensive legal framework for AI-generated creative works. As the nation's economy continues to flourish alongside a burgeoning technology sector, it has a unique opportunity to craft tailored legislation that addresses the complexities of AI while avoiding the pitfalls of retrofitting existing copyright laws. By focusing on authorship, ownership, and the rights of AI systems, Vietnam can create a legal landscape that not only protects creators but also encourages innovation.

The Vietnamese government's commitment to fostering technological advancement—evidenced by support for startups and investment in tech infrastructure—provides an ideal backdrop for these necessary legal reforms. By drawing insights from international models, such as the European Union's proposed Artificial Intelligence Act, Vietnam can implement a balanced regulatory approach that promotes innovation while ensuring ethical considerations are met. This includes adopting a risk-based framework that categorizes AI applications based on their potential impacts, thereby establishing appropriate safeguards.

Moreover, fostering public-private partnerships will be essential in shaping an inclusive dialogue that considers the voices of all stakeholders, ranging from creators to legal experts. By prioritizing educational initiatives that enhance understanding of the legal implications of AI in creative sectors, Vietnam can empower individuals to navigate this evolving landscape effectively.

Ultimately, by taking decisive action now, Vietnam can position itself as a leader in the region regarding the intersection of technology and creativity. This proactive legal framework will not only address current gaps in intellectual property law but also lay the foundation for a vibrant, innovative future where both technology and creativity can thrive harmoniously. Through alignment with international standards and active engagement in global discussions on AI governance, Vietnam can enhance its competitiveness in the global market while fostering a responsible and sustainable growth trajectory for its creative industries.

Works Cited

1. Blodgett, S.L., Barocas, S., Daumé III, H. and Wallach, H., 2020. Language (Technology) is Power: A Critical Survey of "Bias" in NLP. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp.5454-5476.
2. Copyright Act, R.S.C., 1985, c. C-42 (Canada).
3. OECD, 2019. Recommendation of the Council on Artificial Intelligence. Available at: <https://www.oecd.org/science/recommendation-on-artificial-intelligence.html> [Accessed 21 Sep. 20].
4. Oxford Insights, 2023. Government AI Readiness Index, 49.
5. Soc et al., 2022. The Government in Vietnam - Decree 13/2019/ND-CP.
6. U.S. Government Accountability Office, 2021. Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities. GAO-21-519SP.
7. World Bank, Google, Temasek and Bain & Company, 2023. Data.
8. World Economic Forum, 2018. The Future of Jobs Report 2018. Available at: <https://www.weforum.org/reports/the-future-of-jobs-report-2018> [Accessed 21 Sep. 2024].

Small Modular Reactors: Advancing Electrical Grid Flexibility and Reliability

By Janak Bhuta

Abstract

As it becomes increasingly difficult to construct large scale nuclear power plants, Small Modular Reactors (SMR) become more viable as a source of nuclear energy. This paper examines how policy and economic incentives can promote the adoption of SMRs in the U.S. electrical grid and compares their benefits and drawbacks as compared to conventional nuclear power plants. With the U.S. electrical grid facing significant challenges, such as aging infrastructure and higher demand, SMRs present a viable solution to unify the electrical grid due to their adaptability and modularity. In order to optimize the energy produced, this paper also analyzes and compares the use of different coolants and their roles in the SMR. Based on the results in this paper, the integration of SMRs into the U.S. electrical grid will be better than that of a traditional power plant for the following reasons: lower fuel cost, lower carbon emissions, ability to adapt energy output, and compatibility with other renewable energy sources.

Introduction

This paper answers the following question: How can policy and economic incentives promote the adoption of SMRs as a key technology in the development of the United States electrical grid, and how do these benefits compare to those offered by a traditional power plant?

Like most other countries, the United States faces the dual challenge of meeting increasing energy demands while transitioning to low-carbon sources. In 2023, natural gas made up approximately 32% of the country's energy consumption, followed by coal at 26%, and nuclear energy at 22%. Other renewable sources, such as solar and wind, have recently grown but still constitute a relatively small fraction of the energy mix (U.S. Energy Information Administration). Around the world, coal, and natural gas have also been declining since 2019, leaving room for other clean energy sources to grow (Energy Institute Releases Statistical Review of World Energy).

In 2023, carbon emissions grew by 1.1%, further increasing the emissions to 410 million tons, causing global concern (Iea). This increasing global concern about climate change has made it imperative to move away from fossil-fuel-dependent sources of energy, as shown with movements such as the Paris agreement, with countries like the U.S., China, and India joining ("List of Parties That Signed the Paris Agreement on 22 April). These sources are one of the main contributors to greenhouse gas emissions, which are driving rising temperatures, among other environmental impacts. Nuclear energy is a carbon-neutral alternative that could supply large amounts of electricity.

Nuclear power plants generate electricity through a process known as nuclear fission. In this process, the nuclei of Uranium-235 atoms are split apart with a neutron, releasing a large amount of energy in the form of heat. This heat is then used to boil water, producing steam that

drives turbines to generate electricity. Despite this efficiency, the construction of new plants has been hindered by a variety of factors.

The high initial costs of setup have been one of the major deterrents in the establishment of nuclear power plants. The costs involved are not only with respect to building the facility itself but also the nuclear fuel and implementation of strict safety measures. Also, long periods taken to plan, license and build nuclear power plants have made them unattractive to investors.

Within the last decade, a new approach has emerged in the field of nuclear power: Small Modular Reactors (SMRs), which have several advantages over conventional nuclear power plants. SMRs, which have been around since the 1950s, were prominently used for naval use, but have just been growing in terms of popularity for mainstream energy production.

Their design supports fabrication in factories and subsequent delivery to sites for deployment; this approach reduces the need to perform massive on-site construction tasks. This could substantially shorten the building process by decreasing construction costs.

Also, SMRs have been designed and conceptualized as inherently safer and more flexible than typical reactors. Their design provides advanced safety features, such as a simpler design, larger fractions of coolant, and a reactor core with lower power (Small Modular Reactors Explained). These safety features can also be applied to different uses, from electricity generation through district heating to desalination. With this, the SMRs can offer a real pathway toward more capacities in nuclear energy in the United States, enabling a cleaner and more sustainable energy future.

1 Literature Review

1.1 An Overview of SMRs

Among the most important advantages of SMRs is the potential to lessen construction time and costs. The modular design allows for components to be prefabricated in factories, reducing extensive on-site construction and eliminating delays (Locatelli, 2014). This can accelerate new nuclear capacity deployment and help meet growing energy demands. The standardized components used in the SMRs will further drive economies of scale in reducing costs (Locatelli, 2014).

Beyond the economic advantages, SMRs offer greater flexibility. Because of their size, they can be used in more locations, which include inaccessible areas or places with poor grid infrastructure. That makes this reactor suitable for very different applications, ranging from electricity generation to district heating and even desalination. In addition, SMRs can be scaled up or down depending on energy needs, making them more adaptable to changes in market conditions (Locatelli, 2014).

Safety is of primary importance in the nuclear energy sector, and SMRs have the capability to introduce an improved level of safety. Most of the designs of SMRs feature passive safety systems independent of electrical supplies from external sources, which reduce the occurrence of accidents and enhance the general reliability of the plants. These passive safety

systems involve using natural forces, like gravity and natural circulation, to shut down the reactor without human intervention. Furthermore, the reduced scale of SMRs can limit the consequences of any accident that may occur, as the amount of radioactive material released would decrease (Locatelli, 2014).

However, while SMRs offer a number of advantages, there are also challenges to be overcome. First, there is considerable upfront investment in developing and licensing systems. New reactor designs tend to be expensive to develop, while regulatory approval processes can be lengthy. Moreover, the economic competitiveness of the SMR compared to traditional nuclear power plants is yet to be debated. Factors such as economies of scale, construction costs, and operational expenses will play a crucial role in determining the competitiveness of SMRs (Locatelli, 2014).

Despite these challenges, the potential benefits of SMRs make them one of the most prospective technologies for the near future. Their ability to provide clean, reliable, and flexible energy sources can contribute to a more sustainable and resilient energy landscape. As research and development progress, SMRs may play a significant role in meeting the growing demand for energy while reducing greenhouse gas emissions.

1.2 Licensing and Economics/Policy

The effective integration of SMRs into the United States electrical grid will depend very heavily on technological advances, regulatory support, and economic incentives. Above all, it is crucial for policymakers and industry participants to work together toward creating an optimal environment both for development and deployment.

One important determinant will be an enabling regulatory regime. Well-designed and clear licensing regimes can drastically reduce deployment time and cost for the SMRs (Kumar). A standard set of safety guidelines and performance characteristics will ensure clarity in the regulatory environment to investors and developers.

This can also be facilitated by economic incentives that favor the adoption of SMRs. Subsidies from the government, tax credits, or loan guarantees reduce the investment cost at the very initial stages of the project. In addition, carbon pricing mechanisms would make low-carbon energy sources, such as SMRs, more economically attractive in relation to conventional fossil-fuel-based power plants (Kumar).

1.3 Transportation

There is an effective dependence of road transport feasibility for SMR modules on specific weights, lengths, and widths due to regulatory constraints. Exceeding the limits raises the transportation cost substantially and may also imply special permits or alternative modes of transport. Module subdivision, where the SMRs can be subdivided up to 3 times in any dimension to fit transport limits (Lloyd), helps to cope with transportation problems, although it introduces some extra costs connected with site assembly.

Transportation costs can be very high and can eat into the overall economic viability of SMR projects. While the modular design greatly reduces extensive on-site construction, costs for transportation of modules to various sites must be carefully assessed, as the modular design can reduce construction costs by 10%-20% depending on the project (Lloyd). Distances involved, mode of transportation, and the availability of infrastructure will influence transportation costs. This is mainly because of the large amount of local construction and transport of heavy components.

To promote the deployment of SMRs, policy makers can offer a range of incentives to build transportation infrastructure as well as enhance overall economic viability through tax incentives, regulatory streamlining, public-private partnerships, among other policy options. Such infrastructure investments in road, rail, and port transportation networks could enable efficient module transport of SMRs. All these can lead to transport cost reductions, further raising the economic viability of SMRs (Lloyd). Tax incentives comprise tax breaks or credits during transport, lowering the overall cost and giving an incentive for its adoption. These projects could appear more viable to investors and developers. Streamlining the regulations will mean fewer administrative burdens, and the project timelines will be shorter; all these steps can reduce the cost and time to deploy an SMR.

1.4 Challenges Regarding Deployment

There are many potential applications of SMRs in a microgrid. To start, the architecture of the SMRs is modular; hence, they can be flexibly deployed at different sites, such as in very remote locations or within urban centers. Additionally, this technology has potential application in enhancing grid resilience through supplementary electricity supplies when demand is at its peak or when renewable supplies are variable (Michaelson). SMRs' capability to work at different power levels can help balance supply and demand within the microgrid.

However, the incorporation of SMRs into renewable energy microgrids also presents several challenges. One very important aspect involves the control and coordination of SMRs with other parts that make up the microgrid, such as renewable energy sources, energy storage systems, and traditional sources of power (Michaelson). It is vital to develop strong control strategies to ensure grid stability and optimization in terms of energy management.

Another challenge lies in sizing and optimization issues related to SMR-renewable microgrid components. The appropriate capacity of SMRs, renewable sources, and energy storage systems must be derived by careful analysis of load profiles, renewable energy availability, and system dynamics (Michaelson).

Furthermore, the remote operation of SMRs in microgrids necessitates advanced instrumentation and control systems. Ensuring reliable monitoring, fault detection, and autonomous operation capabilities is crucial for safe and efficient operation (Michaelson).

1.5 Power Output

Because they can operate at a range of power outputs, SMRs are better positioned than traditional reactors to integrate with intermittent renewable sources. An SMR whose power output varies in step with fluctuations in wind or solar output reduces pressure on the grid and cuts the number of fossil fuel-fired peaking plants (Locatelli, 2018). The latter aspect of flexibility is particularly important for areas with high amounts of renewable energy because SMRs could serve as a reliable source of baseload power, accommodating the variability of renewable power generation.

Voltage regulation is another important feature of grid integration. The SMRs should be capable of maintaining the level of voltage within a narrow range of stability and not allow it to sag or surge too high, which may lead to damage to equipment and disrupted power quality. Frequency control is another important feature for grid stability; SMRs are supposed to assist in maintaining the grid frequency within the specified range (Locatelli, 2018).

Moreover, SMRs have the potential to be coupled with cogeneration facilities to optimize their energy production. The production of certain valuable byproducts, such as hydrogen, employs surplus thermal energy derived from these SMRs. Hydrogen can be stored for later use or integrated into the energy grid, thereby affording greater flexibility and alternative sources of income (Locatelli, 2018). Combining cogeneration facilities with SMRs will improve the efficiency and financial viability of these projects.

1.6 Comparison to a Traditional Power Plant

From an energy generation perspective, SMRs can generate a significant amount of electricity, specifically 300 Mwe (megawatt electrical) at 45-50% thermal (Locatelli, 2018), or over 120,000 homes for a year (What Is a Megawatt). The amount would, of course, depend on the type of design adopted for the SMR, but in general, they do have generating capacity in the hundreds of megawatts. This may be juxtaposed against normal power plants that produce in the range of gigawatts (Locatelli, 2018). On the other hand, SMRs have the advantage of smaller, more modular units that can be deployed in many different locations and scaled up to meet particular energy demands.

While the conventional power plants have larger production capacities of energy, the SMRs provide a more flexible and efficient solution to integrating clean energy onto the grid. The capability for load following, operation at a range of power output levels, and production of valuable byproducts make them a very important factor toward a sustainable and resilient energy future (Locatelli, 2018).

Thus, the advantages afforded by SMRs compared to conventional power plants are multilayered. They could avail of a more flexible and modular approach to energy generation, deploying on lesser scales and integrating more seamlessly with existing grids. Its increased safety and lesser environmental impact will also make it more favorable to policy and decision makers, as well as the general public (Kumar).

1.7 Nuclear Waste from SMRs

It was found that SMRs can generate up to 30 times more nuclear waste per unit of energy produced than normal nuclear reactors, as there is a lot more neutron leakage, leading to more nuclear waste (Krall). This increased volume is a problem both in terms of storage and disposal. Special materials, such as metallic sodium, metallic uranium, and uranium tetrafluoride and coolants in SMR designs can also produce more chemically complex waste streams, which would require specialized handling and disposal methods, needing to be developed (Krall).

Moreover, even though SMRs produce less radioactive waste by total radioactivity, specific composition may be much more problematically valued. Various designs of SMRs can generate more long-lived radioisotopes, which increase the long-term risks connected with the disposition of these wastes (Krall).

These findings raise one important concern about the environmental impact of SMRs and also the long-term costs associated with nuclear waste management. However, it has to be put into perspective that even SMRs emit considerably less greenhouse gas than conventional fossil fuel-based power plants. Thus, while the nuclear waste challenge is a factor to consider, the possible benefits offered by SMRs in relation to clean energy production and mitigation of climate change should not go unnoticed.

It should also be noted, though, that the particular quantity of nuclear waste to be produced by an SMR would depend on several factors, including reactor design, type of fuel, and operation conditions. In light of this, it should be noted that some new designs for SMRs might reduce the amount of waste produced using new technologies or innovative fuel cycle strategies (Krall). Further research and development are important in order to understand all possible environmental impacts and to develop adequate solutions to manage such waste.

While nuclear waste challenges remain an issue, SMRs are, nonetheless, one of the most promising options for clean energy generation. Smaller in size, with a modular design, and featuring improved safety features, they become very attractive to deploy at multiple locations. Accordingly, with investment in research and development regarding the challenge of nuclear waste, the full range of SMRs' potential can be realized—a very important sustainable energy source.

Results

Economically, SMRs can reduce construction costs by up to 20% through prefabrication and modular designs, but transportation challenges, particularly module subdivision, can increase costs (Fig. 1). Additionally, high initial licensing and development expenses pose financial barriers to widespread adoption, as since the modular design of the SMRs can change the safety functions of the remaining components as well as altering the hazards and risks presented, this makes it harder to get a license.

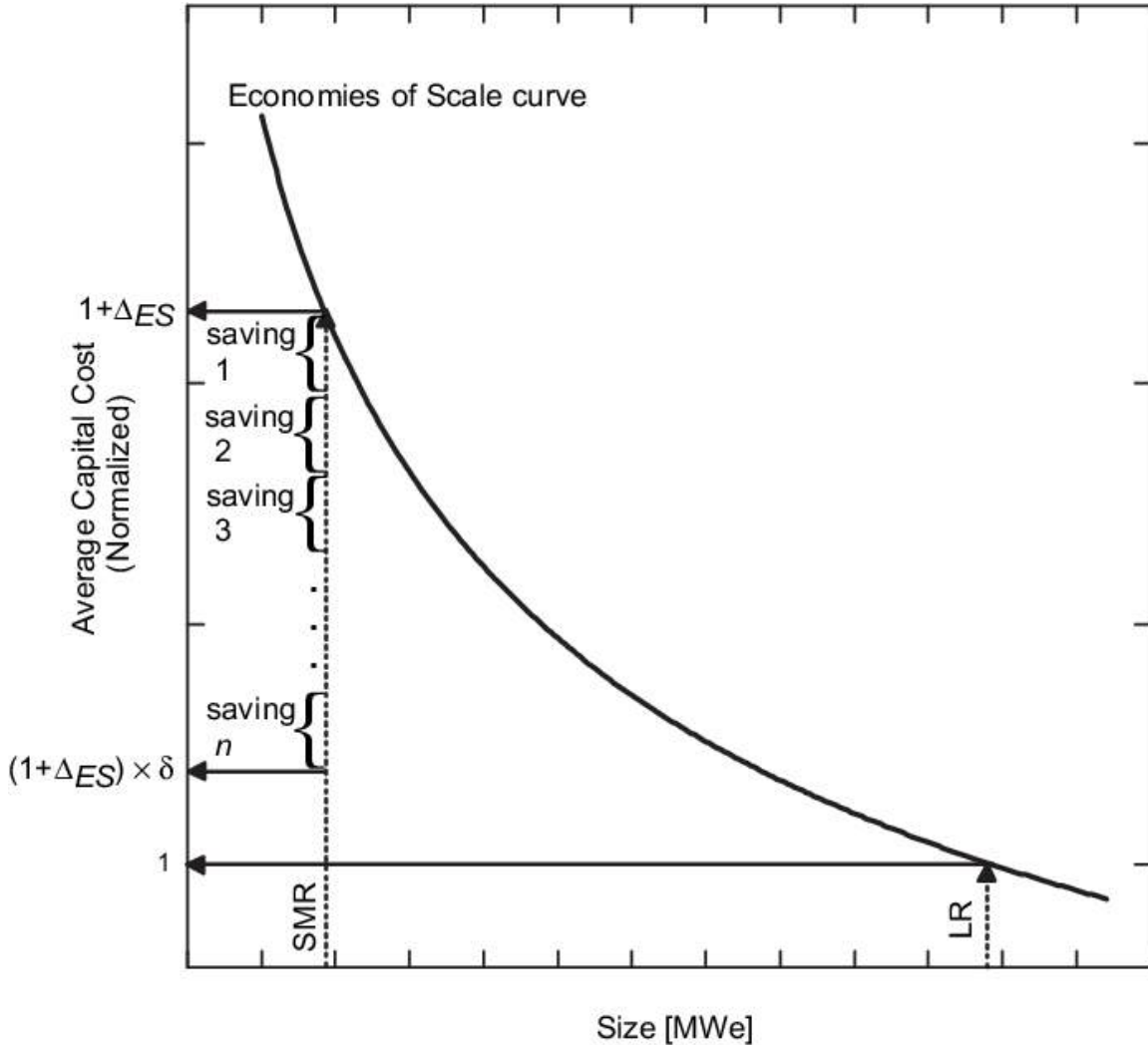


Figure 1: Cost versus Size (Locatelli)

The above graph highlights how, as size in terms of power increases for nuclear reactors, the average capital cost decreases. This means that as these reactors get bigger, they are making less return on investment they receive. Additionally, the graph also emphasizes that SMRs, marked by the dotted line, make the most average capital cost out of any nuclear reactors.

In terms of safety, SMRs incorporate passive safety systems, smaller reactor cores, and larger coolant fractions, significantly lowering the risk and impact of accidents compared to traditional reactors. However, their smaller size may reduce economies of scale for operational efficiency.

From a grid integration perspective, SMRs provide load-following capabilities, allowing them to adjust power output based on fluctuations in renewable energy sources like wind and solar. SMRs can generate up to 300 MWe at 45-50% efficiency and support cogeneration for hydrogen production, offering flexibility in energy production (Fig. 2).

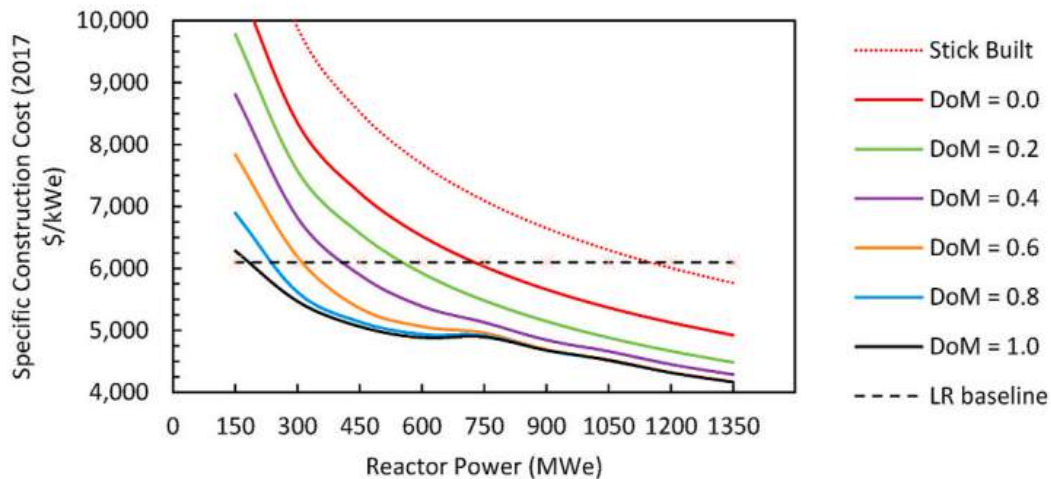


Figure 2: Overnight Construction Cost (OCC)¹² versus Power Output for multiple Degrees of Modularity (DoM) (Lloyd)

The above graph shows that, for multiple different degrees of modularity, the construction cost decreases as the reactor power goes up. It also highlights that for large reactors, the construction cost stays the same no matter how high the reactor power goes.

However, SMRs generate up to 30 times more nuclear waste per unit of energy compared to conventional reactors due to increased neutron leakage, as there is less time for neutrons to get absorbed — neutrons that reach the boundary of the core are more likely to escape than to be absorbed because of the smaller diameter (Krall). This waste, often more chemically complex, requires specialized disposal methods since SMRs are incompatible with existing nuclear waste disposal technologies. Despite these challenges, SMRs produce significantly fewer greenhouse gasses than fossil fuel power plants.ⁱ

Analysis/Discussion

SMRs present a balance of significant advantages, such as large amounts of clean energy and transportability, as well as notable challenges, like disposal of nuclear waste. Economically, while their modular designs reduce construction costs by up to 20%, high initial licensing and transportation costs remain substantial barriers. These costs, particularly those associated with module subdivision and transport logistics, could offset the benefits of modular construction. As such, economic incentives like subsidies, tax breaks, and streamlined regulatory processes are essential to make SMRs competitive with both traditional nuclear plants and renewable energy sources.

In terms of safety, SMRs offer clear improvements over traditional reactors. Their passive safety systems and smaller reactor cores significantly lower the risk and impact of accidents.

¹² OCC — The cost of construction if no interest was incurred during construction, as if the project was completed "overnight."

However, these benefits come at a cost—smaller reactors limit economies of scale, potentially raising operational and maintenance costs compared to larger, conventional reactors. Since the SMRs have to refuel every 1.5-3 years, the energy has to be spread out more to more reactors, leading to more maintenance (Michaelson). While SMRs' safety features make them ideal for deployment in a variety of settings, including urban and remote areas, this may be offset by higher energy production costs on a per-unit basis.

The flexibility of SMRs in grid integration is a major strength. Their load-following capabilities allow them to adjust power output based on fluctuations in renewable energy sources, such as wind and solar. This positions SMRs as valuable assets in grids with high renewable energy penetration, where power supply variability is a major concern. Additionally, SMRs' potential for cogeneration, particularly in hydrogen production, adds economic value. However, the lower power output of SMRs (300 MWe) compared to traditional reactors means that multiple units may be required to meet large energy demands, potentially reducing their cost-effectiveness when scaled up. Even though this might be the case, they are still much more cost-effective than conventional gas-fired plants, as the generation cost, nuclear fuel accounts for only about 12% of generation costs, whereas it accounts for approximately 70-80% of the generation cost for a conventional gas-fired plant.

A significant drawback of SMRs is their production of nuclear waste. SMRs can generate up to 30 times more waste per unit of energy compared to conventional reactors due to increased neutron leakage. Additionally, this waste is more chemically complex, requiring specialized storage and disposal methods. While SMRs emit far fewer greenhouse gasses than fossil-fuel plants, their long-term environmental impact from waste remains a critical issue. Further research into advanced fuel cycles and waste management technologies will be necessary to mitigate this challenge.

Conclusions

SMRs are a promising way to advance nuclear energy, efficiently and expansively answering growing global energy demands. The modular construction reduces construction time and cost; more important, though, it enables deployment even to the most remote or least-developed areas. Additionally, they are designed to facilitate the addition of renewable energy sources like wind, solar, and others; they will thereby be in a position to add value toward a more flexible and resilient energy grid. These advantages place SMRs in a critical position to further decarbonization and answer challenges of modern energy grids.

Yet, SMRs cannot go without challenges. High front-end licensing and development costs and logistics transport issues are some of the major economic drawbacks. One of the most serious disadvantages of small modular reactors is that they produce more atomic waste per unit of energy produced compared to conventional reactors. This happens to be chemically more complex and requires a different means of disposal compared to conventional waste emanating from older reactors. This raises concerns about the long-term environmental impact, emphasizing the need for continued research into advanced waste management solutions.

Economically incentivized installations through tax credits, subsidies, and lean regulations can make SMRs more viable, even beyond traditional power plants and renewable energy. By addressing economic and environmental challenges, SMRs are able to become a cornerstone of energy production in the United States, being able to provide clean energy through the electrical grid.

Acknowledgements

Thanks to Nathaniel Tricarico for mentorship and peer review.

Works Cited

- “Energy Institute Releases Statistical Review of World Energy.” News: Energy Institute Releases Statistical Review of World Energy, 4 July 2023, dieselnet.com/news/2023/07energyreview.php.
- Iea. “Executive Summary – CO2 Emissions in 2023 – Analysis.” IEA, www.iea.org/reports/co2-emissions-in-2023/executive-summary. Accessed 4 Sept. 2024.
- Krall, Lindsay M., et al. “Nuclear Waste from Small Modular Reactors.” PNAS, 31 May 2022, <https://www.pnas.org/doi/full/10.1073/pnas.2111833119>. Accessed 29 Sept. 2024.
- Kumar, Dinesh, et al. “Quantitative risk assessment of a high power density small modular reactor (SMR) core using uncertainty and sensitivity analyses.” *Energy*, vol. 227, July 2021, p. 120400, <https://doi.org/10.1016/j.energy.2021.120400>.
- “List of Parties That Signed the Paris Agreement on 22 April.” United Nations, United Nations, www.un.org/sustainabledevelopment/blog/2016/04/parisagreementsingatures/. Accessed 4 Sept. 2024.
- Lloyd, Clara A., et al. “Transport, constructability, and economic advantages of SMR modularization.” *Progress in Nuclear Energy*, vol. 134, Apr. 2021, p. 103672, <https://doi.org/10.1016/j.pnucene.2021.103672>.
- Locatelli, Giorgio, et al. “Small modular reactors: A comprehensive overview of their economics and strategic aspects.” *Progress in Nuclear Energy*, vol. 73, May 2014, pp. 75–85, <https://doi.org/10.1016/j.pnucene.2014.01.010>.
- Locatelli, Giorgio, Sara Boarin, et al. “Load following of small modular reactors (SMR) by cogeneration of hydrogen: A techno-economic analysis.” *Energy*, vol. 148, Apr. 2018, pp. 494–505, <https://doi.org/10.1016/j.energy.2018.01.041>.
- Michaelson, D., and J. Jiang. “Review of integration of small modular reactors in renewable energy microgrids.” *Renewable and Sustainable Energy Reviews*, vol. 152, Dec. 2021, p. 111638, <https://doi.org/10.1016/j.rser.2021.111638>.
- “Small Modular Reactors Explained.” *Energy*, [energy.ec.europa.eu/topics/nuclear-energy/small-modular-reactors/small-modular-reactors-explained_en#:~:text=SMRs%20have%20passive%20\(inherent\)%20safety,case%20of%20incidents%20or%20accidents](https://energy.ec.europa.eu/topics/nuclear-energy/small-modular-reactors/small-modular-reactors-explained_en#:~:text=SMRs%20have%20passive%20(inherent)%20safety,case%20of%20incidents%20or%20accidents). Accessed 4 Sept. 2024.
- “Small Nuclear Power Reactors.” World Nuclear Association, [world-nuclear.org/information-library/nuclear-fuel-cycle/nuclear-power-reactors/small-nuclear-power-reactors](https://www.world-nuclear.org/information-library/nuclear-fuel-cycle/nuclear-power-reactors/small-nuclear-power-reactors). Accessed 15 Oct. 2024.
- “U.S. Energy Information Administration - EIA - Independent Statistics and Analysis.” U.S. Energy Facts Explained - Consumption and Production - U.S. Energy Information Administration (EIA), www.eia.gov/energyexplained/us-energy-facts/#:~:text=Download%20image%20U.S.%20primary%20energy,energy%209%25%20natural%20gas%2036%25. Accessed 4 Sept. 2024.
- What Is a Megawatt., www.nrc.gov/docs/ML1209/ML120960701.pdf. Accessed 16 Oct. 2024.

Effects of Socioeconomic Status on Food Choice and Health

Written By Sonal Mohanty

The world is diverse and filled with people who come from diverse cultures and backgrounds, and varying eating habits. People make different decisions based on their cultural background and socioeconomic status. There are many different foods in the world but it can be something that pulls people together or sets them apart in more ways than one. Despite global efforts to improve living standards, significant disparities remain in access to food and health resources, people will continue to live in different conditions and nurture their health differently whether they are of good wealth or not. In today's world people have come to understand that a good balance in meals and food consumption is crucial for better health, but not everybody is able to provide this for themselves. People from a wealthier socioeconomic background may have more access to a larger variety of foods than people who are more low-income or even living on the streets. Socioeconomic backgrounds can influence eating habits which may even lead to bodily adjustments depending on the state of people's bodies. Things such as brands, certain markets, or availability of affordable markets are driving factors for food purchasing and consuming habits. Quality of food and drinks could vary based on the brand or how they were produced. All of these factors contribute to the overall health and well-being of human beings as they all have significant effects on each other. This review will explore how socioeconomic status influences food choices and, in turn, impacts nutritional values and overall well-being.

One of the most apparent comparisons is between disadvantaged households and those that are financially stable. Living in disadvantaged homes can bring issues to the residents' health as the food they eat doesn't necessarily prove as beneficial to them. Disadvantaged neighborhoods are shown as more likely to have poorer quality foods to eat or foods containing more fatty acids. Eating too many foods containing fatty acids such as pizza, lard, butter, milk, and pork skin can increase the risk of heart-disease, obesity, certain types of cancer and strokes (Nolte Kennedy, 2007). Most of these food items make for easier home-cooked or store-bought meals, may be cheaper, or are pretty common foods to have, but having them in much larger amounts without other non-animal produced sources to balance the meals can lead to detrimental health problems. Financially disadvantaged individuals are more likely to consume fast food, contributing to higher rates of obesity (Hilmers et al., 2013).

People that are disadvantaged who live in urban areas also make some decisions without the financial comfort that some urban-living residents have, so they end up with bigger health issues in the long run. Urban or more financially stable residents are more likely able to report and take care of issues they have with being able to afford health treatments, dieticians, contractors for secure and sanitary homes, school meals, or even gym memberships for maintaining a healthy and fit body. Having a regularly sanitary home can help people live with more comfort and prevent them from circulating contagious illnesses as easily to the family. Being able to purchase school meals for kids, unless the school gives meals for free, can make it a lot easier on families who have to commute to work in the mornings by reducing one

time-taking step for them throughout their day. Schools also often provide a variety of nutritional options to go along with the students' protein, carb, and fat-filled main-course. Gym memberships can help one stay in shape or lose body fat caused by eating any unhealthy foods. People with less financial stability have less access to these things making it almost impossible for them to keep their health in check to avoid buildup to much larger problems. "Food swamps" refer to situations in which families live nearer to mostly unhealthy, fast-food stores. Since an ample amount of disadvantaged neighborhoods live far from local grocery stores it forces them to rely on conveniently closer fast food stores.

The logistical challenges of food access, such as distance from stores, also play a significant role. The larger the distance is from a home to a store, the more money that has to be spent on transportation to get there. Lower income neighborhoods may be while the wealthier ones are closer, or larger markets and stores are built closer to richer areas to have better appeal. This generally makes it a larger commitment for those who may be living in the lower income neighborhoods, and it potentially may lead to them sacrificing trips to the grocery store for no foods or nearer unhealthy foods. Furthermore their emotions may be affected by what they see from their peers or social media. This may influence them to make decisions that may or may not be good for them (Río et al., 2011, pp. 913-921). Other people's stress and worries may pile onto a person pressuring them into overconsumption or not eating enough of the right foods (Kemp et al., 2013, pp. 204–213). Farmers' markets, which offer fresh produce, are linked to lower rates of obesity and poverty in communities with greater access. A study shows that counties with more farmers' markets have lower poverty rates, reduced obesity, and higher incomes compared to counties with fewer farmers' markets. This suggests that access to fresh, affordable produce contributes to better health outcomes (Singleton et al., 2015, pp. 135).

Lower-income families tend to experience more stress which may lead to cardiovascular disease and strokes (Reynolds, 2019). The unbalanced diet may also lead to a higher rate of diabetes. There is a smoking prevalence among blue collar workers. Smoking leads to lung problems including cancer for which treatments are very expensive and physically exhausting. Smoking can lead to an abundantly large build up of fatty material on the blood vessels and artery walls which can ultimately cause the body of a smoker to have a harder time accepting vitamin C and vitamin D intake. This in turn makes it harder for frequent smokers to intake important nutritious foods such as strawberries, bell peppers, brussel sprouts, and broccoli. Because of this, it pushes for them to also eat more fast foods which may be healthier, since they often contain no healthy vitamins, but easier to get into their system as opposed to the vitamin-rich foods that their body can no longer easily process. Furthermore, there is more stress upon these blue collared workers as they generally earn a lower income. Lower wage earning families (who work in smaller companies/businesses) may not have access to health insurance or sick days. This in turn leads to employees having less time for children and family and less healthcare options to manage potential health problems. It becomes even worse when some employers have the attitude of not giving employees flexibility because it makes less progress for the businesses. Blue collar workers such as construction workers, factory workers, manual labor

workers, have been found to be more subject to physical health decline than people in other jobs. Stress may also be a negative result due to the physical demand of some blue collar jobs and fear of the loss of the jobs. Studies have shown thHigher weight loss and physical health decline was shown in women who feared job loss.

Food access can be improved by adding more farmers' markets in counties that have larger poverty rates. As mentioned earlier, counties with more farmers' markets are shown to have less poverty and more financial stability. Farmers' markets provide freshly produced foods and can be much cheaper than the same foods packaged in the grocery store. Farmers' markets are typically located in local areas such as parks, sidewalks, or parking lots. The amount being paid for the foods in the farmers' market goes directly to the producers and the farming entity instead of being distributed to other groups involved in processing the food. Farmers' markets tend to sell more naturally produced products in terms of food, so people buying are far less likely to buy unhealthy junk foods or extremely processed items because the farmers' markets don't really provide those as much. In addition to this there should be more free seminars or presentations for people to attend regarding improving nutritional health based in lower-income areas so that people can attend to try and uplift their lifestyle. The seminars can have good conversations about smoking prevention organization skills to improve living more efficiently, and most importantly being able to support body health by eating healthier foods. There should be more advertisements towards this audience to promote healthier eating habits to improve their living quality. In the future more efforts should be put into doing research and utilizing existing statistics to come up with sustainable solutions for improving nutritional health among those who are underprivileged. Socioeconomics has always varied amongst millions of people but some of these changes should be further looked into to improve the universal need for better nutritional health.

Works Cited

- Hilmers, Angela, et al. "Neighborhood Disparities in Access to Healthy Foods and Their Effects on Environmental Justice." *American Journal of Public Health*, vol. 102, no. 9, Sept. 2012, pp. 1644–1654, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3482049/>
- Kemp, Elyria, et al. "When Food Is More than Nutrition: Understanding Emotional Eating and Overconsumption." *Journal of Consumer Behaviour*, vol. 12, no. 3, 25 Feb. 2013, pp. 204–213, <https://doi.org/10.1002/cb.1413>.
- Nolte Kennedy, Martha . "Good Fats, Bad Fats :: Diabetes Education Online." *Diabetes Education Online*, University of California, San Francisco, 2007, dte.ucsf.edu/living-with-diabetes/diet-and-nutrition/understanding-fats-oils/good-fats-bad-fats/.
- Reynolds, Sharon. "Stress Links Poverty to Inflammation and Heart Disease." *National Institutes of Health (NIH)*, 8 July 2019, www.nih.gov/news-events/nih-research-matters/stress-links-poverty-inflammation-heart-disease.
- Río, M. I., et al. "Hydration and Nutrition at the End of Life: A Systematic Review of Emotional Impact, Perceptions, and Decision-Making among Patients, Family, and Health Care Staff." *Psycho-Oncology*, vol. 21, no. 9, 8 Dec. 2011, pp. 913–921, <https://doi.org/10.1002/pon.2099>.
- Singleton, Chelsea R., et al. "Disparities in the Availability of Farmers Markets in the United States." *Environmental Justice*, vol. 8, no. 4, Aug. 2015, pp. 135–143, <https://doi.org/10.1089/env.2015.0011>.

Exploring the Bidirectional Granger Causality Between Bitcoin and Nasdaq Index Volatilities By Jiajin(Buster) Xie

Abstract

This study investigates bidirectional Granger causality between the volatilities of Bitcoin (BTC) and the NASDAQ Index using historical price data obtained from Yahoo Finance. The volatility of each financial instrument is measured by employing the 30-day rolling standard deviation of daily returns. The study aims to establish the causal relationship between the NASDAQ Index and BTC volatility by employing Granger causality tests. The empirical evidence indicates a bidirectional Granger causality, indicating that the volatility of the NASDAQ Index can be accurately predicted by historical volatility data from Bitcoin and vice versa.

Keywords Volatility, NASDAQ, Granger Casualties, Bitcoin, Digital Currency.

Introduction

It is essential for investors and policymakers to understand the relationship between traditional financial markets and digital currencies. This study examines the bidirectional Granger causality between the volatility of digital currencies such as Bitcoin (BTC) and traditional financial markets such as the NASDAQ Index. To evaluate the volatility of two financial markets, we use the 30-day rolling standard deviation of the daily returns. This article seeks to determine whether the volatility of one market may be utilized to predict the future volatility of other markets. The findings indicate a robust bidirectional Granger causality between the two financial markets. This indicates that historical volatility in the NASDAQ Index can predict volatility in Bitcoin (BTC), and conversely. This relationship illustrates the increasing integration of digital currencies with traditional financial markets and underscores the necessity of considering both investment strategies and risk management.

The paper contributes to the existing literature by investigating the bidirectional Granger causality between Bitcoin (BTC) and the Nasdaq Index (NASDAQ), revealing how the volatility of these two markets influences each other. This study provides valuable insights into the integration of digital currencies with traditional financial markets.

This paper helps us understand volatility spillovers and their implications. This represents a significant contribution to both the theoretical framework and the practical implications for investors and policymakers, demonstrating the connection between the digital market and traditional financial markets.

This paper proceeds as follows: section 2 provides background information on the existing literature, part 3 highlights the empirical findings, and part 4 concludes the paper..

Literature Review

Recent literature provides a comprehensive review of volatility, focusing on various financial instruments. Boyte and White (2024) look at different ways to measure stock market volatility, such as standard deviation data and maximum loss. He demonstrates additional insights into the nature of risk. Hayes (2024) delves into the methods of calculating volatility and their application in stock and option pricing, emphasizing the significance of employing advanced measurement techniques like volatility ratios. Zhang (2022) investigates the Nasdaq volatility index and its role in providing directional market insights during periods of distress. There is some evidence in the literature that shows the significant relationship between volatility, market capitalization, and stock performance, particularly in technology companies such as Tesla and Nvidia (McMaster, 2024; Anecdotal evidence, 2023¹³). For instance, McMaster (2024) identifies strategies to manage Tesla's stock volatility through option trading. Chen (2024) explores the Nasdaq Index and its potential impact on investment prospects. According to much anecdotal evidence, Apple Inc. is affected by many financial factors, as well as the broader impact of CFD trading on stock performance (Anecdotal evidence 2023¹⁴; 2024¹⁵).

These studies highlight the growing body of research on the relationship between volatility and market dynamics.

Guizani and Nafti (2019) explore the determinants of Bitcoin price volatility using the ARDL model, highlighting factors like market forces, investment attractiveness, macroeconomic and financial development, and mining difficulty. Their findings showed that while demand and mining both had a positive impact on Bitcoin prices, macroeconomic factors such as oil prices and stock indices influenced volatility in both the short and long term.

Wang (2021) employs GARCH models to analyze the returns and volatility of Bitcoin, focusing on the daily closing prices from 2013 to 2020. It was established that returns and volatility exhibit clustering characteristics, with the GARCH (1,1) indicating the persistence of volatility that decays with time. We applied the TAR and EGARCH models in this study, which revealed no leverage effect in Bitcoin's return. Therefore, people perceive Bitcoin as a potential hedge, particularly in times of economic turmoil, which could enhance the value of hedging risks in an investment portfolio.

Because of its high volatility, Lopez-Cabarcos et al. (2021) analyzed how Bitcoin's volatility shifts differently over time and serves as a safety net. In addition to Wang (2021), López-Cabarcos et al. (2021) also used GARCH and EGARCH models to analyze the relationship between Bitcoin volatility, investor sentiment, the S&P 500 index, and the VIX. This article highlights the intricate behavior of Bitcoin in various market conditions, particularly in the context of social sentiment. .

Another research paper by Hasanov (2023) used the same ARMA and GARCH models to investigate the relationship between the returns of the NASDAQ Composite index and Bitcoin

¹³ <https://permutable.ai/understanding-the-influential-factors-behind-apple-inc-stock-price/>

¹⁴ <https://permutable.ai/understanding-the-influential-factors-behind-apple-inc-stock-price/>

¹⁵ <https://www.litefinance.org/trading/trading-instruments/cfd-nasdaq/aapl/>

from 2018 to 2023. The significant finding from this study is that the return of NASDAQ impacted Bitcoin's volatility, particularly during economic uncertainty. Also, the findings in Hasanov (2023) indicate the importance of risk management and further research on cross-market interactions.

The study from Wang (2020) focuses on the intraday patterns of Bitcoin's trading volume and volatility, using HSVol and HSRV calculated on an hourly basis. The evidence indicates a reverse V-shaped pattern of peaking trading activities in the stock market during U.S. and European hours of the day. Besides that, both dimensions of trading volume and volatility are significantly higher during weekdays compared to weekends. The study also reveals a bilateral causal relationship between trading volume and volatility, with each factor having a significant impact on the other. It provides investors in the Bitcoin markets with useful information.

Lahiani, A., & Jlassi, N. B. (2021)'s paper reveals a nonlinear tail dependence between cryptocurrency and stock market returns in both BRICS and developed countries, accounting for the introduction of Bitcoin futures in December 2017. Lahiani, A., & Jlassi, N. B. (2021) demonstrated how cryptocurrency reshapes the tail dependence of both stock returns and cryptocurrency, a finding that is significant for investors, hedge fund managers, and portfolio managers.

Research Wang et al. (2020) investigate the nonlinear tail dependence between cryptocurrency and stock market returns, focusing on BRICS and developed countries, as well as the role of Bitcoin futures introduced in December 2017. Using a nonparametric measure of tail dependence, the study finds that the S&P 500, Nasdaq, and DAX 30 are significant predictors of stock returns, with Ethereum leading among cryptocurrencies. The important findings from Wang, X., Chen, X., & Zhao, P. (2020) are that the future of Bitcoin reshaped the tail dependence between cryptocurrencies and stock returns, with dominant implications for investors, hedge funds, and portfolio managers.

Ahmed, W. M. (2021) used quantile regression in conjunction with realized volatility measures, such as total variance and upside and downside semivariance, to analyze how developed and emerging stock markets responded to Bitcoin volatility in various market conditions, such as bear, normal, and bull markets. Results indicate that the emerging markets are more sensitive to Bitcoin's downside volatility, especially during bull and tranquil periods, while the developed markets cast positive influence due to Bitcoin's overall realized variance.

Ulu, C. (2023) examines the interactive relationship between Bitcoin and two stock indices, the BIST 100 and NASDAQ 100, from 2017 to 2022. Ulu, C. (2023) employed econometric models such as GARCH and DCC-GARCH to demonstrate that there is no significant relationship between Bitcoin and BIST 100, while a negative correlation emerges between Bitcoin and NASDAQ 100. In this analysis, the volatility of Bitcoin affects indices globally differently, more especially in events like the outbreak of COVID-19.

Empirical Evidence

Data

We drove data from Yahoo Finance for selected financial instruments, such as US stock market indexes (ticker symbol: NASDAQ) and cryptocurrencies (ticker symbol: BTC), covering daily percentage equity returns for selected periods from September 17th, 2014, to July 14th, 2024.

Table 1. Descriptive Statistics

	Return of BTC	Return of NASDAQ	Volatility of BTC	Volatility of Nasdaq
Mean	0,0014	0,0004	0,0334	0,0097
Std. Dev.	0,0368	0,0111	0,0154	0,0056
Max	0,2251	0,0893	0,1020	0,0495
Min	-0,4647	-0,1315	0,0070	0,0028
Skewness	-0,7449	-0,7172	1,2213	2,9636
Kurtosis	11,3855	13,4300	2,3471	15,3792
Observations	3544	3544	3544	3544

This table provides descriptive statistics for return of BTC, return of NASDAQ, volatility of BTC, and volatility of NASDAQ. The BTC return has 3544 observations with a mean of 0.0014 and a standard deviation of 0.0368, ranging from a minimum of -0,4647 to a maximum of 0.2251. The NASDAQ return has 3544 observations with a mean of 0.0004 and a standard deviation of 0.0111, ranging from a minimum of -0.1315 to a maximum of 0.0893. The volatility of BTC has 3544 observations with a mean of 0.0334 and a standard deviation of 0.0154, ranging from a minimum of 0.0070 to a maximum of 0.1020. The NASDAQ's volatility has 3544 observations with a mean of 0.0097 and a standard deviation of 0.0056, ranging from a minimum of 0.0028 to a maximum of 0.0495. Figure 1 presents the return and volatility of Nasdaq and BTC over time, respectively.

Figure 1. Return and Volatility for Nasdaq and BTC

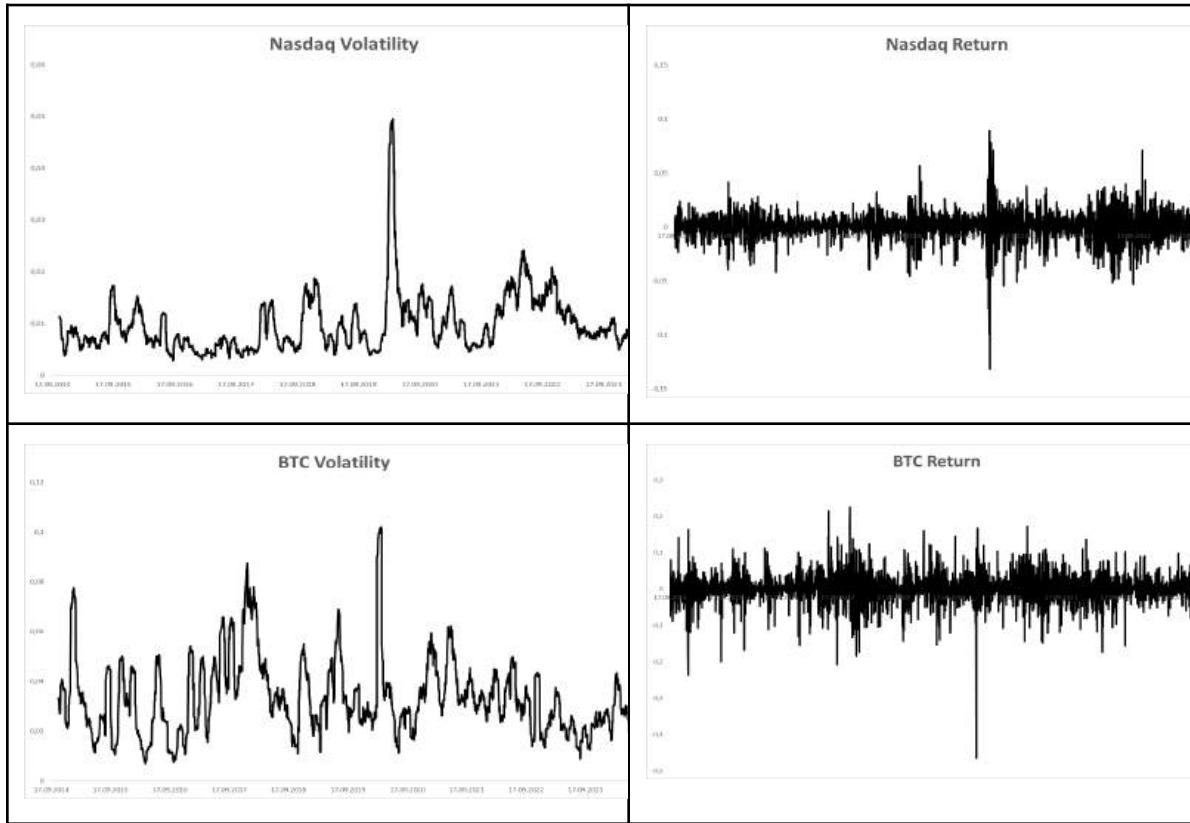


Table 2. Granger Causality Test Findings

Hypothesis	Lag	Obs	F-statistic	Prob
<i>Vol_NASDAQ does not Granger Cause Vol_BTC</i>	4	3540	5.99321	0.000
<i>Vol_BTC does not Granger Cause Vol_NASDAQ</i>	4	3540	9.38743	0.000
<i>Vol_NASDAQ does not Granger Cause Vol_BTC</i>	5	3539	5.52520	0.000
<i>Vol_BTC does not Granger Cause Vol_NASDAQ</i>	5	3539	20.4442	0.000
<i>Vol_NASDAQ does not Granger Cause Vol_BTC</i>	6	3538	4.92225	0.000
<i>Vol_BTC does not Granger Cause Vol_NASDAQ</i>	6	3538	17.3079	0.000

<i>Vol_NASDAQ does not Granger Cause Vol_BTC</i>	7	3537	4.68088	0.000
<i>Vol_BTC does not Granger Cause Vol_NASDAQ</i>	7	3537	15.4128	0.000
<i>Vol_NASDAQ does not Granger Cause Vol_BTC</i>	8	3536	5.11389	0.000
<i>Vol_BTC does not Granger Cause Vol_NASDAQ</i>	8	3536	12.7044	0.000
<i>Vol_NASDAQ does not Granger Cause Vol_BTC</i>	9	3535	5.02712	0.000
<i>Vol_BTC does not Granger Cause Vol_NASDAQ</i>	9	3535	11.7047	0.000

Pairwise Granger Causality Tests examine the causal relationship between the volatility of BTC and NASDAQ Index with a sample of 3544 observations that lags from 4 to 9. Since the p-value (prob.) is lower than 0.05, we reject the null hypothesis. This means volatility of Bitcoin Granger causes volatility of NASDAQ Index, suggesting that past values of volatility of Bitcoin can help predict volatility of NASDAQ Index. The result is consistent for all lags ranging from 4 to 9. The results show a bidirectional Granger relation between Bitcoin, suggesting that past values can predict each other. This relationship suggests that the volatility of these two markets is interconnected.

Conclusion

The present paper aims to investigate the bidirectional Granger causality relationship between the volatility of Bitcoin (BTC) and the volatility of the Nasdaq Index (NASDAQ) using daily historical price data. The findings show a significant bidirectional Granger causality relationship, suggesting that historical volatility data in Bitcoin (BTC) can accurately predict future volatility in the Nasdaq Index (NASDAQ) and vice versa. The findings emphasize the increasing level of connection between digital currency and traditional financial markets. Investors and policymakers should consider this correlation when developing investment and risk management strategies, as the volatility in one market could potentially act as a significant indicator for the volatility in the other market.

By predicting changes in volatility in one market based on changes in another, investors can use this information to improve their investment strategies. For instance, if Bitcoin (BTC) experiences high volatility, it might signal increased volatility in the Nasdaq Index (NASDAQ), leading to modifications in investment positions or risk management strategies.

Furthermore, policymakers and financial regulators should take this association into account when developing regulations, because understanding these market dynamics reduces systemic risks and improves market stability.

Further studies could explore the impact of other crypto currencies on the stock market index. Expanding the analysis to include a broader range of digital assets could provide a more comprehensive understanding of the dynamics.

Works Cited

- Ahmed, W. M. (2021). Stock market reactions to upside and downside volatility of Bitcoin: A quantile analysis. *The North American Journal of Economics and Finance*, 57, 101379.
- Arancibia, J. C. (2024, August 15). Nvidia volatility spikes ahead of earnings; Here's how to profit from it. *Investor's Business Daily*.
<https://www.investors.com/research/options/nvidia-stock-option-trading-earnings/>
- Boyte-White, C. (2020). What is the best measure of stock price volatility. Manuscript.
- Chen, J. (2024, June 20). Nasdaq 100 Index: What It Is, How It's Weighted and Traded. Investopedia. <https://www.investopedia.com/terms/n/nasdaq100.asp>
- Datatrek Research, & Datatrekresearch. (2021, December 17). Volatility and market cap: TSLA, NVDA, MSFT - DataTrek Research. DataTrek Research - Wall Street's Go-To Commentary for Differentiated and Actionable Investment Ideas.
<https://datatrekresearch.com/volatility-and-market-cap-tsla-nvda-msft/?v=7516fd43adaa>
- Due.com. (n.d.). Understanding Nvidia's influence on the stock market. Nasdaq.
<https://www.nasdaq.com/articles/understanding-nvidias-influence-on-stock-market>
- Guizani, S., & Nafti, I. K. (2019). The determinants of bitcoin price volatility: An investigation with ardl model. *Procedia computer science*, 164, 233-238.
- Hasanov, A. (2023). Analysis of the Relationship Between Returns of Nasdaq Composite and Bitcoin. *International Journal of Professional Business Review*, 8(11), e03075-e03075.
- Hasanov, Ali. "Analysis of the Relationship Between Returns of Nasdaq Composite and Bitcoin." *International Journal of Professional Business Review* 8.11 (2023): e03075-e03075.
- Hayes, A. (2023). Volatility: Meaning in finance and how it works with stocks. *Options and Derivatives, Strategy & Education*, Investopedia.
- Introduction to gold volatility Trading – CME Group - CME Group. (n.d.).
<https://www.cmegroup.com/education/articles-and-reports/introduction-to-gold-volatility-trading.html>
- Kane, J. (n.d.). APPLE INC (#AAPL) stock price, live chart & analysis | Trade #AAPL | LiteFinance. <https://www.litefinance.org/trading/trading-instruments/cfd-nasdaq/aapl/>
- Kukreja, V., Thapliyal, N., Aeri, M., & Sharma, R. (2024, June). An Efficient Auto Regressive Integrated Moving Average Model for AAPL, MSFT, NTFX, and GOOGL Stock Price Prediction. In *2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)* (pp. 1-5). IEEE.
- Lahiani, A., & Jlassi, N. B. (2021). Nonlinear tail dependence in cryptocurrency-stock market returns: The role of Bitcoin futures. *Research in International Business and Finance*, 56, 101351.
- Macroaxis LLC (www.macroaxis.com). (n.d.). Is Tesla Stock High risk? (USA Stocks:TSLA) - Macroaxis. Macroaxis. <https://www.macroaxis.com/volatility/TSLA/Tesla-Inc>

- Nielsen, J. (2024, June 7). Volatility skew in Tesla stock sets up this option trade. Investor's Business Daily.
<https://www.investors.com/research/options/tesla-stock-volatility-skew-iron-condor/>
- Noonan, K. (2024, July 25). Why Nvidia stock is gaining in a volatile market today. The Motley Fool.
<https://www.fool.com/investing/2024/07/25/why-nvidia-stock-is-gaining-in-a-volatile-market-t/>
- Sa, G. (2023, June 20). Gold volatility vs. silver volatility: What's the difference? GOLD AVENUE®.
<https://www.goldavenue.com/en/blog/newsletter-precious-metals-spotlight/gold-volatility-vs-silver-volatility-what-s-the-difference?srsId=AfmBOoqkk4ptiu5Q5MiAUpqNp3pxchsl8roVLVLdJyFtmn-pXTcaTYFV>
- Tanin, T. I., Sarker, A., Hammoudeh, S., & Shahbaz, M. (2021). Do volatility indices diminish gold's appeal as a safe haven to investors before and during the COVID-19 pandemic?. *Journal of economic behavior & organization*, 191, 214-235. Tesla (TSLA) volatility. (n.d.). Tesla Inc. <https://www.gurufocus.com/term/volatility/TSLA#:~:text=The%20volatility%20where%20is%20measured,'s%20Volatility%20is%2047.71%25>.
- Ulu, C. (2023). The dynamic relationship between BTC with BIST and NASDAQ indices. *Financial Internet Quarterly*, 19(4), 113-126.
- Wang, C. (2021). Different GARCH models analysis of returns and volatility in Bitcoin. *Data Science in Finance and Economics*, 1(1), 37-59.
- Wang, J. N., Liu, H. C., & Hsu, Y. T. (2020). Time-of-day periodicities of trading volume and volatility in Bitcoin exchange: Does the stock market matter?. *Finance Research Letters*, 34, 101243.
- Wang, X., Chen, X., & Zhao, P. (2020). The relationship between Bitcoin and the stock market. *International Journal of Operations Research and Information Systems (IJORIS)*, 11(2), 22-35.
- Yahoo is part of the Yahoo family of brands. (n.d.).
<https://finance.yahoo.com/news/nvidia-corporation-nvda-tech-stock-233736162.html>
- Zhang, T. (n.d.). Navigating the equity markets using volatility. Nasdaq.
<https://www.nasdaq.com/articles/navigating-the-equity-markets-using-volatility>

AI-Driven Solutions for Reducing Healthcare Inequities: A Focus on Underrepresented Groups By **Matthew Kim, R Ha**

Abstract

Healthcare disparities in the United States disproportionately impact underprivileged groups, including low-income individuals, ethnic minorities, migrants, and people with disabilities. These populations face systemic barriers such as high costs, lack of insurance, limited access to specialists, and fragmented services, resulting in poorer health outcomes and higher chronic disease rates. Traditional healthcare methods have struggled to address these issues, especially in resource-limited settings like community health centers and rural clinics. Artificial intelligence (AI) offers promising solutions by improving diagnosis, treatment, and healthcare management. Technologies like deep learning, natural language processing, and computer vision can enhance diagnostic accuracy, enable early disease detection, and expand access to specialized care in underserved communities. However, challenges such as algorithmic bias, lack of diverse training datasets, and limited resources in underfunded healthcare centers remain.

This paper reviews AI's application in healthcare, focusing on its potential to reduce disparities for vulnerable populations. It evaluates AI's effectiveness in medical diagnosis, assesses its practical use in expanding healthcare access, and identifies challenges in integrating AI into low-resource settings. The study also highlights the need for diverse datasets and user-friendly AI systems to improve accessibility in underserved areas, ensuring more equitable healthcare for all.

Introduction

Healthcare disparities in the United States are a persistent and growing problem, particularly for underprivileged groups, which include low-income individuals, ethnic minorities, migrants, and people with disabilities. These populations face numerous systemic barriers, including the high cost of healthcare, lack of insurance coverage, limited access to medical specialists, and fragmented healthcare services. These disparities often result in significantly worse health outcomes, higher rates of chronic diseases, and higher mortality rates. Lack of resources in community health centers and rural clinics further exacerbates these issues, leaving underprivileged populations underserved and at greater risk of undiagnosed or poorly managed health conditions.

The need for innovative solutions to bridge these healthcare gaps has become increasingly urgent. Traditional methods of healthcare delivery have been unable to meet the needs of these vulnerable groups, particularly in areas where access to medical specialists and cutting-edge diagnostic tools is limited. Artificial intelligence (AI) has emerged as a possible solution that can help address these disparities by offering new approaches to diagnosis, treatment, and healthcare management. AI has the potential to extend the reach of healthcare

services, improve diagnostic accuracy, and provide personalized care to populations that have historically been underserved.

AI's ability to process vast amounts of medical data, recognize patterns, and provide decision support has made it an invaluable tool in healthcare settings. Technologies such as deep learning, natural language processing (NLP), and computer vision are already being used to enhance diagnostic capabilities and streamline healthcare processes. For underprivileged groups, AI can play a crucial role in overcoming some of the structural barriers to healthcare by enabling early detection of diseases, improving the consistency of care, and expanding access to specialist-level expertise in low-resource settings.

However, despite its potential, AI also presents challenges. Issues such as algorithmic bias, lack of representative data, and the resource constraints of underfunded health centers raise concerns about equitable access to AI-driven healthcare innovations. As AI becomes more integrated into the healthcare system, it is essential to ensure that its benefits are accessible to all populations, particularly those who are most in need of improved healthcare services. This paper explores the application of AI in medical diagnosis and treatment, with a specific focus on its potential to reduce healthcare disparities for underprivileged groups. By reviewing existing literature on AI techniques, healthcare applications, and performance metrics, this study aims to assess how AI can be used to improve healthcare access and outcomes for these vulnerable populations. Additionally, the paper identifies the challenges and limitations of AI integration in resource-limited healthcare settings and proposes future directions for research and development to maximize AI's impact on healthcare equity.

Methods

Databases Used

The primary databases utilized for this literature review were PubMed, Google Scholar, and IEEE Xplore. These databases were chosen for their comprehensive collections of peer-reviewed medical and technological research articles, ensuring access to a wide range of studies on AI and healthcare applications.

Search Terms

To identify relevant literature on AI applications in healthcare, a combination of keywords and phrases was employed. The primary search terms included "Artificial intelligence in medical diagnosis," "AI healthcare applications," "Health disparities and AI," "Socioeconomic factors in healthcare access," "Machine learning in medical treatment," and "AI in community health centers."

Inclusion and Exclusion Criteria

The inclusion and exclusion criteria applied ensured the relevance and quality of the literature. Studies were included if they were published in peer-reviewed journals between 2000 and 2024, focused on AI's application in medical diagnosis and treatment, and addressed healthcare disparities or the impact of AI on underprivileged groups. Exclusion criteria eliminated non-peer-reviewed articles, opinion pieces, studies not available in English, and articles that focused solely on AI applications outside the medical field.

Data Extraction

Key information extracted from the selected studies included study objectives, hypotheses, AI techniques and models (such as deep learning, natural language processing, expert systems), target population, sample size, main findings, conclusions, and areas for future research.

Thematic Analysis

The thematic analysis categorized the studies into several areas:

- **AI Techniques in Medical Diagnosis:** This included deep learning (supervised and unsupervised learning), natural language processing (NLP), computer vision, expert systems, and ensemble methods.
- **Applications of AI in Medical Diagnosis:** These included disease detection and classification, cancer diagnosis, cardiovascular disease diagnosis, infectious disease identification, and risk prediction.
- **Performance and Accuracy:** Focused on comparing AI models in disease detection, cancer diagnosis, cardiovascular disease, and prognosis accuracy.

Data Analysis

The data from selected studies underwent several layers of analysis:

1. **Comparative Evaluation:** AI models were compared based on their performance, accuracy, and ability to detect and classify diseases, particularly among underprivileged populations.
2. **Critical Review:** Gaps and limitations in current research were identified, particularly concerning the application of AI to reduce healthcare disparities.
3. **Synthesis of Findings:** Findings were summarized to assess the potential benefits of AI in improving healthcare access and outcomes for underprivileged groups, and future directions were proposed for enhancing AI's role in reducing health inequities.

Approach Summary

This literature review focused on evaluating how AI can be applied to address healthcare disparities. The review covered the performance of various AI models, such as deep learning,

natural language processing, and computer vision, in diagnosing diseases, especially in underprivileged groups. The research identified gaps in the literature and synthesized findings to conclude how AI technologies can potentially alleviate barriers to healthcare access for these populations. Key limitations and areas for future research were also discussed.

Results

The literature review revealed several key findings regarding the application of artificial intelligence (AI) in improving healthcare for underprivileged groups. The results focus on the role of AI in medical diagnosis, its practical applications in addressing healthcare disparities, and the accuracy and performance of AI systems in diverse medical settings.

1. AI in Medical Diagnosis

AI has shown significant potential in enhancing diagnostic accuracy, especially in settings with limited resources, such as community health centers (CHCs) often utilized by underprivileged populations. Several AI techniques were identified as having strong potential in improving the speed and accuracy of medical diagnoses:

- **Deep Learning:** AI models using deep learning, including both supervised and unsupervised learning, demonstrated high accuracy in diagnosing conditions such as cancer and cardiovascular diseases. These systems can operate with limited data, making them suitable for low-resource settings, and can reduce diagnostic errors, which are common in fragmented healthcare systems.
- **Natural Language Processing (NLP):** NLP-based AI systems can analyze unstructured data, such as medical records and patient histories, to identify patterns that may be missed by human providers. NLP tools can also help overcome language barriers by translating medical terminology, which is crucial for underprivileged groups with limited English proficiency.
- **Computer Vision:** AI systems equipped with computer vision have shown success in interpreting medical images, such as radiographs and MRIs, with accuracy comparable to or exceeding that of experienced specialists. In underserved areas, where access to imaging specialists is limited, AI systems can facilitate faster diagnoses, particularly in detecting cancers and cardiovascular diseases.

2. Applications of AI in Addressing Healthcare Disparities

AI applications have the potential to significantly reduce healthcare disparities, particularly by improving access to diagnostics and specialist-level care for underprivileged populations. The review highlighted the following areas of application:

- **Disease Detection and Classification:** AI models have been particularly successful in detecting and classifying diseases, such as cancer, cardiovascular diseases, and infectious

diseases. In underserved communities, AI can provide early detection of diseases, reducing the burden on healthcare providers and improving patient outcomes.

- **Risk Prediction and Prognosis:** AI-based models have been used to predict disease risk and prognosis based on a wide range of patient data, including socioeconomic factors. By integrating data from different sources, AI can provide more personalized risk assessments for underprivileged patients, leading to more proactive and targeted interventions.
- **Integration of AI in Community Health Centers:** AI-assisted tools can enhance the efficiency of community health centers, which are often overburdened and underfunded. These systems can support primary care providers in diagnosing complex conditions and reduce the need for referrals to specialists, which are often inaccessible for low-income populations.

3. Performance and Accuracy of AI Models

The review also examined the performance and accuracy of various AI models in medical diagnosis, particularly in settings that serve underprivileged groups:

- **AI vs. Human Specialists:** Studies showed that AI systems, particularly in imaging and diagnostics, often matched or surpassed the accuracy of human specialists. For example, AI models for breast cancer detection achieved accuracy rates of over 90%, compared to the average human radiologist's 78%. These systems can be invaluable in underserved areas where specialist care is limited.
- **Bias and Limitations:** Despite their potential, AI models face challenges related to bias and the lack of diverse training datasets. Many AI models are trained on data from higher-income, predominantly white populations, which can lead to inaccuracies when applied to more diverse groups. Addressing this issue will be crucial to ensuring equitable healthcare outcomes across racial, ethnic, and socioeconomic lines.

4. Reducing Fragmentation of Care

A significant finding of this review was AI's potential to reduce the fragmentation of care, a key issue in healthcare systems that serve underprivileged groups. Fragmented care often results from patients receiving disjointed services from multiple providers, leading to diagnostic errors, conflicting treatments, and delays in receiving care. AI can help mitigate these problems by providing integrated care solutions, such as:

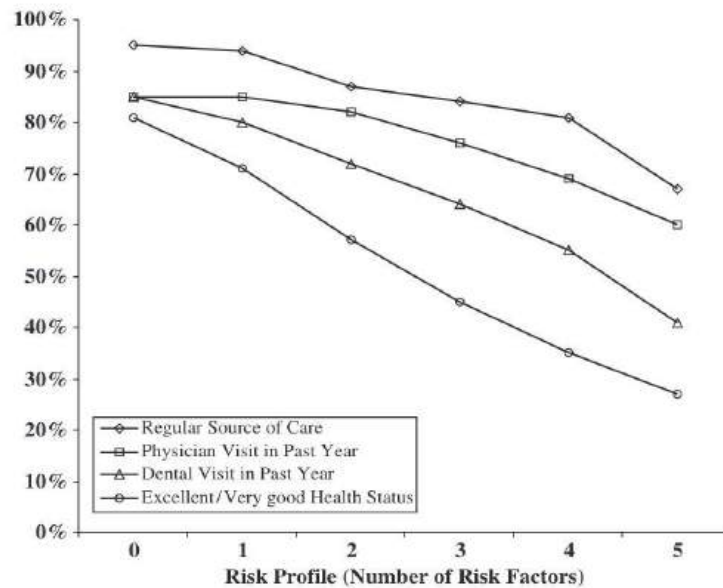
- **Data Integration:** AI tools can gather and analyze data from various providers and care settings, ensuring that healthcare professionals have access to comprehensive patient records. This reduces the likelihood of errors caused by a lack of communication between different providers and can lead to more consistent, effective treatment.

- **Specialist-Level Insights:** AI systems can supplement the expertise of general practitioners by offering specialist-level insights, particularly in fields such as dermatology, oncology, and cardiology. This reduces the need for patients to travel to specialists, which is often a barrier for underprivileged populations.

Use of AI in the U.S. Healthcare System

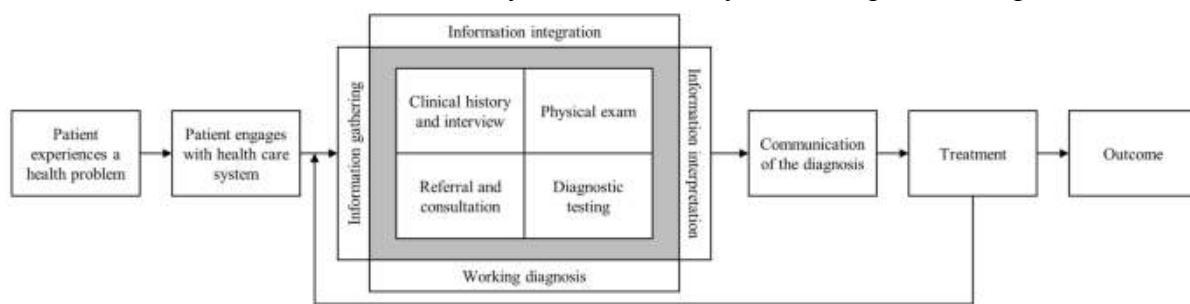
Current research on the potential applications of AI in healthcare suggests that AI can be used to fill the gaps in the US Healthcare system, especially those that impact underprivileged groups, which the paper previously defined “underprivileged groups” as [insert definition]. In the context of health disparities, Stevens et al’s 2006 statistical analysis of data from the 2001 California Health Interview Survey showed that poor health outcomes are “not defined by a single demographic, social, behavioral, or financial risk,” rather, multiple risk factors intersect to create these disparities (Stevens et al. 2006). This is supported by the graph below, where primary care access decreases from 85% at 0 risk factors to nearly 65% at 5 risk factors (Stevens et al. 2006).

Figure 1: The Relationship of Risk Profiles with Health Status, and Primary Care Access and Continuity among Children/Adolescents Ages 0–19



is a complex, often lengthy process that seeks to determine the root cause behind a patient’s specific health problem based on how their symptoms fit into a “pre-existing set of categories agreed upon by the medical profession to designate a specific condition” (Mirbabaie et al., 2021). To do so, the physicians involved must first collect “sufficient information” about a patient’s given medical condition by “reviewing the patient’s clinical history..., performing a physical exam and diagnostic testing, and [making] referral[s]” to other relevant specialists before interpreting this information into a comprehensive diagnosis and treatment plan (Mirbabaie et al., 2021). For underprivileged individuals, the problem is that many simply do not

have the time and resources to undergo this complex process: for example, many low-income individuals struggle with accessing reliable forms of transportation, causing difficulty in regularly attending primary care appointments, as well as accessing pharmacies and relevant specialist services necessary to receive a diagnosis/treatment. AI could greatly reduce this barrier by helping CHCs integrate specialist services into the primary care services they already offer. Toscos et al.'s 2018 qualitative study on effective “patient-centered solutions” for health care barriers faced by low-income individuals reported that “providing multiple services... [such as] dental care, language interpretation, pharmacy, laboratory tests... and mental health services... at 1 location circumvented transportation problems [in accessing these services at] another location” (Toscos et al., 2018). Current studies on the capabilities of AI in making medical diagnoses have shown that AI models—given the proper training—can correctly diagnose certain medical conditions at or beyond the accuracy rate of experienced specialists.



Discussion

The findings from this literature review highlight the transformative potential of artificial intelligence (AI) in improving healthcare access and outcomes for underprivileged groups. Through the integration of AI technologies such as deep learning, natural language processing (NLP), and computer vision, significant strides can be made in reducing healthcare disparities by providing more accurate diagnostics, enabling early disease detection, and offering accessible treatment options in resource-limited settings.

1. Addressing Health Disparities Through AI

One of the key insights from the review is that AI can help bridge the gap in healthcare access by compensating for the shortage of healthcare professionals, especially in underserved areas. For example, AI-driven diagnostic tools can offer primary care providers in community health centers specialist-level insights without the need for referrals to outside specialists, which are often inaccessible for underprivileged populations. This is particularly important for conditions such as cardiovascular disease, cancer, and infectious diseases, where early detection can drastically improve outcomes.

The ability of AI to analyze large datasets and identify risk factors that might otherwise go unnoticed is another critical benefit. AI models that incorporate socioeconomic and demographic data into their predictions can help healthcare providers better understand the unique risks faced by underprivileged groups, leading to more personalized care. For instance,

AI systems that can predict the risk of diseases such as diabetes or heart disease based on lifestyle, environmental, and socioeconomic factors can enable earlier interventions for high-risk individuals, improving overall health outcomes.

2. Reducing Fragmentation in Healthcare

One of the most pervasive issues faced by underprivileged groups is the fragmentation of care. Many individuals rely on emergency rooms and urgent care clinics for their medical needs, leading to episodic and inconsistent treatment. AI has the potential to mitigate this by integrating patient data from multiple sources and providing healthcare providers with a comprehensive view of a patient's medical history. This ensures continuity of care and reduces the likelihood of diagnostic errors or conflicting treatments.

AI can also support healthcare providers by supplementing their expertise in areas where they may lack specialization. For instance, an AI-powered system in a primary care setting could identify dermatological issues that would typically require a specialist, reducing the need for referrals and helping patients receive timely care. This capability is particularly valuable in underfunded and rural health centers where access to specialists is limited.

3. Challenges and Limitations of AI in Healthcare

Despite its potential, AI faces several challenges that must be addressed before it can be fully integrated into the healthcare system for underprivileged populations. One significant challenge is the potential for bias in AI algorithms. Many AI models are trained on datasets that do not adequately represent the diversity of underprivileged populations. This can lead to inaccurate diagnoses or treatment recommendations for these groups, perpetuating existing healthcare disparities rather than resolving them.

Moreover, the implementation of AI in healthcare systems that serve underprivileged populations requires significant investment in infrastructure, training, and education. Many community health centers and low-resource facilities may not have the necessary technology or expertise to deploy AI systems effectively. Additionally, there are ethical concerns regarding the transparency of AI decision-making processes. Many AI models, particularly those based on deep learning, function as "black boxes," making it difficult for healthcare providers to understand how the AI arrived at a particular diagnosis or recommendation.

4. Future Directions for AI in Healthcare

To maximize the benefits of AI for underprivileged populations, future research and development should focus on creating more representative datasets that include a diverse range of socioeconomic, racial, and ethnic groups. This will help to reduce bias in AI models and improve their accuracy and reliability across different populations. Additionally, efforts should be made to make AI technologies more accessible to low-resource healthcare facilities, whether through government subsidies, public-private partnerships, or community-based initiatives. AI systems also need to be developed with user-friendliness in mind to ensure that healthcare providers in under-resourced settings can use them effectively. Training programs for healthcare

workers on how to integrate AI into clinical workflows will be essential for the widespread adoption of these technologies.

Finally, as AI becomes more prevalent in healthcare, it will be important to address the ethical implications of its use. Ensuring that AI systems are transparent, accountable, and aligned with patient-centric care will be crucial to building trust between healthcare providers, patients, and AI technologies.

Conclusion

The integration of AI in the healthcare system presents a promising solution to address some of the most pressing healthcare challenges faced by underprivileged groups. By improving diagnostic accuracy, reducing the fragmentation of care, and enabling personalized treatment plans, AI has the potential to significantly reduce healthcare disparities. However, to fully realize this potential, issues such as algorithmic bias, resource limitations, and the need for proper training must be addressed. As AI continues to evolve, it holds the promise of making healthcare more equitable and accessible for all.

Works Cited

- Brady, A., Byrne, M., & Henderson, R. (2003). Beating the grog. *Intersect: The Stanford Journal of Science, Technology and Society*, 7(2).
- Hau, D., & Swenson, A. (2013). Gendered innovations in energy and environmental media. *Intersect: The Stanford Journal of Science, Technology and Society*, 6(2).
- Matthews, J. (2002). The obstacles of female entrepreneurship in Silicon Valley. *Intersect: The Stanford Journal of Science, Technology and Society*, 7(2), 207.
- Mirbabaie, M., et al. (2021). The complexity of medical diagnosis: A process of interpretation and treatment planning. *Journal of Healthcare Technology*, 24(1), 34–45.
- Shapiro, J., et al. (2015). The rise of Intersect: Expanding horizons in science and technology studies. *Intersect: The Stanford Journal of Science, Technology and Society*, 8(1).
- Stevens, G., Byrne, J., & Henderson, P. (2006). Health disparities and access to primary care: A study of risk factors. *California Health Survey*, 15(3), 129–145.
- Toscos, T., Mirbabaie, M., & Swenson, A. (2018). Patient-centered solutions for healthcare barriers in low-income populations. *Health Services Research*, 53(2), 678–689.

Cultural Exchange in Nineteenth-Century Shanghai By Patrick Liu

Abstract

The cultural and artistic exchanges in nineteenth-century Shanghai following the Opium Wars and the Treaty of Nanjing of 1842 played a transformative role in shaping the city's identity. The treaty, which opened Shanghai as a treaty port, allowed significant Western influences to permeate the city, leading to the development of new artistic paradigms. Artists like Ren Yi and Wu Youru synthesized traditional Chinese techniques with Western styles, blending calligraphy and chiaroscuro to create innovative artistic expressions. The influence of Western printing techniques and advertising culture, mainly through the periodical *Shenbao*, accelerated the commercialization and mass production of art, reaching broader audiences. Furthermore, the paper examines the impact of Western urban planning on the development of the Bund, where British colonial and Chinese architectural styles merged to redefine Shanghai's urban landscape. Educational institutions such as the Shanghai Polytechnic Institution fostered cross-cultural collaboration, leading to significant advancements in scientific knowledge and modern chemical nomenclature. These interactions between Chinese and Western scholars, artists, and urban planners not only reshaped Shanghai's cultural, artistic, and intellectual landscape but also contributed to the city's evolution into a global center for commerce and creativity. The study highlights the broader implications of Western imperialism on local cultures and underscores the lasting legacy of these exchanges in Shanghai's modernization.

Introduction

In the chaotic aftermath of the Opium Wars, the Chinese city of Shanghai emerged as a place where Chinese and Western artistic and institutional practices met in a collision of cultures. Against the backdrop of British imperialism in China inaugurated by the Opium Wars and a declining Qing dynasty, the port of Shanghai was opened to the West. In 1842, the Treaty of Nanjing established the city as a "treaty port," opening trade to British merchants and allowing them to establish a sovereign concession outside the purview of Chinese laws. Designed by the British to counteract China's earlier restrictive policies regarding foreign presence in China, this treaty port opened Shanghai to international trade, visitation, and residency. The city was divided into two jurisdictions: the old walled city of Shanghai in the south, which remained under Chinese control, and the area to the north, which was allocated to foreign powers, including Great Britain and the United States (Wue, *Art Worlds: Artists, Images, and Audiences in Late Nineteenth-Century Shanghai*). Thus, Shanghai became an economic and cultural hub that was situated at the crossroads of traditional Chinese and Western cultures. As a result, the opening of the city's ports precipitated an increased presence of Western influences in fields such as art, architecture, and science, leading Shanghai artists, officials, and intellectuals to incorporate foreign cultural styles into Chinese paradigms, resulting in new ways of viewing, understanding, portraying, and interacting with the world around them.

The Collision of Chinese and Western Art in Post-1842 Shanghai

Increased Western influences in post-1842 Shanghai played a pivotal role in fostering new artistic styles among the city's artists. For instance, Ren Yi, also known as Ren Bonian (1840-1895), was a Shanghai painter renowned for his novel painting style, which blended Chinese and Western influences. After moving to Shanghai from a rural village, Ren was exposed to Western artistic methods, such as Western-style pencil-sketching techniques, which he learned from a brother in the Roman Catholic Church, itself a Western import (Lai). Eventually, he began combining traditional Chinese artistic techniques with Western ones, particularly in his portraits, in which he rendered facial features using chiaroscuro, a Western method of depicting light and shade while portraying draperies and settings in a more Chinese calligraphic style ("Art in Late-Nineteenth-Century Shanghai"). In her article "Remapping Borders: Ren Bonian's Frontier Paintings and Urban Life in the 1880s," art historian Yu-chih Lai contends that, in his portraits, Ren utilized "techniques probably learned from his father, adding to these Western shading techniques inspired by the wide variety of Western visual materials present in the dazzlingly international environment of Shanghai (Lai)." In addition, she states, Ren adopted the Western "penchant for three-dimensional volume (Lai)." Overall, as art historians Julia F. Andrews and Kuiyi Shen assert in their book chapter titled "Chinese Art in the Age of Imperialism," Ren's "personal style resulted from an unprecedented synthesis of the rich artistic traditions that met in Shanghai," including "the appeal of Chinese folk art" and "the realism and bright color of Western art (Andrews)." Besides Ren, another Shanghai artist who was influenced by Western artistic methods was Wu Youru (1850-1894), a painter from the present-day Jiangsu Province who utilized Western techniques in his representation of light and shadow as well as in his depiction of perspective—techniques that he combined with the Chinese style of fine-brush painting. In addition, Wu's paintings sometimes commingled objects from East and West. For instance, in his work *Banquet in Yu Garden* (1880, see Fig. 1) Kun portrays a banquet that features Chinese chopsticks alongside Western tableware (Kun). In these cross-cultural coalescences of style and subject, the works of Ren Yi and Wu Youru are examples of new ways that, in reaction to Western influences, Shanghai artists portrayed the world around them.

After the Treaty of Nanjing, Shanghai's emerging mercantile class developed a growing interest in purchasing art as a way to showcase their wealth. In response, local artists began utilizing Western printing and advertising techniques to mass-produce and promote their works. One of the leading publications in which these artists promoted their work was *Shenbao*, founded in 1872 by the British entrepreneur Ernest Major and became China's foremost newspaper during the late nineteenth century (Wue, *Art Worlds: Artists, Images, and Audiences in Late Nineteenth-Century Shanghai*). According to art historian Roberta Wue, this Western-style periodical, which was published in Shanghai, has been widely recognized for "its pioneering journalism and national impact" as well as "for its influence in defining and participating in Shanghai culture and identity (Wue, *Art Worlds: Artists, Images, and Audiences in Late Nineteenth-Century Shanghai*)." For example, one advertisement that appeared on the front page

of this influential newspaper, which promoted an artwork by the renowned Ren Yi, read, “[Ren Yi]’s painting brush is quite different and unlike what people usually hang in their reception halls and [yet] is quite suitable (Wue, “Selling the Artist”).” Indeed, works like the one advertised were unlike any previous artworks made in China in that they were produced using the Western printing technique of lithography, which, not unlike the Western-style pencil-sketching techniques taught to Ren by a Roman Catholic monk, was introduced to China by Jesuit missionaries (Wue, “Selling the Artist”). Chinese artists in Shanghai utilized Western print culture not only to advertise their works but also to produce them on an unprecedentedly large scale using, as another advertisement in *Shenbao* described it, an “amazing and clever Western technique [that] has never before been seen in China (Wue, “Selling the Artist”)!” In this way, the residents of Shanghai were newly exposed to the Western phenomenon of mass advertising. While the printing press had existed in China centuries before its first use in Europe, this Western way of producing and distributing mass-produced images radically redefined media in the city (Cooper). Although *Shenbao* was an excellent way to advertise artists’ works, it generally only included famous artists. Nonetheless, Wue states that lesser-known artists still found ways to promote themselves in Shanghai by self-printing and displaying advertisements throughout the city. One British expatriate observed that “notices and advertisements” were frequently placed near gateways and on streets in order to promote goods ranging from kerosene to medicine to works of art (Wue, *Art Worlds: Artists, Images, and Audiences in Late Nineteenth-Century Shanghai*). Mass-produced print advertisements created new profit opportunities, allowing artists to reach a wider audience, at the same time that Western printing techniques allowed the large-scale production of various versions of an image that were, depending on factors like size and color, sold for different prices, allowing households of different income levels to purchase the work (Wue, “Selling the Artist”). This mass production of artwork, along with the advertisements in *Shenbao*, demonstrates the entrepreneurial reaction of Shanghai artists to Western influence. By leveraging Western printing technologies and advertising techniques, these artists developed new ways of interacting with their clientele and disseminating their works throughout Shanghai and beyond.

The Bund: A Symbol of British Influence in Shanghai’s Urban Development

Another major way in which Western influences shaped Shanghai after 1842 was through urban planning and architecture, particularly the development of the waterfront area known as the Bund. Located on the banks of the Huangpu River, the Bund, according to scholars of architecture Yingchun Li and Weijen Wang, had become, by the early twentieth century, “the prime icon of Chinese economic strength and vigor (Li and Weijen).” Originally, however, this iconic urban landscape was a British Settlement, as specified by the Treaty of Nanjing. Indeed, as Li and Wang state in their 2010 presentation “Shaping the Bund: Public Spaces and Planning Process in the Shanghai International Settlement, 1843-1943,” the idea of developing the Bund “into the foremost urban center was not begotten in Shanghai” but “has roots deep in the practice of British colonial town planning (Li and Weijen).” As in other British colonial port cities,

including Bombay, Hong Kong, Singapore, and Yokohama, the Bund in Shanghai became the commercial and financial center of the metropolis. In the decades after the creation of the British Settlement, residents completed a number of projects to make the area look and feel more like a “British town,” including the construction of sporting clubs, dramatic societies, a racetrack, and a library (Wasserstrom). The Settlement’s municipal council and police force were based on those that existed in Great Britain, and the Public Garden, completed in 1867, was based on the urban planning notion, popular in Victorian Britain, that such a park was needed in order “to serve as the ‘lungs’ of a city (Wasserstrom).” Over time, an increasing number of Victorian features were added to the garden, including a bandstand, a decidedly British architectural element that would not generally be found in a Chinese city (Wasserstrom). The construction of a baroque pavilion, the cultivation of flower beds with seeds that had been shipped from England, and the placement of iron lounging seats that had also been imported from England further contributed to the British look and feel of the Bund (Li and Weijen). At the same time, however, the Chinese influence on the area was substantial, as the Land Regulations that governed development in the Bund were promulgated by the British Consul only after two years of negotiation with Shanghai’s governor. Most notably, the governor insisted that the towing path along the river “should be reserved both physically and functionally, and [that] its standard width must be 30 feet according to the requirement of the Chinese government (Li and Weijen).” In addition, he specified that the midpoint of the Bund Line—that is, the stretch of the Huangpu River along which the Bund is located—would be the location of the Chinese customs house (Li and Weijen). Thus, the urban planning and architecture of the Bund reflected a mixture of British and Chinese styles, demonstrating new ways in which Chinese officials came to visualize urban space as a result of British influences.

The Role of the Shanghai Polytechnic Institution in Cross-Cultural Knowledge Exchange

The Treaty of Nanjing also opened up Shanghai to foreign academics, allowing the city to serve as a nexus of collaboration between foreign scholars and their Chinese counterparts at newly founded educational institutions designed to facilitate cross-cultural learning. For instance, the establishment of the Shanghai Polytechnic Institution and Reading Rooms was initially proposed by Walter Medhurst, the British consul in the city, “to extend the knowledge of the Chinese in regard to Foreign countries and topics generally, and thereby to promote good feeling between foreigners and Chinese (Wright).” John Fryer, a British sinologist who served as the first director of the Shanghai Polytechnic, actively collaborated with Chinese educators, including the natural scientist Xu Shou and the mathematicians Li Shanlan and Hua Hengfang. Through these collaborations, Western scientific terms were translated into Chinese, a task that many Western scientists claimed was impossible because, as the historian of science David Wright notes, they thought that the Chinese language “was too ‘vague’ to be able to carry the precise terminology of science (Wright).” However, Fryer and his Chinese collaborators not only proved these naysayers wrong but also, in the process of translation, developed a system of naming chemical compounds that formed the basis of modern chemical nomenclature (Wright).

Thus, cooperation between British and Chinese intellectuals in Shanghai not only served to spread Western knowledge to China but also made an essential contribution to the global scientific community. Beginning in 1885, Chinese boys between the ages of ten and fourteen were admitted to classes at the Shanghai Polytechnic in order to provide them with the skills and knowledge required for advanced scientific study (Biggerstaff). Thus, Shanghai became a place where not only established Chinese academics but also budding ones could benefit from Western knowledge. In these ways, Western influences led Chinese scholars to develop new ways of understanding the sciences and interacting with educational institutions.

Conclusion: The Transformation of Shanghai into a Cosmopolitan Hub

Noting that “Shanghai accommodated one of the first and most important Western settlements in China”—one that drew “merchants, migrants, and foreign visitors from every part of the globe”—Yu-chih Lai characterizes nineteenth-century Shanghai as one of “the most exciting and complex cities in the land (Lai).” βIndeed, the ratification of the Treaty of Nanjing, which opened up this crucial Chinese port city to foreign trade and visitation, marked a pivotal moment in Shanghai’s history and played a major role in shaping its nineteenth-century development. By leading to the emergence of new artistic styles, the mass production of works of art and print advertisements, the confluence of Chinese and British architecture and urban planning in the Bund, and the establishment of joint Chinese-British educational institutions to promote cross-cultural learning, these influences led Shanghai artists, officials, and intellectuals to blend foreign styles with their own, thereby ushering in novel perspectives, interpretations, representations, and engagements with their surroundings and contributing to the transformation of Shanghai into a dynamic, cosmopolitan global hub. Questions remain, however, regarding the extent to which Western influences permeated other cultural spheres, such as literature, music, law, and government.

光緒庚辰春三月
 德國皇孫海哪哩
 來游滬上 觀察
 劉公觴於豫園中
 峯迴路轉城市山
 林之仰山堂是日
 也重軒洞開冠帶
 雲集福勳承命供
 張亦幸叨陪末坐
 俯仰禮儀之備盱
 衡品類之盛洵一
 時嘉會也爰屬畫
 工繪圖以紀之畝
 成並綴數語於紙
 尾 錢塘陳福勳題



Figure 1: *Banquet in Yu Garden* by Wu Youru

Works Cited

- Andrews, Julia Frances. *The Art of Modern China*. Berkeley : University of California Press, 2012, http://archive.org/details/artofmodernchina0000andr_y9v7.
- “Art in Late-Nineteenth-Century Shanghai.” *The Metropolitan Museum of Art Bulletin*, vol. 58, no. 3, 2001, p. 10. DOI.org (Crossref), <https://doi.org/10.2307/3269181>.
- Biggerstaff, Knight. “Shanghai Polytechnic Institution and Reading Room: An Attempt to Introduce Western Science and Technology to the Chinese.” *Pacific Historical Review*, vol. 25, no. 2, May 1956, pp. 127–49. DOI.org (Crossref), <https://doi.org/10.2307/3635292>.
- Cooper, Helen. “The Origins of the Early Modern.” *Journal for Early Modern Cultural Studies*, vol. 13, no. 3, 2013, pp. 133–37.
- Kun, Zhang. “Making a Scene.” *China Daily*, 7 Apr. 2023, https://www.chinadaily.com.cn/a/202304/07/WS642f40b9a31057c47ebb8ba5_5.html.
- Lai, Yu-chih. “Remapping Borders: Ren Bonian’s Frontier Paintings and Urban Life in 1880s Shanghai.” *The Art Bulletin*, vol. 86, no. 3, Sept. 2004, p. 550. DOI.org (Crossref), <https://doi.org/10.2307/4134446>.
- Li, Yingchun, and Wang Weijen. *Shaping the Bund: Public Spaces and Planning Process in the Shanghai International Settlement, 1843-1943*. 2010. Semantic Scholar, <https://www.semanticscholar.org/paper/SHAPING-THE-BUND%3A-PUBLIC-SPACES-AND-PLANNING-IN-THE-Li-Weijen/490b24e1de9c5164bb83b56cd08a99749e1699d9>.
- Shanghai Chinese Polytechnic Institution and Reading Rooms. Report. 1875, <https://www.google.com/books/edition/Report/GhILAAAIAAJ?gbpv=0>.
- Wasserstrom, Jeffrey N. *Global Shanghai, 1850-2010: A History in Fragments*. Routledge, 2009.
- Wright, David. “John Fryer and the Shanghai Polytechnic: Making Space for Science in Nineteenth-Century China.” *The British Journal for the History of Science*, vol. 29, no. 1, 1996, pp. 1–16.
- Wue, Roberta. *Art Worlds: Artists, Images, and Audiences in Late Nineteenth-Century Shanghai*. Hong Kong University Press, HKU, 2014, <https://muse.jhu.edu/pub/175/monograph/book/35849>.
- . “Selling the Artist: Advertising, Art, and Audience in Nineteenth-Century Shanghai.” *The Art Bulletin*, vol. 91, no. 4, 2009, pp. 463–80.

Protection or punishment: analyzing how policy shapes experiences of homelessness in New York City and Miami By Matthew Pollak

Abstract

Homelessness remains a critical issue in American cities, with over 650,000 people living without stable shelter in 2023, and little consensus exists on the best policy approaches to address this crisis. Homelessness is not only an important subject of academic study in sociology, urban studies, and a variety of other disciplines, but it is also a humanitarian crisis, a major challenge for policymakers, and a critical issue for urban residents. Informed by an in-depth policy analysis alongside key informant interviews with individuals experiencing homelessness, this paper investigates the contrasting homelessness policies of New York City and Miami, and the lived experiences of homeless people under each policy framework. This paper analyzes the policy frameworks of both Miami and New York, comparing how their approaches have shaped the experiences of homeless individuals and assessing the effectiveness, ethics, and sustainability of each model. While Miami's punitive approach offers a short-term reduction in visible homelessness, it does so at the expense of human rights and long-term stability, and homeless residents speak of frequent police encounters and harassment. Though hindered by a housing affordability crisis that leads homeless individuals to experience an overcrowded shelter system, New York City's supportive framework provides a more compassionate and potentially more effective path forward, placing long-term housing and dignity ahead of short-term visibility. The findings indicate that although New York City's model is more humane and effective in addressing the homeless crisis, both cities would benefit from: expanding mental health and addiction treatment programs; mitigating the gentrifying effects of urban redevelopment; implementing scalable supportive housing programs; and reevaluating policing practices related to homelessness.

Introduction

Homelessness is one of the trenchant—and vexing—issues facing American cities today. Whether living in tents, sleeping in cars, or lying on a doorstep, homeless Americans are suffering, but identifying effective policies to address the growing crisis remains an elusive goal. Two cities, Miami and New York City, have implemented contrasting policies to meet the challenge of managing a heterogeneous population unable to live anywhere but on the streets. Miami has opted for a punitive approach, relying on criminalization and law enforcement to reduce visible homelessness, raising concerns about displacing vulnerable populations and eroding their basic human rights. In contrast, New York City's supportive social framework for the homeless, backed by legal mandates, prioritizes shelter and long-term housing solutions, although it struggles under the weight of an overwhelming demand for affordable homes. Both cities claim progress, but their methods reveal a deep divide between visible, short-term results and long-term sustainability.

Miami's approach largely focuses on reducing visible homelessness through criminalization, including policies like anti-camping ordinances and increased policing. This has resulted in a significant decrease in street homelessness but at the cost of weakening social protections for the homeless. Though Miami has also implemented the Housing First model through the Homeless Trust, this effort exists alongside its punitive strategies, creating a tension between housing assistance and ongoing displacement caused by criminalization.

New York City, by contrast, has approached homelessness through a legal mandate to provide shelter, established by the *Callahan v. Carey* New York County Supreme Court decision. This policy has led to the creation of the largest shelter system in the U.S., and the city has attempted to reduce reliance on shelters through permanent housing solutions like Housing First. Yet, despite these efforts, New York faces rising homelessness due to a severe lack of affordable housing. The supportive policies in place are further strained by the city's expensive rental market, leaving many stuck in shelters without viable long-term housing options.

This paper analyzes the policy frameworks of both Miami and New York, comparing how their different approaches have shaped the experiences of homeless individuals and assessing the effectiveness, ethics, and sustainability of each model. The paper shows that while Miami's punitive approach offers a short-term reduction in visible homelessness, it does so at the expense of human rights and long-term stability, and homeless residents speak of frequent police encounters and harassment. Though hindered by a housing affordability crisis, which leads homeless individuals to experience an overcrowded shelter system, New York City's supportive framework provides a more compassionate and potentially more effective path forward, placing long-term housing and dignity ahead of short-term visibility.

The paper proceeds with a review of relevant academic literature that examines homeless policy in U.S. cities, followed by an explanation of the research methods used in the paper. It goes on to first explain, and then to analyze, the homelessness policy of New York and Miami in turn. After the policy analysis, the paper turns to the lived experiences of homeless individuals in each city, and how their experiences have been shaped by policy approaches. Bringing New York and Miami into direct conversation in the penultimate section, the paper evaluates the homelessness policy of the two cities in direct comparison. Finally, the paper suggests policy reform for both cities. It shows that although Miami would benefit from a shift from criminalization to comprehensive support systems and New York City from increased investment in affordable housing, both cities would benefit from: expanding mental health and addiction treatment programs; mitigating the gentrifying effects of urban redevelopment; implementing scalable supportive housing programs; and reevaluating policing practices related to homelessness.

Literature Review

In the academic literature on homelessness policy, the two main approaches—punitive measures and the Housing First model—are widely debated. Wilking et al. offer a detailed case study of Chico, California, where anti-camping ordinances are used to remove homeless

individuals from public spaces. This method focuses on reducing the visibility of tents and personal belongings in urban areas, a strategy similarly employed in cities like Miami. However, these punitive measures do not address the root causes of homelessness, such as housing insecurity and mental health issues, leading to a cycle of displacement rather than long-term resolution (Wilking et al. 2). Robinson expands on this critique by showing that criminalization policies not only fail to resolve homelessness but often worsen conditions by funneling individuals into the criminal justice system, where they remain marginalized and unsupported (Robinson 83). These perspectives highlight the limitations of punitive measures, which prioritize public order over individual welfare.

The Housing First model takes a fundamentally different approach by prioritizing housing without preconditions such as sobriety or employment. Baker and Evans argue that Housing First offers a more humane and effective framework for addressing chronic homelessness, focusing on stability and long-term support rather than punishment. They note that Housing First programs have shown success in reducing homelessness in cities like New York, but the model faces challenges, particularly regarding affordability and scalability (Baker and Evans 12). Ly and Latimer corroborate this, emphasizing that while Housing First has a positive impact on reducing homelessness, the economic constraints of implementing the model—especially in cities where affordable housing is scarce—hinder its broader application (Ly and Latimer 62). Despite these challenges, the Housing First model remains a widely supported alternative to punitive measures.

Speer (2019) further contributes to this discourse by examining the intersection of homelessness and urban redevelopment, focusing on the displacement caused by gentrification efforts. In cities undergoing rapid urban renewal, homeless populations are frequently forced out of public spaces to make way for beautification projects. Speer notes that this displacement often occurs without offering viable long-term housing solutions; further entrenching the cycle of homelessness (Speer 580). Augustine and Kushel support this analysis, showing that urban beautification and gentrification often prioritize economic interests over the needs of marginalized populations, pushing homeless individuals into even more precarious situations (Augustine and Kushel 45).

Together, these studies highlight the deep divide between punitive and supportive homelessness policies. Wilking et al. and Robinson critique the short-term focus of criminalization, while Baker and Evans, along with Ly and Latimer, advocate for Housing First as a more sustainable, humane solution. In both the studies by Speer and Augustine and Kushel (2019, 2022), additional insight is provided into how urban development exacerbates the issue, showing that without structural changes to affordable housing and urban planning, neither approach will fully resolve homelessness.

Method

This research adopts a comparative analysis of homelessness policies, focusing on the differing strategies employed by Miami and New York City. The analysis combines secondary

sources, such as academic literature and public data on homelessness policies and outcomes, with primary data gathered from interviews with key informants. These informants include unhoused individuals living in both cities who provided critical insights into the life of the unhoused affected by each city's policies. Interviews were conducted in various locations, including shelters, public spaces, and areas where unhoused populations frequently reside, ensuring a diverse range of perspectives and experiences. These interviews focused on the participants' interactions with law enforcement, access to housing and shelters, and their overall experiences under the distinct frameworks of punitive versus supportive homelessness policies. The analysis of these interviews is complemented by publicly available data on homelessness trends, city reports, and legal documents. Policy analysis is also a key method in this research. This mixed-methods approach allows for a holistic understanding of the impact of each city's homelessness policies on individual well-being and broader social outcomes.

Developing policy frameworks

Development of Miami's Policy Framework

Miami's approach to homelessness has historically been shaped by a desire to reduce the visibility of homelessness in public spaces, leading the city to adopt punitive measures aimed at enforcing public order. The roots of Miami's current homelessness policy can be traced back to the *Pottinger v. City of Miami* case in 1992, which has fundamentally altered how the city handled homelessness for decades. Brought on behalf of homeless individuals, the lawsuit argued that the city's practice of arresting people for performing life-sustaining activities—such as sleeping or eating in public—is unconstitutional. The court ruled in favor of the plaintiffs, resulting in the Pottinger Agreement barring the city from arresting homeless individuals for these daily, life-sustaining activities unless adequate shelter is available (Daynes).

The Pottinger Agreement was seen as a victory for homeless rights advocates. The legislation placed limits on the city's ability to enforce public space ordinances, much to the frustration of developers and business owners seeking to revitalize Miami's downtown. Over time, as the city prioritized urban development and beautification projects, pressure mounted to weaken the agreement. The tide shifted in 2014 when a settlement allowed the city to resume arrests in cases where shelter was available, effectively rolling back many of the protections afforded to homeless individuals under Pottinger (Daynes; Guido and Weinstock). This weakening of the statute announced a shift toward criminalization, wherein visible homelessness was increasingly treated as a public order issue rather than a humanitarian crisis.

At the heart of Miami's enforcement efforts is the city's reliance on anti-homelessness laws that effectively push homeless individuals out of public view without providing sustainable long-term solutions. This tactic has been closely linked to the city's broader urban beautification initiatives, prioritizing the appearance and cleanliness of public spaces. Developers and city planners have used these initiatives to transform Miami's downtown area from urban blight characterized by abandoned buildings, uncared-for streets, and visible rubbish to upscale

neighborhoods with high-end restaurants, commerce and manicured urban spaces. Displacing homeless populations to less visible parts of the city removes them from sight but does nothing to shore up the problem. Critics argue that this strategy only serves to marginalize homeless individuals further, as it fails to address the structural issues that contribute to homelessness, such as a lack of affordable housing (Daynes; Guido and Weinstock).

As Miami continues to grapple with its homelessness crisis, the city's policies have come under increasing scrutiny. While the Homeless Trust, established in 1993, has made strides in providing housing through its Housing First model, which focuses on placing homeless individuals in permanent housing without preconditions, the simultaneous focus on criminalization undermines these efforts. Funded through taxes placed on food and beverages, the initiative has helped reduce homelessness by providing more permanent solutions, yet these gains are often offset by the displacement caused by enforcement tactics (Homeless Trust 2; Guido and Weinstock). As such, Miami's homelessness policy remains caught between competing priorities: addressing homelessness as a social issue versus treating it as a public nuisance.

Development of New York City's Policy Framework

New York City's approach to homelessness is rooted in a supportive legal framework that distinguishes it from many other major cities. At the core of New York City's policies is the right to shelter, established through landmark litigation, which obligates the city to provide shelter to any homeless individual seeking refuge. However, despite this legal obligation, the city continues to face immense challenges related to homelessness, driven by issues like the migrant crisis, housing affordability, and the enduring effects of the COVID-19 pandemic (Maroño Porto).

The migrant crisis has been a significant driver of the recent increase in New York City's homeless population. In 2023, the city's shelter system reported over 90,000 people, including thousands of asylum seekers, many from Latin America (particularly Venezuela), as well as individuals from Africa, the Caribbean, and Mexico. Many of these migrants are fleeing from political instability, criminal violence, and economic hardships in their countries of origin. Recent academic analysis highlights that the surge of migrants into New York City's shelters has been driven by U.S. policy toward Latin America, including sanctions and economic conditions that have exacerbated crises in countries like Venezuela, pushing millions to flee. In fact, this large-scale migration has put substantial strain on New York's shelter system, making it one of the city's most pressing humanitarian challenges (Gonzalez). However, when accounting for the migrant population, the number of local homeless individuals has shown signs of stabilization. This distinction highlights the complexity of New York's homelessness crisis, which is shaped by both external factors like migration and internal challenges related to housing affordability (Otokiti, 3).

Affordability continues to be one of the city's primary drivers of homelessness. The rental market in New York City is among the most expensive in the world, and a significant

portion of the population faces an extreme rent burden, with families spending more than 30% of their income on housing (Otokiti 2), a situation which easily becomes unsustainable when costs for medical care rise or employment revenues decrease. The impact of gentrification has further worsened the situation, displacing many low-income families who can no longer pay for rent or keep pace with rising associated costs, for example, when cheaper food stores are replaced by high-end shops catering to refined tastes. Gentrification is particularly an issue in rapidly changing neighborhoods, such as Brooklyn and Harlem (Kirchheimer 613). The COVID-19 pandemic exacerbated these affordability issues, leading to widespread unemployment and evictions, further straining the city's already overstretched shelter system (Otokiti 1).

New York's policy framework for addressing homelessness was significantly affected by the *Callahan v. Carey* case in 1981. A landmark ruling, it gives every person the right to shelter, which made New York City legally responsible for providing shelter to all homeless individuals. This legal mandate transformed New York's shelter system into a vast public social function, which expanded rapidly in response to an exponentially growing homeless population. By 1985, the city's shelter expenditure had risen from \$8 million to over \$100 million annually, and the number of shelters had multiplied (Kirchheimer 608).

The city's shelter system, initially focused on emergency services, has grown to include transitional housing and specialized shelters for various homeless subpopulations, such as families, battered women, veterans, and individuals with mental health issues. The broadening of services reflects the city's commitment to providing not just temporary relief, but also pathways toward more stable housing for populations with specialized needs. However, the right to shelter has placed immense pressure on the city, with the number of individuals in shelters rising to over 62,000 by 2014 (Otokiti 2).

In addition to its shelter system, over the last four decades, New York has adopted several Housing First and other supportive housing policies to combat the rise in chronic homelessness, defined as people who have experienced homelessness for at least a year — or repeatedly — while struggling with a disabling condition such as a serious mental illness, substance use disorder, or physical disability (National Alliance to End Homelessness). Among the most notable initiatives was Mayor Bloomberg's Five-Year Plan (2004), which aimed to reduce chronic homelessness by transitioning individuals from shelters to permanent housing. The Five-Year Plan focused on creating supportive housing units for individuals with disabilities and chronic mental and physical illnesses, providing them with stable homes alongside access to essential services, such as mental health and substance abuse treatment (Markee 2).

Despite these efforts to create new and innovative housing solutions, the affordability crisis has curtailed the overall impact of these supportive initiatives. Recent research argues that the shortage of affordable housing continues to be the main barrier to reducing homelessness, with rent increases far outpacing wage growth in the city so that working individuals can no longer meet monthly rental payments (Otokiti 1). The result is that the sheer rise in homelessness neutralizes gains in aiding the homeless populations. For example, while Mayor Bill de Blasio's Turning the Tide initiative (2017) attempted to reduce street homelessness by building 90 new

shelters and expanding supportive housing, research suggests that these efforts have been insufficient given the exponential scale of the city's housing crisis (Otokiti 4).

Homelessness in New York City remains persistently high, with little chance for leveling off in sight. As of 2023, the total number of individuals in the shelter system, including both local residents and asylum seekers, exceeds the city's ability to wrestle with the problem (Otokiti 1). While counting the number of homeless remains an intractable problem, due to invisible homeless, such as families who double up by sleeping three or more to a bedroom, as well as clandestine individuals whose civil status is unclear, it is clear that New York is overwhelmed by the problem. The city's right to shelter ensures that no one is left without a bed, but the sheer volume of demand has led to overcrowding and stretched resources, including shelters and social funds, leaving many homeless individuals in a state of limbo, cycling between shelters and the streets.

Policy analysis

Effects of Programs and Legislation in Miami

Miami's adoption of punitive measures to address homelessness has been met with mixed outcomes, despite reductions in visible homelessness. While the Council on Homelessness reported a decline in the number of unsheltered individuals from 5,773 in 2005 to 3,440 in 2022, the true extent of these figures shows that many homeless individuals were simply pushed into other areas near Miami, such as Tampa Bay, Lee County, and Collier County, where the homeless population has risen dramatically in the past three years. These areas have seen sharp increases in homelessness, particularly in the past few years, with Collier County experiencing a 52% rise in homelessness in 2023 alone. This displacement results in hidden homelessness, where individuals are forced into areas with less public visibility but without access to essential services, like shelters or healthcare, exacerbating the problem (Karsen & Pellegrino).

Moreover, as Kalesnikate and Zamor have argued, this displacement creates a new cycle of criminalization. Incarceration becomes frequent for those unable to escape homelessness when the city applies anti-camping and loitering laws, such as HB 1365, which "prohibits local governments in Florida from allowing any person to 'regularly engage' in public sleeping or camping on public property, buildings, and/or sidewalks and streets" (Hodges). This criminalization of homelessness is not offset by sufficient affordable housing options; exacerbating the cycle where the homeless population has nowhere to go but the streets. Miami's policies ignore long-term solutions to homelessness, focusing instead on immediate, visible reductions in the downtown areas and public spaces prioritized for tourism and economic development.

The effect of urban development and gentrification, in particular, have contributed significantly to the city's homelessness crisis since developers equate homelessness with lost value and entrenched urban blight. Goldfischer notes that as Miami's downtown areas undergo rapid revitalization, the homeless population becomes collateral damage in the city's quest for

beautification. Gentrification has driven up property values, displacing low-income families and individuals who once occupied affordable housing, perhaps for generations, near these redeveloped spaces, and who lose not only their housing but also their communities. The rise in rent forces many individuals, including significant numbers of children, into homelessness when their parents no longer have resources to keep them housed. The city's Homeless Trust and housing initiatives, such as Camillus House, Miami Rescue Mission, and Lotus House, struggle to keep pace with demand.

The harsh effects of these programs and legislation on Miami's homeless population have left human rights advocates criticizing Miami's use of criminalization and its impact on the dignity and rights of homeless individuals. According to Kalesnikate and Zamor, Miami's aggressive policing tactics, including frequent encampment sweeps, violate the basic rights of the homeless, treating them as a public nuisance rather than a vulnerable population in need of support. Little, if any, consideration is given to the humanitarian crisis when the focus is on protecting rising real estate values. These concerns highlight the need for a policy shift toward more inclusive, human-centered approaches to homelessness.

Effects of Programs and Legislation in New York City

In contrast to Miami, New York City's right-to-shelter law, established through *Callahan v. Carey*, has also created dependency on the city's already overwhelmed shelter system. Individuals who have no other option but to return to the shelters again and again, as the safety net is unable to keep them housed. The cycle nature of shelter-to-street risks inuring individuals to their own homelessness, triggering mental-health issues or worsening pre-existing issues, in some cases, making it all the more difficult for the population to find a way back into society. With shelter populations exceeding 132,000 by 2024, the city's focus on temporary housing has limited the effectiveness of these efforts in providing long-term solutions (Coalition for the Homeless).

While New York City's Housing First initiatives aim to reduce reliance on shelters by transitioning individuals into permanent supportive housing, these programs have not been able to keep pace with the rising number of homeless individuals entering the system. There are simply not enough housing units to absorb the need, so the homeless rely on shelters pushed beyond their limits. The shelter system remains a stopgap rather than a sustainable solution, as families and individuals continue to cycle in and out of shelters without access to affordable housing (Coalition for the Homeless).

Consequently, New York City's most significant challenge in addressing homelessness is its housing affordability crisis. The NYU Furman Center has highlighted that nearly half of New York City renters are considered rent-burdened, spending more than a third of their monthly income on housing, a situation which is untenable, and suggests that rather than slowing the movement from housed to unhoused, circumstances may continue to worsen. Bretherton and Pleace (2023) note that without significant investment in affordable housing, New York City will continue to face a growing homelessness problem, as the number of housing units affordable to

low-income residents is far outweighed by demand. Converting hotels into shelters, which are costly and unsustainable in the long term, is just another temporary measure.

An important aspect to note in the recent, massive rise in New York’s homeless population is the migrant crisis that began in the summer of 2022, with thousands of asylum seekers entering NYC Department of Homeless Services (DHS) shelters, essentially causing the shelter population to inflate up to unfathomable numbers. The incremental uptick in shelter population is illustrated in the graph below, provided by the New York City Office of the Comptroller. This influx in the shelter population caused the City to impose 30- and 60-day limits on shelter stays for many of the new asylum seekers. While New York continues to lead the nation in providing shelter, the system’s capacity is unable to keep up with the rising demand, leaving many individuals in unsustainable, temporary housing situations.

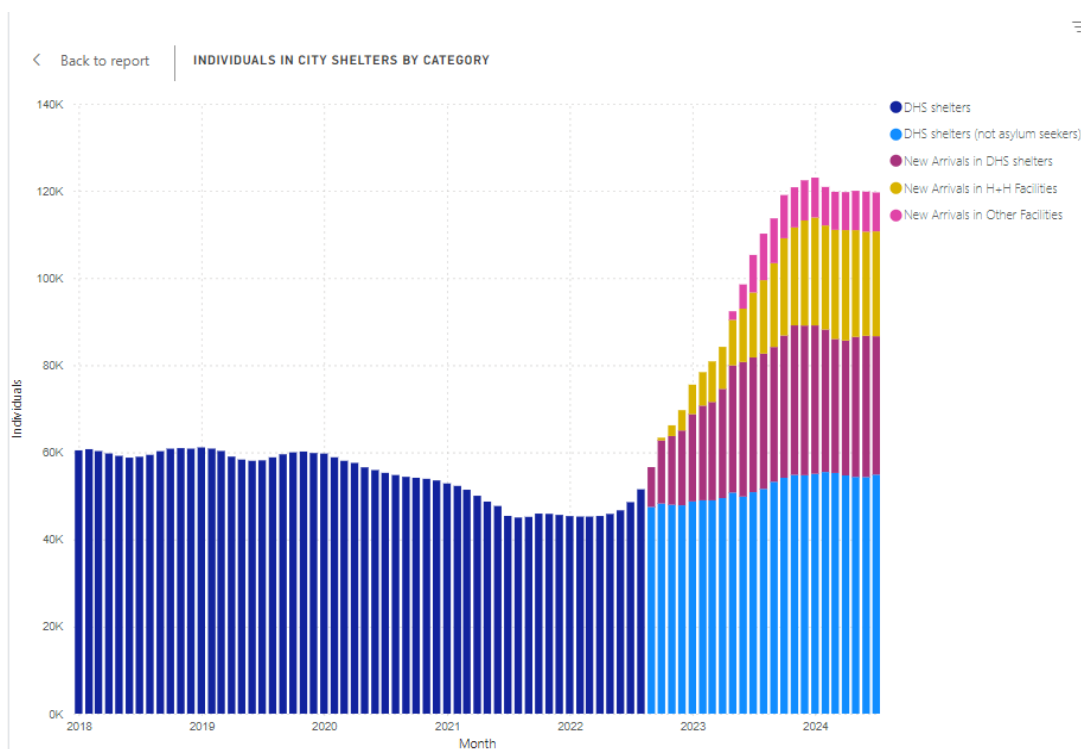


Figure 1: The number of individuals in New York City shelters by various categories from 2018 to 2024 (New York City Office of the Comptroller)

Lived experiences of homeless policy

Experiences Shaped by Policies in Miami

Interviewing unhoused individuals in Miami gives first-hand accounts and insights regarding Miami’s reliance on punitive measures and reveals the harsh realities of living under these policies. With such intense struggle to find a shelter for most unhoused individuals, they are forced to set up encampments or tents for the night. Despite the reduction in visible

homelessness, punitive policies such as anti-camping ordinances like HB 1365 and frequent police raids have displaced many individuals without addressing the root causes of homelessness, including affordable housing shortages and employment instability.

Melvin, who has lived on the streets of Miami for five years, shared his frustration with the shelter system and the police. After being hospitalized for a stroke, he returned to find that his belongings had been stolen and his spot in a local shelter given away. “They stole all my stuff, my clothes, shoes, property,” he explained, visibly upset. The police were called not to help but to tell him to leave the area. “I was fighting for my life in the hospital, but when I got back, my stuff was gone, and they told me to get out,” Melvin recalled. His story is emblematic of the instability and lack of support many face when trying to access shelter in Miami.

Amos, who has been in Miami for nine years, also discussed the constant struggle to secure a place in the city’s shelters. “The line for emergency housing is so long,” he remarked, emphasizing how the shelter system is often overwhelmed by the sheer number of people needing help. Despite being placed on a waiting list, Amos continues to face uncertainty, not knowing when he might get a permanent place to stay. “It’s just hard, man,” he added, expressing the emotional toll of living in a state of perpetual displacement.

For many, public perception has been shaped by the visibility of homelessness and the criminalization of encampments, worsening the situation. Darryl, another individual who has experienced homelessness for three years, expressed how the city’s laws have made it increasingly difficult for homeless people to find stability. “They start October 1st, nobody can be homeless anywhere in the city of Miami,” he explained, referring to new ordinances targeting people sleeping in public spaces. This crackdown leaves individuals like Darryl with fewer options for safety and shelter, increasing their vulnerability to harassment and violence.

Edgardo, who came to Miami from Texas, shared how police harassment compounds the challenges posed by homelessness. “They’ll wake you up in the middle of the night, tell you you can’t be here, and you’re just sleeping,” he said, noting how officers often target homeless individuals for merely existing in public spaces. “I told them, ‘Y’all need to stop harassing me. I’m not doing anything wrong,’ but they don’t care,” Edgardo added, displaying the sense of powerlessness many feel when trying to navigate Miami’s punitive policies.

The criminalization of homelessness in Miami, coupled with the lack of adequate shelter and social services, creates a cycle of displacement and exclusion. As Melvin explained, “We didn’t choose to be out here. We were pushed out here because the economy is high, the bills are high, but the pay rate is low.” His statement captures the systemic nature of homelessness in Miami, where individuals are forced onto the streets by economic pressures, only to face further marginalization under the city’s punitive policies.

Experiences Shaped by Policies in New York

Interviews I conducted with unhoused individuals in New York City reveal the gap between the intent of New York City’s supportive housing policies and the reality faced by

homeless individuals. While the city's shelter system aims to provide immediate relief, overcrowding, theft, and bureaucratic challenges illustrate the limitations of these policies.

Supportive housing and shelter systems in New York City offer refuge from the harsh winters and dirty city floors, but they also present significant challenges, particularly in terms of quality of shelter life and transitioning into permanent housing. Marvin, a homeless man I interviewed in Washington Square Park, discussed the difficulties of simply securing a bed, emphasizing how the shelter system leaves many feeling forgotten. "If people knew what we had to go through to get a bed at night, they'd pay more attention," Marvin exclaimed. Others, like Randy, highlighted the insecurity in City shelters, recalling how his belongings were stolen during his short stay: "They took my clothes, my food, everything." Such experiences reflect the broader issue of safety and stability within the shelters.

For many, transitioning from shelters to permanent housing remains a significant obstacle. Larry, another interviewee, expressed his frustration with the bureaucratic barriers in the housing system: "I've been in the system for years, but the housing just isn't there, man." His experience, which echoes the calls of many, reveals the inefficiencies in securing permanent housing, leaving many stuck in a state of temporary shelter without viable long-term solutions.

Despite New York City's progressive policies aimed at reducing homelessness, police harassment continues to be a persistent issue. Many homeless individuals, like Marvin, shared experiences of frequent police encounters, even when they were not engaging in illegal activities. "Even if you're just sitting there, they come around, tell you to move," Marvin said. The policing of unhoused individuals often contradicts the city's efforts to provide support through shelters and housing programs. This ongoing tension between law enforcement and the homeless population displays the limits of supportive policies when they are not paired with the correct style of policing in regards to how they interact with the homeless. As Marvin's account shows, these experiences of police harassment contribute to the broader social exclusion that many homeless individuals face, making it even harder for them to access services or maintain a sense of security.

Evaluating New York City and Miami in comparison

The contrasting approaches to homelessness adopted by New York City and Miami reveal both the strengths and limitations of their respective policy frameworks. New York City's right to shelter and supportive housing policies are rooted in a legal obligation to provide emergency shelter to all homeless individuals. This approach has ensured that tens of thousands of individuals are not left without a bed, addressing immediate needs. However, the housing affordability crisis and the overburdened shelter system have led to long-term challenges in transitioning individuals from temporary shelters into permanent housing. Despite the city's progressive policies, police harassment continues to be an issue, as recounted by interviewees like Marvin, highlighting the disconnect between policy intent and on-the-ground realities.

In contrast, Miami's reliance on punitive measures and criminalization of homelessness has led to a visible reduction in street homelessness but at the expense of human rights and

long-term stability. The frequent displacement of individuals from public spaces, combined with aggressive policing, has led to social exclusion and repeat homelessness. The city’s simultaneous focus on criminalizing homelessness and providing housing through initiatives like the Homeless Trust creates a complex dynamic. While some are housed through Miami’s housing programs, many remain stuck in a cycle of displacement, as evidenced by the interviews with individuals like Melvin and Darryl, who spoke about repeated encounters with law enforcement and lack of stable housing solutions.

A comparison of the two cities shows that while New York City offers a more compassionate and structured framework, its systemic issues, such as overcrowded shelters and a lack of affordable housing, hinder long-term solutions. Miami’s approach, on the other hand, prioritizes public order and aesthetics, often at the cost of the dignity and safety of its homeless population. The result is a stark divide in the experiences of homeless individuals in each city, where New York offers immediate relief but struggles with long-term housing, and Miami focuses on short-term visibility with little regard for sustainable support.

While both cities are making efforts to address homelessness, the supportive framework of New York City, despite its challenges, offers a more humane and potentially sustainable model when compared to Miami’s punitive approach. However, without addressing underlying issues such as affordable housing and police reform, both cities will continue to face significant barriers in effectively reducing homelessness and improving the lives of their homeless populations.

Aspect	New York City	Miami
Policy	Right to Shelter, Supportive Housing	Criminalization, Housing First through Homeless Trust
Main Focus	Long-term housing, legal obligation	Immediate reduction of visible homelessness
Challenges	Overcrowded shelters, housing affordability	Social exclusion, displacement, rights violations
Police Interaction	Instances of harassment despite progressive policies	Many reports of aggressive policing, frequent displacement
Long-term Outcomes	Limited scalability of supportive housing	High rates of repeat homelessness, short-term relief

Figure 2: Table comparing homeless policy between New York City and Miami
Future of Homeless Policy

To create meaningful, long-term solutions to homelessness, both Miami and New York City must take a step back and reimagine their current approaches. Each of their strategies reveal significant gaps in their systems that must be addressed to reduce homelessness effectively and sustainably.

Shift from Criminalization to Comprehensive Support Systems (Miami)

Miami's heavy reliance on punitive measures, such as anti-camping ordinances and law enforcement, only displaces the problem without addressing the root causes of homelessness. Rather than focusing on visibility, the city should expand its Housing First model, currently offered through the Housing Trust, which has already shown success in providing permanent housing without preconditions. This approach should be coupled with robust support services, including mental health care, addiction treatment, and job placement programs, to help individuals transition out of homelessness. Decriminalizing homelessness and redirecting funds from policing to social services would promote a more compassionate and effective strategy.

Increase Investment in Affordable Housing (New York City)

New York City's right-to-shelter law has ensured that no one is left without a bed, but the city's shelters are overburdened, and the lack of affordable housing exacerbates the cycle of homelessness. A more aggressive push to build affordable housing is essential. This includes expanding the scope of the city's existing Housing First initiatives to secure more units dedicated to low-income individuals and families. Furthermore, policies that regulate rent and prevent evictions, such as stronger tenant protection laws and rent stabilization, are necessary to slow the influx of newly homeless individuals.

Implement Scalable Supportive Housing Programs (Both Cities)

Both Miami and New York City face challenges in scalability. While supportive housing programs have proven effective, the demand far exceeds supply. Future policy must focus on creating scalable solutions. This can be achieved by partnering with nonprofits and private developers to create affordable, mixed-income housing communities. By integrating affordable housing into broader urban development projects, cities can ensure long-term housing stability while fostering economic diversity.

Reevaluate Policing Practices Related to Homelessness (Both Cities)

In both cities, the relationship between the homeless population and law enforcement creates additional barriers to stability. New York's supportive policies are often undermined by persistent police harassment, while Miami's heavy-handed policing displaces individuals without offering meaningful alternatives. Both cities should adopt policies that emphasize harm reduction policing, where law enforcement officers work in tandem with social services to connect homeless individuals to housing and care rather than pushing them further into the margins. Training police officers to engage with homeless populations in a supportive and non-punitive manner could dramatically improve trust and access to services.

Address Urban Redevelopment and Gentrification (Both Cities)

Gentrification and urban redevelopment continue to displace low-income families and exacerbate homelessness in both Miami and New York City. Policymakers must prioritize

inclusive urban planning that considers the needs of vulnerable populations. Inclusionary zoning laws should mandate a significant percentage of new housing developments to be affordable, ensuring that gentrification doesn't continue to displace individuals. Additionally, developers should be incentivized through tax credits or subsidies to invest in affordable housing projects as part of broader redevelopment efforts.

Expand Mental Health and Addiction Treatment Programs (Both Cities)

A significant portion of the homeless population in both cities struggles with mental illness and substance use disorders, often worsening their ability to secure stable housing. Expanding access to mental health care and addiction treatment through community-based programs would not only address these issues but also reduce long-term reliance on shelters. Integrating health services directly into housing programs could offer a more holistic approach to stability, addressing both the physical and mental well-being of homeless individuals.

Ultimately, the future of homelessness policy in both Miami and New York City depends on addressing the root causes of homelessness—economic inequality, lack of affordable housing, and inadequate support services. Miami must move away from criminalization toward a more human-centered approach, while New York must focus on improving the scalability of its housing programs and addressing its housing affordability crisis. Both cities need to reevaluate their policing practices, invest in mental health and addiction services, and ensure that urban redevelopment does not displace vulnerable populations. Only through comprehensive, scalable, and compassionate policies can the homelessness crisis be sustainably addressed.

Works Cited

- Amos. Personal interview. Conducted by Matthew Pollak, 28 Sep 2024.
- Anonymous A. Personal interview. Conducted by Matthew Pollak, 28 Sep 2024.
- Anonymous B. Personal interview. Conducted by Matthew Pollak, 12 Aug 2023.
- Augustine, Dallas, and Margot Kushel. "Community Supervision, Housing Insecurity, & Homelessness." *The Annals of the American Academy of Political and Social Science*, vol. 701, no. 1, 2022, pp. 152-171. PubMed.
<https://doi.org/10.1177/00027162221113983>.
- Baker, Tom, and Josh Evans. "'Housing First' and the Changing Terrains of Homeless Governance." *Urban Studies*, vol. 53, no. 6, 2016, pp. 1181–1198.
- Bretherton, Joanne, and Nicholas Pleace, editors. *The Routledge Handbook of Homelessness*. Taylor & Francis, 2023.
- Brimm, Marvin. Personal interview. Conducted by Matthew Pollak. 12 Aug 2023.
- Coalition for the Homeless. "Basic Facts About Homelessness: New York City." Coalition for the Homeless,
www.coalitionforthehomeless.org/basic-facts-about-homelessness-new-york-city/.
[Accessed 9/24/2024].
- Darryl. Personal interview. Conducted by Matthew Pollak, 13 Jan 2024.
- Daynes, Leigh. "Pottinger v. City of Miami and the Criminalization of Homelessness." *Miami Law Review*, vol. 70, no. 4, 1992, pp. 455–460.
- Gonzalez, Juan. "The Current Migrant Crisis: How U.S. Policy Toward Latin America Has Fueled Historic Numbers of Asylum Seekers." *New Labor Forum*, 1 May 2024,
<https://newlaborforum.cuny.edu/2024/05/01/the-current-migrant-crisis-how-u-s-policy-toward-latin-america-has-fueled-historic-numbers-of-asylum-seekers/>. Accessed [9/15/2024]
- Goldfischer, Eric. *From Encampments to Hotspots: The Changing Policing of Homelessness*. Routledge, 2020.
- Goodman, Laurie, et al. "Homelessness Prevention in New York City: On Average, Is It Cost Effective?" Urban Institute, 2016.
- Grainger, Clint. "Punishment, Support, or Discipline? Taking Stock of Recent Debates about Homeless Governance in Neoliberal Cities." *Housing Studies*, vol. 36, no. 8, 2021, pp. 1127-1143.
- Hauber, Jonathan. "Criminalization of the Unhoused: A Case Study of Alternatives to a Punitive System." *Journal of Urban Affairs*, vol. 45, 2023, pp. 20–35.
- Hodges, Kody. "No Shelter Here: Florida's New Anti-Vagrancy Law Misses the Point." Florida Policy Institute, 26 June 2023,
<https://www.floridapolicy.org/posts/no-shelter-here-floridas-new-anti-vagrancy-law-misses-the-point>. Accessed [9/15/2024].
- Homeless Trust. *Miami-Dade County Homeless Trust 2022 Annual Report*. Miami-Dade County, 2022.

- Kalesnikate, Vaiva, and Jean-Claude Garcia-Zamor. "The Sustainable Development of Miami: Tackling the Homeless Problem." *Current Urban Studies*, Scientific Research Publishing, 18 Sept. 2014, www.scirp.org/html/4-1150080_49825.htm.
- Karsen, Nicholas, and Joey Pellegrino. "Rising Homelessness in Collier County, Especially Among Kids, Seniors." *Wink News*, 3 May 2023, <https://winknews.com/2023/05/03/rising-homelessness-in-collier-county-especially-among-kids-seniors/>. Accessed [9/30/2024].
- Kirchheimer, David. *Sheltering Homeless: New York City's Struggle with Right to Shelter and Affordable Housing*. Columbia University Press, 1989.
- Larry. Personal interview. Conducted by Matthew Pollak. 12 Aug 2023.
- Ly, Angela, and Eric Latimer. "Housing First Impact on Costs and Associated Cost Offsets: A Review of the Literature." *Psychiatric Services*, vol. 66, no. 7, 2015, pp. 62–73.
- Markee, Patrick. "Turning the Tide: New York City's Response to Homelessness." *New York University Furman Center for Real Estate and Urban Policy*, 2017.
- Maroño Porto, Alex. "Understanding New York City's Right to Shelter." *West Side Rag*, 17 Aug. 2023, <https://www.westsiderag.com/2023/08/17/understanding-new-york-citys-right-to-shelter-2>. Accessed [9/15/2024].
- Melvin. Personal interview. Conducted by Matthew Pollak. 28 Sep 2024.
- Mitchell, Don. *Homelessness, American Style*. University of Georgia Press, 2011.
- Otokiti, Ahmed. "Challenges Faced by Homeless Populations in New York City." *Urban Social Issues Review*, 2018.
- Oscar. Personal interview. Conducted by Matthew Pollak. 16 Feb 2024
- Quilgars, Deborah, and Nicholas Pleace. "Housing First and Social Integration: A Realistic Aim?" *Social Policy & Administration*, vol. 50, no. 1, 2016, pp. 25–42.
- Randy. Personal interview. Conducted by Matthew Pollak. 12 Aug 2023.
- Robinson, Tony. "No Right to Rest: Police Enforcement Patterns and Quality of Life Consequences of the Criminalization of Homelessness." *Urban Affairs Review*, vol. 55, no. 1, 2019, pp. 83–112.
- Romi, Edgardo. Personal interview. Conducted by Matthew Pollak, 28 Sep 2024.
- Speer, Joanna. "Urban Makeovers, Homeless Encampments, and the Aesthetics of Displacement." *City & Community*, vol. 18, no. 1, 2019, pp. 44–65.
- Tsemberis, Sam, et al. "Housing First, Consumer Choice, and Harm Reduction for Homeless Individuals with a Dual Diagnosis." *American Journal of Public Health*, vol. 94, no. 4, 2004, pp. 651–656.
- Wiling, Jennifer, et al. "Understanding the Implications of a Punitive Approach to Homelessness: A Local Case Study." *Social Science Research Network*, 2018, pp. 1–30.

Factors of Type 2 Diabetes in the 21st Century (United States and India)

By Aleena Bakerywala

Abstract

Type 2 diabetes is one of the most prevalent conditions worldwide and has major health implications for those affected. The multifactorial nature of this condition makes its prevalence widespread and worthwhile to investigate between nations. The United States and India are two major nations, with a combined population of 1.7 billion people. Among these people, a significant portion of them have type 2 diabetes. Although the disease is the same and the symptoms are consistent in both countries, its causes and effects on people differ significantly. Determinants such as obesity prevalence, age, socioeconomic status, genetic predisposition, lifestyle, and healthcare resources of each country were investigated in this paper. In India, the age of type 2 diabetes diagnosis is lower compared to the United States (Mohan et.al. 2007). Differences in healthcare between the US and India are apparent. In the United States, lower socioeconomic status increases the risk of getting the disease (Agardh et.al. 2011). One's genetic predisposition also affects their risk of developing type 2 diabetes (Leong et.al. 2016), (Mohan et.al. 2021), (Radha et.al. 2007). The prevalence of obesity differs in both countries, and this has a significant effect on the people who are getting the disease. The lifestyle of Americans and Indians also differs which plays into people's experiences with type 2 diabetes. In both countries, we can see one main similarity, and that is that the increased use of machinery and technology has given people less of a chance to be active, which leads to obesity that may eventually lead to type 2 diabetes. This paper investigates the diverse factors contributing to type 2 diabetes within the United States and India.

Introduction

Type 2 diabetes is a multi-factorial condition that affects hundreds of millions of people worldwide. Although the disease pathophysiology is identical worldwide, the manifestation may differ due to cultural, environmental, and socioeconomic differences. There are two major forms of diabetes: type 1 and type 2. Type 2 diabetes is when your pancreas doesn't produce enough insulin and/or when cells in the body are desensitized to its effects. Insulin is needed to control your body's blood sugars, and when your body doesn't make enough (type 2) or at all (type 1), that leads to diabetes. 90-95% of the people with diabetes have type 2. Symptoms of type 2 diabetes are often not recognized immediately, so it's essential to get tested if you are at risk. If a person is at risk of getting type 2 diabetes, it is highly advised that they make healthy changes to prevent further complications. Changes can include losing weight, adjusting diet, staying active, and if necessary, taking medications to manage type 2 diabetes. Current treatment options for the management of type 2 diabetes include drugs like insulin and metformin. If a person does have type 2 diabetes, they have a choice of taking the insulin by a shot or pills. A person with diabetes should also always check their blood sugar levels and ensure they are at the level they should be. If a person isn't handling the disease correctly, they can be at risk for more medical issues,

including kidney damage, eye damage, heart disease, or strokes. If a high amount of glucose continues circulating in the bloodstream for a long time, it can also lead to disorders in the circulatory, immune, and nervous systems. Type 2 diabetics must be mindful of their glucose intake. Health factors that are known to increase risk of type 2 diabetes are obesity; people who are 45 years or older; having a direct family member, such as a parent or sibling, who has type 2 diabetes; not being physically active; and eating a high carbohydrate diet (“Type 1 vs Type 2 Diabetes | UVA Health”).

There is a major rise in both the United States and India in the number of people affected by type 2 diabetes. According to the IDF Diabetes Atlas, by 2045, it is estimated that 36,289.9 (in thousands) Americans will have type 2 diabetes. Similarly, by 2045, 124,874.7 (in thousands) will have type 2 diabetes in India. The United States and India both have high cases of type 2 diabetes, and the way the people in both countries deal with their health is very different (IDF Diabetes Atlas). In this review, we will examine factors of type 2 diabetes and how they differ between the United States and India including lifestyle, obesity prevalence, age, socioeconomic status, genetic predisposition, and healthcare resources.

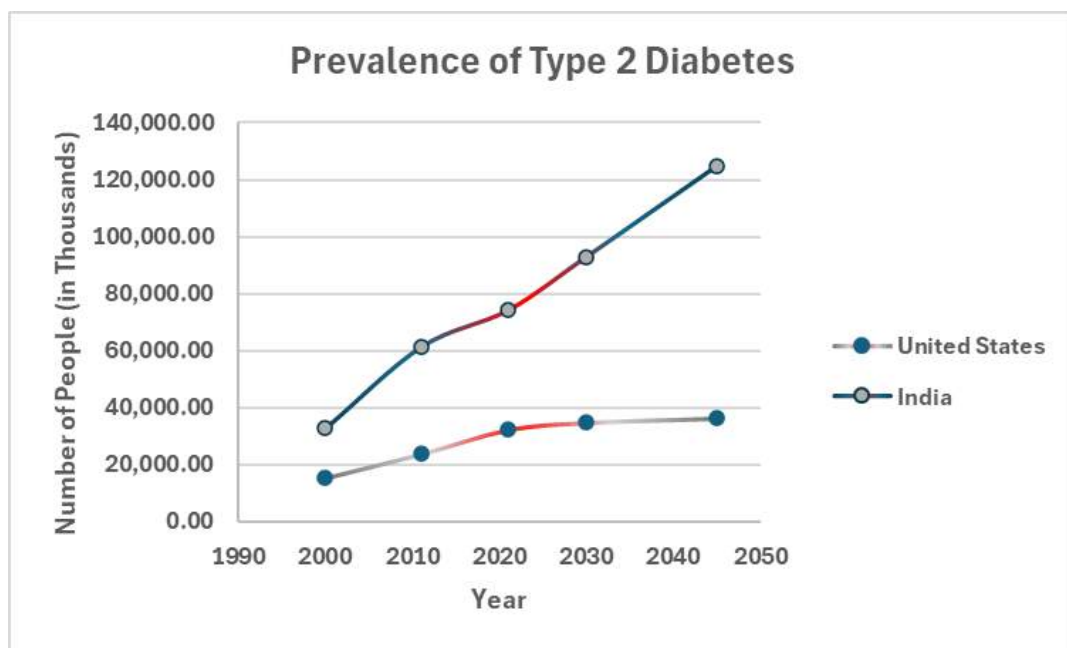


Figure 1: Prevalence of type 2 diabetes in the United States and India from 2000 to 2045. Reported in thousands. Data from IDF Diabetes Atlas

Obesity

Obesity is a leading contributor to type 2 diabetes. In the United States, notably, abdominal obesity is recognized by healthcare professionals as a leading cause of type 2 diabetes. The prevalence of type 2 diabetes, as well as abdominal obesity, significantly increased from 1999 to 2000. From 2013 to 2014, the prevalence of abdominal obesity increased by 8.8%. and commonness of abdominal obesity continues to rise for all age groups, which can potentially

lead to more Americans, children, teenagers, and adults becoming prediabetic or diabetic. Starting in 1999, every 2 years, the National Health Association (NHA) has examined the U.S. nationally by sampling 10,000 people, including residents, civilians, and non-institutionalized populations. The survey included an interview in each person's home and a health examination. The health examination consisted of taking whole blood specimens from people who were over the age of twelve. The survey concluded that obesity is a contributing factor to type 2 diabetes in the United States.

Obesity is also a contributing factor to people getting type 2 diabetes in India. According to the India state-level disease burden report, the root cause for a tenth of the people with diabetes in the country was an unhealthy diet and obesity. To prevent obesity, which can eventually lead to diabetes, people need to start eating healthier diets. More specifically, people should reduce the amount of refined grains, fats, and sugars they eat to help care for their health and prevent diabetes. Doing this can help prevent obesity, which can eventually lead to diabetes (Mohan et.al. 2021).

Overall, to limit obesity and avoid diabetes, people should pay attention to their waist-hip ratio. This is very important, and if people make sure the ratio stays the same, it should protect them from any health conditions that can eventually lead to obesity. In both countries, obesity—specifically from an unhealthy diet—is a substantial contributing factor to many cases of type 2 diabetes.

Age

Age is another major factor for diabetes. In the United States, the mean age for type 2 diabetes diagnosis decreased from 52 to 46 years old from 1988-1994 to 1999-2000, and more children are diagnosed each year (Koopman et.al 2005). Overall, the age for people being diagnosed in the U.S. with type 2 diabetes is decreasing each year, and the number of cases has been going up for the past few decades. This can be changed and managed by increased awareness to the public on what type 2 diabetes is and how to deal with and prevent it. Doctor recognition of symptoms can also help manage prediabetes so that not as many people are being diagnosed.

In India, the mean age for type 2 diabetes diagnosis is between 40 to 64. A report from North India stated that 8% of children under the age of 18 with diabetes have type 2 diabetes (Mohan et.al. 2007). The percentage of children who are getting type 2 diabetes in India is increasing. There has been an increase from 1999 to 2006 in children who are getting the disease. The reason the age of diagnosis is decreasing in India is because of increasing genetic predisposition to insulin resistance.

The decrease in the age of people contracting type 2 diabetes is a primary medical concern for both countries.

Socioeconomic Status

Another factor of type 2 diabetes is a person's socioeconomic status. Socioeconomic position is a significant factor associated with type 2 diabetes. The disease is more prevalent in high-income countries, possibly because people who are part of the lower SEP (Socio-Economic Position) group in these countries tend to be more obese from eating unhealthy foods, which increases rates of diabetes. A review was conducted to see the connection between type 2 diabetes and SEP. This was done by measuring the educational level, occupation, and income of the person and family and the payment of the overall country. Two authors did this study independently, finding different groups of people from various backgrounds and families. Both authors had the same controls, which were population and hospital-based. Once they had finished their data analysis, they concluded that the risk of getting type 2 diabetes is associated with low SEP in high, middle, and low-income countries. The impact of socioeconomic status on people getting type 2 diabetes in India is also severe.

The World Health Organization, or the WHO, an organization within the United Nations that is responsible for the overall health of the world, predicts that by 2025, 57.2 million people in India will have type 2 diabetes (Agardh et.al. 2011). As time goes on, diabetes is becoming a more prevalent disease in India. Most people living in India and getting the disease live in urban areas; in 1970, diabetes was at 2.1%, and in 1996, the cases increased to 11.6% (Rayappa et.al. 1999). The chances for diabetes are also rising in the rural parts of India. A study was done to see how people's socioeconomic status affects people in India getting diabetes. The study asked them about their general background, diagnosis and complications, current treatment and monitoring, cost of care for their diabetes, and quality of life. The results revealed the level of education had a significant effect on diabetes. People who had less of an education were more at risk of having type 2 diabetes than people who received more education.

Genetic Predisposition

Another main contributor to type 2 diabetes in both the United States and India is genetic predisposition. A person's genetic profile may play an essential role in developing type 2 diabetes. Type 2 diabetes can be in a person's genes and that can lead to them obtaining the disease. The genetic factors are fixed at birth and can become prevalent depending on personal and environmental factors. According to the WHO, by 2030, about 80 million people will have diabetes. Genetic profiles can help determine if they are more susceptible to contracting a disease because of a specific gene or if a parent can inherit it.

Indians are more resistant to insulin compared to Europeans and whites and are more susceptible to getting type 2 diabetes. The resistance to insulin from Indian people is commonly called "The Asian Indian or South Asian Phenotype". Indians are more insulin-resistant because of their higher rates of central adiposity, measured by their waist circumference and waist-to-hip ratio. The rate of low birth weight also contributes to this because children's growth accelerates to compensate for their age, which may be associated with insulin resistance in Indian children. The hip-waist ratio is a ratio that is suggested to maintain health and prevent obesity, which leads to other diseases. An increased hip-waist ratio can lead to diabetes and other diseases. (Radha

et.al. 2007). In contrast to India, the genetic predisposition for type 2 diabetes in the U.S. is related to racial minority groups. There was a study done to look at specific parts of the DNA of a group of test subjects and see what factors within their DNA contribute most to the likelihood of getting type 2 diabetes. Researchers looked at DNA variations and saw a sequence called single nucleotide polymorphisms (SNPs), associated with people having a higher risk of getting type 2 diabetes. The research shows that people who have SNPs in any of 38 type 2 diabetes related genes have a higher rate of mortality due to type 2 diabetes. The researchers also calculated the genetic risk score of all the individuals in the study by using the combination of risk alleles they called at each SNP. The risk alleles are the versions of the genes that are associated with having a higher risk of type 2 diabetes. (Leong et.al. 2016). Overall, these researchers analyzed the genetic factors and tried to get more insight into the genetic makeup that leads to or makes a person have a higher chance of developing the disease.

Healthcare Resources

Another main contributor to type 2 diabetes is healthcare resources. The health care resources differ in the United States and India. The United States has a healthcare system that is often very expensive and causes problems for people who can't afford it or struggle to pay. Many people have type 2 diabetes and other diseases and find it difficult to settle. In the United States, patients are often put on diets to help with their diabetes. The doctors give resources and medications to the patients and suggest self-care and self-management of the disease. There are many tools provided to assist with type 2 diabetes. (Whittboodt et.al. 2020). India has a less developed healthcare system, and because of this, more people remain undiagnosed with diabetes. Since people aren't getting diagnosed and aren't on the proper treatment, this leads to other medical complications such as damage to the eyes, kidneys, feet, and nerves. According to the CDC, 7.2 million Americans are undiagnosed with diabetes.

While India is also dealing with the problem of healthcare resources, it is on a much higher scale; many people are undiagnosed with type 2 diabetes; according to an IDF report, about 36.0 million in India are undiagnosed with diabetes. Not all parts of the county have a fully developed healthcare system, and this leads to more significant problems. This is a major problem for the country since more people are getting sick and having other medical issues just because they weren't diagnosed with diabetes. The access to healthcare resources and where a person is living has a significant effect on whether or not they will have type 2 diabetes. According to BMC Public Health, 77 million people in India have type 2 diabetes, not all of these people have access to good healthcare that would be able to take care of the disease they are battling. (Pradeepa, Mohan 2021). Based on research, the prevalence of diabetes in urban areas in India ranges between 12 to 16 percent. From that, we can tell that the environment and their healthcare access affected them medically. To prevent this from continuing rapidly, India needs to offer more healthcare resources so that the people and the country remain healthy and strong.

While healthcare is a problem in both countries, the reason for it is different. In the U.S., the issue is that individuals pay for the care, but it can be financed, which can help with their proper care and treatment. In India, certain parts of the country have fewer opportunities because of the underdeveloped healthcare system.

Lifestyle

Another main contributor to type 2 diabetes in many countries is lifestyle. Lifestyle is a person's behavior, which can contribute to the habits or illnesses a person has. In India, more people are getting type 2 diabetes in urban and rural areas because of the change in people's diets and people getting less physical activity. Now, more mechanical devices are helping people with their work, making them less active. The same applies in the United States. In both countries, there have been considerable advances in technology and machinery, making people less active, eventually leading to diabetes.

Going back to the lifestyle in India, the country has a huge pool of prediabetic people, and without treatment and a change to their diet and lifestyle, they will eventually develop diabetes. To prevent them from getting into the diabetic stage, they should eat healthier and be more active. Indians also have more exposure to environmental pollutants, which leads to other illnesses and, in some cases, this leads to diabetes. The other illnesses they get can cause them to become inactive, which leads to diabetes. To prevent this, the lifestyle of Indians needs to be changed (Mohan et.al. 2021), (Pradeepa, Mohan 2017). The diet of people living in the United States is believed to contribute to the prevalence of diabetes heavily. Daily consumption of sweets can lead to obesity, which can eventually lead to diabetes. More people in the United States spend much of their time in front of a screen. The increased time on the screen doubles the chances of getting diabetes because there is no physical activity that is going on (Kolb, Martin 2017).

Conclusion

Type 2 diabetes is a worldwide prevalent condition that affects many individuals. There are many factors, and they vary depending on the country a person is from. Determinants such as obesity prevalence, age, socioeconomic status, genetic predisposition, lifestyle, and healthcare resources are all factors that affect the people dealing with the disease. This paper inspected the differences in all of these factors within the United States and India and determined the similarities and differences type 2 diabetics face. In both the United States and India, obesity is one of the major contributors to type 2 diabetes. Although there were differences in the average age in both countries, the overall age of people getting the disease was very similar. In terms of socioeconomic status, the results were similar. In the United States, it was observed that people who have type 2 diabetes come from less-developed places. In India, it was observed that people who aren't as educated have a higher chance of having type 2 diabetes. The healthcare systems cause the two countries to have different results; while healthcare is expensive in the U.S., people who are receiving it say the process of taking care of diabetes is more self-care, whereas,

in India, the healthcare system isn't as developed and as a result, many people remain undiagnosed, which leads to more medical problems. Regardless, in both countries, people are undiagnosed because healthcare is expensive or they need access. Lifestyle factors are similar in both countries; people are becoming less active, which leads to obesity and other medical issues, which eventually leads to type 2 diabetes. In both the United States and India, many factors contribute to type 2 diabetes, and the factors sometimes are the same, and they sometimes differ.

Works Cited

- Agardh, Emilie, et al. "Type 2 Diabetes Incidence and Socio-Economic Position: A Systematic Review and Meta-Analysis." *International Journal of Epidemiology*, vol. 40, no. 3, 1 June 2011, pp. 804–818, pubmed.ncbi.nlm.nih.gov/21335614/, <https://doi.org/10.1093/ije/dyr029>.
- Caspard, Herve, et al. "Recent Trends in the Prevalence of Type 2 Diabetes and the Association with Abdominal Obesity Lead to Growing Health Disparities in the USA: An Analysis of the NHANES Surveys from 1999 to 2014." *Diabetes, Obesity and Metabolism*, vol. 20, no. 3, 1 Dec. 2017, pp. 667–671, <https://doi.org/10.1111/dom.13143>.
- International Diabetes Federation. "Diabetes around the World in 2021." *IDF Diabetes Atlas*, 2022, diabetesatlas.org/.
- Kolb, Hubert, and Stephan Martin. "Environmental/Lifestyle Factors in the Pathogenesis and Prevention of Type 2 Diabetes." *BMC Medicine*, vol. 15, no. 1, 19 July 2017, bmcmedicine.biomedcentral.com/articles/10.1186/s12916-017-0901-x, <https://doi.org/10.1186/s12916-017-0901-x>.
- Koopman, Richelle J. "Changes in Age at Diagnosis of Type 2 Diabetes Mellitus in the United States, 1988 to 2000." *The Annals of Family Medicine*, vol. 3, no. 1, 1 Jan. 2005, pp. 60–63, <https://doi.org/10.1370/afm.214>.
- Leong, Aaron, et al. "Type 2 Diabetes Genetic Predisposition, Obesity, and All-Cause Mortality Risk in the U.S.: A Multiethnic Analysis." *Diabetes Care*, vol. 39, no. 4, 16 Feb. 2016, pp. 539–546, care.diabetesjournals.org/content/39/4/539, <https://doi.org/10.2337/dc15-2080>.
- Mohan, Viswanathan, et al. "Type 2 Diabetes in Asian Indian Youth." *Pediatric Diabetes*, vol. 8, no. s9, Dec. 2007, pp. 28–34, <https://doi.org/10.1111/j.1399-5448.2007.00328.x>. Accessed 3 Aug. 2020.
- Mohan, Viswanathan, and Rajendra Pradeepa. "Epidemiology of Type 2 Diabetes in India." *Indian Journal of Ophthalmology*, vol. 69, no. 11, 29 Oct. 2021, p. 2932, www.ncbi.nlm.nih.gov/pmc/articles/PMC8725109/, https://doi.org/10.4103/ijo.ijo_1627_21.
- PH Rayappa, et al. *THE IMPACT of SOCIO-ECONOMIC FACTORS on DIABETES CARE*. Vol. 19, 1 Jan. 1999. Accessed 16 Oct. 2024.
- Pradeepa, R, and V Mohan. "Prevalence of Type 2 Diabetes and Its Complications in India and Economic Costs to the Nation." *European Journal of Clinical Nutrition*, vol. 71, no. 7, 19 Apr. 2017, pp. 816–824, <https://doi.org/10.1038/ejcn.2017.40>.
- Radha, V., and V. Mohan. "Genetic Predisposition to Type 2 Diabetes among Asian Indians." *Indian Journal of Medical Research*, vol. 125, no. 3, 1 Mar. 2007, p. 259, journals.lww.com/ijmr/Abstract/2007/25030/Genetic_predisposition_to_type_2_diabetes_among.7.aspx.
- Saslow, Laura R, et al. "An Online Intervention Comparing a Very Low-Carbohydrate Ketogenic Diet and Lifestyle Recommendations versus a Plate Method Diet in Overweight

- Individuals with Type 2 Diabetes: A Randomized Controlled Trial.” *Journal of Medical Internet Research*, vol. 19, no. 2, 13 Feb. 2017, p. e36, <https://doi.org/10.2196/jmir.5806>.
- Segal, Jodi, et al. “Comprehensive Diabetes Assessment Instrument for Patients with Type 2 Diabetes.” *Journal of Primary Care & Community Health*, vol. 14, 19 Nov. 2023, p. 21501319231204590, www.ncbi.nlm.nih.gov/pmc/articles/PMC10658767/, <https://doi.org/10.1177/21501319231204590>.
- “Type 1 vs Type 2 Diabetes | UVA Health.” *Uvahealth.com*, 2023, uvahealth.com/services/diabetes-care/types.
- Wittbrodt, Eric, et al. “Assessment of the High Risk and Unmet Need in Patients with CAD and Type 2 Diabetes (ATHENA): US Healthcare Resource Utilization, Cost and Burden of Illness in the Diabetes Collaborative Registry.” *Endocrinology Diabetes & Metabolism*, vol. 3, no. 3, 7 May 2020, <https://doi.org/10.1002/edm2.133>. Accessed 9 Aug. 2024.

Effects of Typically Developing Siblings on the Social and Emotional Development of Adolescents with Autism Spectrum Disorder By Navita Prabhakar

Abstract

The effect typically developing (TD) siblings have on adolescents with autism spectrum disorder (ASD) is rarely studied. Current published research mainly highlights the negative impact siblings with ASD have on their TD siblings. The current study's goal was to explore the beneficial effects that TD siblings may have on their siblings with ASD. Using a survey administered to seven parents of adolescents with ASD who have TD siblings, this study investigated how sibling interactions may contribute to the social and emotional development in adolescents with ASD. I ran correlations between several variables measured in the survey: (1) the quality of the relationship between the siblings, (2) the amount of time spent together, (3) the frequency of conflicts, and (4) how the interactions have affected the child with ASD. I found significant correlations in the relationship between the siblings and how the interactions have affected the child with ASD. This suggests that sibling interactions, regardless of their quality or the amount of time spent together, positively influenced the social and emotional development of the child with ASD. Simply, having a TD sibling may contribute positively to the development of the child with ASD. These findings highlight the need for more research on the positive role of TD siblings in the lives of adolescents with ASD.

Keywords Autism Spectrum Disorder, Siblings, Social Development, Emotional Development

Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by a combination of social communication challenges and restricted and repetitive behaviors that usually start earlier in life (Lord et al.; American Psychiatric Association). According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), ASD is diagnosed based on criteria that include repetitive patterns of behavior, interests, or activities, as well as ongoing deficiencies in social communication and interaction (American Psychiatric Association). These deficiencies include trouble developing, maintaining, and understanding relationships, difficulty maintaining a back-and-forth conversation, and abnormalities in eye contact and body language (American Psychiatric Association). The diagnostic criteria also require the presence of restricted and repetitive behaviors, which can include repetitive motor movements, insisting on routines, and fixated interests (American Psychiatric Association). People with ASD may also show additional challenges (echolalia (meaningless repetition of words), extreme sensitivity, hand flapping, pacing, and resistance to change), which can make it difficult for them to communicate and form meaningful relationships with others, including family members and educators (Mazef-sky, Williams, & Minshew; Robledo & Donnellan). Based on the estimates from the CDC's Autism and Developmental Disabilities Monitoring Network, approximately 1 in 36

children are said to have ASD (Maenner). Autism spectrum disorder is reported to occur across all racial, ethnic, and socioeconomic groups (Maenner).

Also, Maenner found that ASD is close to four times more common among boys than girls. However, estimates of the prevalence of ASD in various populations and settings differ based on the methods of sampling used. Frequently, TD siblings play an essential role in each other's cognitive development. This is due to their close relationships and shared experiences (Knott et al.). TD sibling interactions also allow for better negotiation and conflict-resolution skills, which are essential for cognitive development (Azmitia & Hesser). Existing research on the influence of TD siblings on adolescents with ASD is mainly focused on the negative aspects. According to several studies, TD siblings of children with ASD had more behavioral or emotional problems than siblings of TD children or children with other developmental disorders (e.g., Meyer, Ingersoll, Hambrick; Ferraioli; Jones et al.). One study stated that impairments in social communication, as well as other behaviors that characterize ASD, may impact family members and family dynamics, including TD siblings and sibling relationships (Meadan et al.).

Nevertheless, these results tend to overshadow the positive impacts TD siblings have on their siblings with ASD. The objective of this study is to investigate the effects that TD siblings can have on the social and emotional development of adolescents with ASD. While existing research papers primarily focus on the negative impact that adolescents with ASD have on their TD sibling, this study aims to figure out if there are any benefits that TD siblings may provide. I hypothesize that the better the relationship quality between siblings and the more they spend time together, the more positive the impact would be on the social and emotional growth of adolescents with ASD. I also hypothesize that the fewer conflicts the siblings have, the more positive the effect would be on the social and emotional growth of the adolescent with ASD.

Methods

Participants. To understand the effects TD siblings have on the social and emotional development of adolescents with Autism Spectrum Disorder (ASD), I conducted a survey with the feedback of parents. Parents were eligible to take the survey if they had one adolescent child with ASD and one TD adolescent child (Table 1). Both children had to be within the age range of 10-19 years. Parents must also be able to read and understand English.

Table 1. Survey questions

Sibling Relationship Questions	
How would you describe the relationship between your child with autism and their sibling? (Likert scale: 1 - very distant, 5 - very close)	
How frequently do your children spend time together? (Likert scale: 1 - rarely, 5 - daily)	
How often do conflicts occur between your child with autism and their sibling? (Likert scale: 1 - rarely, 5 - daily)	

Social and Emotional Growth Questions	
Interactions with their sibling have improved my child's social communication skills. (Likert scale: 1 - strongly disagree, 5 - strongly agree)	
Interactions with their sibling have helped my child express their emotions better. (Likert scale: 1 - strongly disagree, 5 - strongly agree)	
Interactions with their sibling have helped my child understand other people's feelings better. (Likert scale: 1 - strongly disagree, 5 - strongly agree)	
Interactions with their sibling have helped my child make friends. (Likert scale: 1 - strongly disagree, 5 - strongly agree)	
Interactions with their sibling have helped my child learn coping strategies for managing stress and anxiety from school, sports, ect. (Likert scale: 1 - strongly disagree, 5 - strongly agree)	
Interactions with their sibling have helped my child learn to share and take turns better. (Likert scale: 1 - strongly disagree, 5 - strongly agree)	
Interactions with their sibling has a positive impact on my child's mental health. (Likert scale: 1 - strongly disagree, 5 - strongly agree)	

Open-Ended Questions	
Does your child with autism have any other diagnoses? Please write all diagnoses. Leave blank if not.	
What activities do your children commonly do together? (optional)	
What do your children tend to have conflicts about? (optional)	
Please elaborate on how you think the interactions between your children have influenced your child with autism's social and emotional development. Feel free to comment specifically about any of the social and emotional growth areas you rated above. (optional)	
Do you have any additional observations about the relationship between your child with autism and their sibling? (optional)	

The survey was distributed through various WhatsApp communities dedicated to parenting and autism support. Data was collected from seven parents over two weeks in July of 2024. As stated in Table 2, six adolescents with ASD were male, and one was female. All adolescents with ASD were Asian, and their ethnicity was not Hispanic or Latino. The adolescents with ASD had varying severities of social communication difficulties, restricted interests, and repetitive behaviors.

Table 2. Demographics table for child with ASD

	Mean (SD)	Range
Age (years)	15.86 (4.10)	10-19
	<i>n</i>	Percentage
Biological sex		
Male	6	85.71
Female	1	14.29
Race		
White	0	0
Black or African American	0	0
American Indian or Alaska Native	0	0
Asian	7	100
Native Hawaiian or Other Pacific Islander	0	0
Multiracial	0	0
Ethnicity		
Hispanic or Latino	0	0
Not Hispanic or Latino	7	100
Social Communication Difficulties		
Mild	1	14.29
Moderate	5	71.43
Severe	1	14.29
None	0	0
Restricted Interests and Repetitive Behaviors		
Mild	1	14.29
Moderate	3	42.86
Severe	2	28.57
None	1	14.29

As illustrated in Table 3, two TD adolescents were male, and five were female. All TD adolescents were Asian, and their ethnicity was not Hispanic or Latino. All adolescents lived in the same household with their parents.

Table 3. Demographics table for child without ASD

	Mean (SD)	Range
Age (years)	15 (2.58)	11-19
	<i>n</i>	Percentage
Biological sex		
Male	2	28.57
Female	5	71.43
Race		
White	0	0
Black or African American	0	0
American Indian or Alaska Native	0	0
Asian	7	100
Native Hawaiian or Other Pacific Islander	0	0
Multiracial	0	0
Ethnicity		
Hispanic or Latino	0	0
Not Hispanic or Latino	7	100

Note. All participants live in the same house as their sibling. All participants have one sibling.

Procedures. The survey was sent to parents with at least one TD sibling and one sibling diagnosed with Autism Spectrum Disorder (ASD). Families were informed of the study's purpose and were provided with instructions on how to participate. Participants were allowed to leave the study at any time. No personal identifying information was collected, and confidentiality was maintained throughout the research. After participation, the participants were thanked and the results were collected. Other than the demographic questions, the survey asked parents three questions about their siblings' relationship and seven questions about the social and emotional growth of the sibling with ASD, using Likert scales. The Likert scale questions were rated from 1 to 5, where the points ranged from 1, representing "strongly disagree", "very distant", and "rarely", to 5, representing "strongly agree", "very close", and "daily". Correlations were run using the Likert scales. Further, five open-ended questions were asked.

Analyses. Pearson's correlations were performed between variables measured in the survey. The variables included all questions with likert scales: the quality of the relationship between the siblings (variable 1), the amount of time spent together (variable 2), the frequency of conflicts (variable 3), and how the interactions have affected the child with ASD (variables 4-10). Additionally, I qualitatively summarized the open-ended responses to understand better what the siblings tend to have conflicts over and what activities they do together.

Hypotheses. I hypothesized that the better the relationship quality between siblings (variable 1) and the more they spent time together (variable 2), the more positive the impact would be on the social and emotional growth of the adolescent with ASD (variables 4-10). I also hypothesized that the fewer conflicts the siblings had (variable 3), the more positive the impact would be on the social and emotional growth of the adolescent with ASD (variables 4-10).

Results

Quantitative findings. According to Table 4, the quality of the sibling relationship had a moderately positive correlation with the ability to make friends more easily ($r = 0.65$), learning coping strategies ($r = 0.65$), and understanding how to take turns better ($r = 0.69$). There was a moderately negative correlation between the frequency of conflicts and the overall relationship quality ($r = -0.66$).

The improvement in social communication skills for adolescents with ASD was strongly correlated with their ability to express emotions better ($r = 0.92$), understand the feelings of others better ($r = 0.92$), make friends more easily ($r = 0.83$), learn coping strategies ($r = 0.83$), understand how to take turns better ($r = 0.86$), and improvement of mental health ($r = 0.96$).

The ability to express emotions better was a perfect positive correlation with understanding the feelings of others better ($r = 1$). Expressing emotions better and understanding the feelings of others were strongly positively correlated with making friends more easily ($r = 0.87$), learning coping strategies ($r = 0.87$), understanding how to take turns better ($r = 0.79$), and improving mental health ($r = 0.89$).

Another perfect positive correlation was between making friends more easily and learning coping strategies ($r = 1$). Making friends more easily and learning coping strategies both

correlated positively with understanding how to take turns better and the improvement of mental health for the child with ASD ($r = 0.94$ and $r = 0.68$, respectively). Lastly, there was a strong correlation between taking turns better and improvement of mental health for the child with ASD ($r = 0.71$).

Very weak correlations were observed between a few variables ($r = -0.04$ to $r = 0.31$; Table 4). The amount of time adolescents spent together showed very weak correlations with all other variables. Very weak correlations were also observed between the frequency of conflicts among the adolescents and all the other variables. Additionally, the quality of the relationship between the adolescents had very weak correlations with social communication skills, ability to express emotions, understanding the feelings of others, and the mental health of the child with ASD.

Table 4. Correlations between variables

	Relationship b/w both adolescents	Time that both adolescents spend together	Frequency of conflicts between adolescents	Social communication skills have improved for the child w/ ASD	Child w/ ASD can express emotions better	Child with ASD can understand the feelings of others better	Child w/ASD can make friends more easily	Child w/ ASD has learned coping strategies	Child w/ ASD understands how to take turns better	The mental health for the child w/ ASD has improved
Relationship b/w both adolescents										
Time that both adolescents spend together	-0.06									
Frequency of conflicts b/w adolescents	-0.66	-0.37								
Social communication skills have improved for the child with ASD	0.25	0.02	0.11							
Child w/ ASD can express emotions better	0.31	0.22	0.10	0.92						
Child w/ ASD can understand the feelings of others better	0.31	0.22	0.10	0.92	1					
Child w/ ASD can make friends more easily	0.65	-0.04	-0.16	0.83	0.87	0.87				
Child w/ ASD has learned coping strategies	0.65	-0.04	-0.16	0.83	0.87	0.87	1			
Child w/ ASD understands how to take turns better	0.69	-0.09	-0.27	0.86	0.79	0.79	0.94	0.94		
The mental health for the child w/ ASD has improved	0.01	0.14	0.25	0.96	0.89	0.89	0.68	0.68	0.71	

Qualitative findings. One parent noted that their children engaged in crafting, drawing, and building blocks together. Two parents mentioned that their children enjoyed playing cards and board games together. One parent said their children liked going on walks together. Another parent stated that their children did pretty much everything together. One parent reported that their children do not participate in any activities together. One parent did not respond.

Parents noted that their children tended to have conflicts over entertainment, eating without sharing, and not keeping things clean. Of the seven responses for the children with ASD, one adolescent had OCD, one had cerebral palsy, and one had OCD and anxiety (i.e., did not specify what type of anxiety disorder).

When asked whether interactions between the adolescents have influenced the child with ASD's social and emotional development, one parent stated that "interactions between [the] siblings helped [the] ASD kid understand the emotions of others and value them." Another parent said that the child with ASD mimicked everything the child without ASD did, and the friends of the child without ASD were the friends of the child with ASD. The parent also noted that the child with ASD had someone to share their feelings with other than the parents. Two parents stated that they saw improvement in the social and communication aspects of their child with ASD. Lastly, one parent said there was not much interaction between the children. Thus, there was no influence.

Discussion

To summarize, many parents said that interactions with siblings, regardless of the quality of their interactions, had a positive impact on the social and emotional development of their child with ASD. Siblings without ASD often acted as role models, encouraging their siblings with ASD to engage in social activities and develop their skills. I hypothesized that the better the relationship quality between siblings, the more they spent time together, and the fewer conflicts they had, the more positive the impact would be on the social and emotional growth of the adolescent with ASD. Contrary to my hypothesis, the results showed very weak correlations between the amount of time adolescents spent together and all other variables. The results also showed very weak correlations between the frequency of conflicts among the adolescents and all the different variables.

Additionally, the quality of the relationship between the adolescents had very weak correlations with social communication skills, ability to express emotions, understanding the feelings of others, and the mental health of the child with ASD. Taken together, this demonstrates relationship quality, amount of time spent together, and frequency of conflicts did not have a significant impact on the social and emotional development of the adolescent with ASD, as I predicted. This means that the interactions between siblings are not as strongly connected to the social and emotional development of adolescents with ASD as I expected. This could be because of age gap differences, parental influences, the small sample size, or lack of diversity. Exploring some of these aspects in further studies may be critical to understanding these dynamic

relationships better. It is also possible that rather than the frequency of conflicts, what matters is the intensity of conflict.

Similarly, the quality of time spent together may be more important than the quantity. This would potentially have had a stronger relationship with social communication skills, expressing emotions, understanding the feelings of others, or the mental health of the adolescent with ASD. Future research should examine both the frequency and intensity of conflicts and see if the relationships with these variables are different. Furthermore, siblings without ASD could potentially benefit from intervention programs that aim to improve their ability to interact positively with their siblings with ASD.

Furthermore, parents also said that the child with ASD often mimicked what the child without ASD did. One parent said that the child with ASD learned a lot of social cues from their sibling. This finding illustrates that simply having a sibling, engaging in interactions, and observing what they do contributes positively to the social and emotional development of adolescents with ASD. However, several limitations should be considered in this study. As mentioned earlier, the sample size plays a significant role in the observed correlations, and hundreds of participants would be needed to validate these results. The survey is also biased in terms of race/ethnicity, surveying only Asians. Future research should consider including more diverse demographics that reflect the whole population better. Distribution of the survey through various channels, other than just WhatsApp, would be necessary for obtaining a more diverse sample. Future research should target channels like groups and organizations dedicated to autism support. Future research could also reach out to schools and healthcare providers to recruit participants. I was also not able to investigate how age and gender affect these relationships, as the majority of participants with ASD were male, and the sample size was small. This limitation is significant to address because the interactions between siblings could vary between males and females, especially given sex effects in ASD diagnosis (i.e., ASD occurs more often in males). Future research should explore these gender differences and age-related variations.

Furthermore, this brief survey was intended to take less than 5 minutes. A more detailed survey, in-person interviews, or even dedicated time to observing relationships between siblings first hand would provide better results in the future. To do this, funding will be needed to pay participants for their time. Additionally, I asked parents to think about the relationship between a child with ASD and one child without ASD. It is possible that relationships could differ between other siblings or that more siblings could have a more significant impact. Lastly, a few of the parents stated that their children had other neurological or neurodevelopmental disorders. These comorbidities are additional challenges that may make it more complicated for the adolescent with ASD to interact with their sibling, leading to more conflicts. This could impact the social and emotional development of the adolescent with ASD, influencing the overall results. Therefore, comorbidities should be accounted for in future work.

The findings of this study indicate that interactions between TD siblings and adolescents with ASD can have a positive impact on the social and emotional development of the adolescent with ASD. This aligns with existing research that emphasizes the importance of sibling

relationships in cognitive and emotional growth (Knott et al.; Azmitia & Hesser). While previous studies have focused on challenges faced by TD siblings due to their sibling with ASD, my findings reveal that sibling interactions also have positive contributions. This finding is important because it goes against the prevailing view that sibling interactions are primarily negative and highlights the need for further research into these interactions.

Works Cited

- American Psychiatric Association. "Diagnostic and Statistical Manual of Mental Disorders." *Diagnostic and Statistical Manual of Mental Disorders*, vol. 5, no. 5, 2013, [dsm.psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596](https://doi.org/10.1176/appi.books.9780890425596), <https://doi.org/10.1176/appi.books.9780890425596>.
- Azmitia, Margarita, and Joanne Hesser. "Why Siblings Are Important Agents of Cognitive Development: A Comparison of Siblings and Peers." *Child Development*, vol. 64, no. 2, Apr. 1993, p. 430, <https://doi.org/10.2307/1131260>.
- Ferraioli, Suzannah J., and Sandra L. Harris. "The Impact of Autism on Siblings." *Social Work in Mental Health*, vol. 8, no. 1, 11 Dec. 2009, pp. 41–53, <https://doi.org/10.1080/15332980902932409>.
- Jones, Emily A., et al. "When One Sibling Has Autism: Adjustment and Sibling Relationship." *Journal of Child and Family Studies*, vol. 28, no. 5, 26 Mar. 2019, pp. 1272–1282, <https://doi.org/10.1007/s10826-019-01374-z>.
- Knott, Fiona, et al. "Sibling Interaction of Children with Autism: Development over 12 Months." *Journal of Autism and Developmental Disorders*, vol. 37, no. 10, 21 Feb. 2007, pp. 1987–1995, [link.springer.com/article/10.1007/s10803-006-0347-z](https://doi.org/10.1007/s10803-006-0347-z), <https://doi.org/10.1007/s10803-006-0347-z>.
- Lord, Catherine, et al. "Autism Spectrum Disorder." *Nature Reviews Disease Primers*, vol. 6, no. 1, 16 Jan. 2020, pp. 1–23, [www.nature.com/articles/s41572-019-0138-4](https://doi.org/10.1038/s41572-019-0138-4), <https://doi.org/10.1038/s41572-019-0138-4>.
- Maenner, Matthew J. "Prevalence and Characteristics of Autism Spectrum Disorder among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2020." *MMWR. Surveillance Summaries*, vol. 72, no. 2, 24 Mar. 2023, pp. 1–14, [www.cdc.gov/mmwr/volumes/72/ss/ss7202a1.htm](https://doi.org/10.15585/mmwr.ss7202a1), <https://doi.org/10.15585/mmwr.ss7202a1>.
- Mazefsky, Carla A., et al. "Variability in Adaptive Behavior in Autism: Evidence for the Importance of Family History." *Journal of Abnormal Child Psychology*, vol. 36, no. 4, 10 Jan. 2008, pp. 591–599, <https://doi.org/10.1007/s10802-007-9202-8>.
- Meadan, Hedda, et al. "Review of Literature Related to the Social, Emotional, and Behavioral Adjustment of Siblings of Individuals with Autism Spectrum Disorder." *Journal of Developmental and Physical Disabilities*, vol. 22, no. 1, 3 Dec. 2009, pp. 83–100, <https://doi.org/10.1007/s10882-009-9171-7>.
- Meyer, Katherine A., et al. "Factors Influencing Adjustment in Siblings of Children with Autism Spectrum Disorders." *Research in Autism Spectrum Disorders*, vol. 5, no. 4, Oct. 2011, pp. 1413–1420, <https://doi.org/10.1016/j.rasd.2011.01.027>.
- Robledo, Jodi, and Anne M Donnellan. "Properties of Supportive Relationships from the Perspective of Academically Successful Individuals with Autism." *Intellectual and Developmental Disabilities*, 1 Aug. 2008, [https://doi.org/10.1352/1934-9556\(2008\)46\[299:posrft\]2.0.co;2](https://doi.org/10.1352/1934-9556(2008)46[299:posrft]2.0.co;2).

Smith, Laura O., and Jennifer H. Elder. "Siblings and Family Environments of Persons with Autism Spectrum Disorder: A Review of the Literature." *Journal of Child and Adolescent Psychiatric Nursing*, vol. 23, no. 3, 12 Aug. 2010, pp. 189–195, <https://doi.org/10.1111/j.1744-6171.2010.00240.x>.

Perceptions on Artificial Intelligence, Mental Health, Emotion Detection Technology

By Saanvi Movva

INTRODUCTION

The world continues to digitize with new advancements, research, and technological developments. Within this digitization process, artificial intelligence has taken the world by storm, creating machines that think and mimic the roles of humans. These innovations paved the way for the creation of emotion detection technology. As its name suggests, this technology uses sensors and machinery to analyze human emotions.

Existing research demonstrates that early detection of emotional symptoms significantly correlates with improved outcomes in managing depression and mental illness, with a high success rate observed in intervention strategies. Emotion detection technology has the potential to detect symptoms at an early stage, thereby contributing to improved outcomes in managing mental health conditions.

Mental illness encompasses conditions such as depression, anxiety, stress, and suicidal tendencies. American Psychiatric Association states, “1 in 5 U.S. adults experience some form of mental illness”(What Is Mental Illness?, n.d.), with the CDC’s 10-Year *Youth Risk Behavior Surveillance Analysis* stating that more than 4 in 10 adolescents experience poor emotions with 1/3 experiencing poor mental health. Beyond feeling happy, sad, or angry, emotions deeply influence human interactions and well-being. Many individuals fail to recognize that their emotions, stress levels, and anxiety may not be within the normal range, and there should be a means to help them regain balance through technology. This is where emotion detection technology could have an impact. This technology can be a mobile application that can identify and react to human emotions, providing customized assistance and offering capabilities such as real-time emotion analysis, sentiment tracking, and visualization of emotional data.

The intersection of health and technology has varying perspectives amongst age groups; therefore, for emotion detection technology to be effective, it is important to research ***How is the implementation of and the need for AI-powered mobile emotion detection software to improve mental health outcomes perceived by participants under and above 18 years of age?***

LITERATURE REVIEW

To understand the rationale behind this study’s analysis of the perception and necessity of emotion detection technology, it is essential to analyze the existing research regarding the development and usage of the technology and the psychology of individuals under and over 18.

Psychology of Adolescents Versus Adults

A study from NIH explains that during the process of an adolescent's brain development, an “increase in stress-related psychological dysfunctions, such as anxiety, depression, and drug abuse” is often observed (Romeo R. D. 2013). On the contrary, adults have encountered many more hardships and stressful situations, reducing their stress levels and making their brains more

adaptive. This is shown by an NIH study that tested this concept on animals, showing that “an adult animal repeatedly exposed to the same stressor (homotypic stress) displays a habituated hormonal response compared to an adult exposed to that stressor for the first time”(Romeo R. D. 2013). An adult experiencing the stressor for the first time is similar to an adolescent encountering a stressful moment for the first time. This difference in stress levels is apparent as a study by *APA* shows that teen stress surpasses adult stress and is far from the healthy standard. (“American Psychological Association Survey Shows Teen Stress Rivals That of Adults,” 2014).

The differences in stress between teens and adults might explain the difference in perceptions between teens and adults. This stress dynamic underscores the importance of detecting and preventing stress before it escalates into depression, anxiety, PTSD, and other deteriorations in life. Being able to detect stress and symptoms of mental health illness early on can have significant benefits to an individual's future.

Functionality and Features of Emotion Detection Technology

Standard features include facial recognition, personalized chatbots, speech analysis, and sentiment analysis tracking.

Speech Emotion Recognition

Speech emotion recognition involves analyzing acoustic features to infer emotional states from speech signals. An article on speech communication overviews the key components in speech emotion recognition systems (Schuller et al. 2011). It discusses acoustic features, such as prosodic, spectral, and voice quality, and how they are extracted and selected for emotion recognition tasks. The authors also present different machine-learning approaches, including hidden Markov models (HMMs) and support vector machines (SVMs), for classifying emotions based on these features.

Personalized Chatbots

Chatbots equipped with emotion detection analyze user input to respond with appropriate emotional sensitivity, mimicking human interaction. Research by the Rajagiri School of Engineering and Technology on Emotion Recognition-Based Mental Healthcare Chatbots highlights their potential to reduce stress and assist those with emotional vulnerabilities across all age groups. These chatbots use specific sentiment analysis and machine learning algorithms to efficiently detect sentiment and generate suitable responses. The study concludes that chatbots can engage in constructive conversations and continuously monitor users’ emotional and behavioral patterns (Antony et al., n.d.).

Facial Recognition

Facial Recognition is driven by “Action Units that describe movements of certain facial muscles and muscle groups to classify emotions” (Pao, n.d.). Examples of action units include raising the eyebrow, scrunching the nose, pouting lips, etc. James Pao, a researcher at Stanford

University, further explains the different approaches to facial recognition. One approach, Fisherface, involves storing data on typical emotional expressions like happiness, anger, and sadness and matching test images to this stored data. The system then uses mathematical techniques to extract key features from these emotions, creating a set of 'Fisherfaces.' Finally, the system compares new face images to the Fisherfaces to determine which emotion is expressed (Pao, n.d.).



This is an example of the Fisherfaces. (Pao, n.d.)

Sentiment analysis tracking

Natural Language Processing (NLP) enables the analysis of textual data for affective content, focusing on two main concepts: sentiment analysis and emotion detection. Sentiment analysis classifies subjective information in text into positive, negative, or neutral categories. For example, the start-up CompanionMX, emerging from MIT, developed an app that analyzes voice and phone metadata (including call logs, text logs, and geolocation) to assist clinicians in making informed decisions for mood disorder patients. Such applications allow early detection of disorders and conditions, significantly benefiting healthcare (CompanionMX | STEX, n.d.).

The articles above offer insights into the complexities of developing and utilizing various features supported by emotion detection technology. While essential for a foundational understanding of this technology, they lack information concerning individuals' perceptions before and after using it and the outcomes of implementing these features on individuals.

Perception of AI in healthcare and Benefits

The integration of AI into healthcare has brought about significant changes in the execution of procedures, diagnoses, medical data storage, and more. However, there is a common concern regarding its perception. While many view it as a groundbreaking advancement for society, others perceive it as unethical.

In a survey conducted by the Pew Research Center, 60% of American adults expressed discomfort with the idea of their doctor depending on AI for medical diagnosis and believed it would hinder doctor-to-patient relations (Akinrinmade et al., 2023). Furthermore, a study by Vo et al. (2023) systematically reviewed 7,490 different records. It concluded that there was an overall positive sentiment, but many seemed concerned about AI's ability to deliver empathetic care. On the contrary, Fritsch et al. 2022 conducted a study on 452 participants using a paper-based questionnaire, which resulted in 53.18% of the respondents having a positive perception of the use of artificial intelligence in medicine but strongly believing that a healthcare

provider should be present in the usage.

Advancement in this technology can benefit several fields, specifically mental health, by tailoring interventions to individuals' specific needs, storing an individual's emotional data, and detecting symptoms early on, which can “help reduce the severity of an illness and interruptions in quality of life and functions”(Warning Signs of Mental Illness, n.d.).

NIH article Khullar et al. (2022) highlights that AI overall has a positive societal perception. However, potential issues such as misdiagnosis, dependency on artificial intelligence, lack of research, increased costs, and privacy concerns could arise. Moreover, Akinrinmade et al. (2023) explain that common misconceptions about AI usage lie in online sensationalism, as many advertisements, social media, and news narrow in on myths and unrealistic expectations and perpetuate that AI is potentially dangerous.

It is noticed that the studies regarding perceptions did not have an analysis comparing adolescents with adults and rather gave a generalistic view of society as a whole.

GAP IN RESEARCH

Emotion recognition systems, such as mobile applications, offer immense potential to revolutionize the understanding of human psychology and mental healthcare through machine sensing of affective states. While emotion detection technology is still nascent, its potential impact on mental health could be significant. To maximize its benefits in safeguarding individuals from deteriorating mental health conditions, it is imperative to expand research in this field. There is a general lack of research on the perceptions surrounding the utilization of AI-based emotion detection technology for identifying and addressing mental health symptoms, and it is unclear what factors may change these perceptions, especially age. Diverse perceptions and needs persist when comparing individuals below and above 18 years as psychological issues, stressors, and comprehension evolve with age and maturity. Recent research has presented a comparison of psychological issues amongst both age groups. Still, the comparison of perceptions amongst both age groups regarding AI-based emotion detection technology is lacking.

To fully understand the implications of AI-based emotion detection apps on the mental health space, it is important to investigate age-based differences in how they are viewed, as these could be instrumental in designing effective and targeted features and functionalities for the application. This research aims to answer: ***How is the implementation of and the need for AI-powered mobile emotion detection software to improve mental health outcomes perceived by participants under and above 18?***

METHODOLOGY

Hypothesis

This study hypothesizes that a significantly higher percentage of adolescents would positively perceive the use of emotion detection technology in the mental health sector and indicate a stronger need for it than adults.

Method

A mixed-method approach was utilized to execute the purpose of this research. A cross-sectional descriptive questionnaire with a thematic analysis was used to characterize the participants' attitudes, perceptions, behaviors, and experiences regarding various aspects of emotions, mental health, and technology. Thematic analysis is a method for identifying and analyzing themes in data that can provide trustworthy analysis and be flexible for the needs of the research conducted, which is effective for this research as it groups the varying perceptions analyzed into groups (Nowell et al., 2017b). Qualitatively based and binary questions were needed to capture the nuances of the participant's thoughts on the issue and their mental health evaluations. While allowing respondents to express their feelings more specifically, it also allows responses to be sorted into thematic groups based on keywords and identifying information; such as age and response type. The major themes explored by the questionnaire are 1) Receptiveness to AI-based emotion detection and 2) Self-assessment of Mental Health. These themes provide an organizational structure to qualitative questions, making it easy to distinguish responses based on the technology's knowledge from responses based on the perceived need for this technology. These groupings can be used to perform a frequency analysis.

Frequency distributions, a quantitative tool, reveal significant differences in responses between adults and adolescents and corroborate these differences across themes and trends in both experimental groups.

Using the frequency distribution data gathered, a two-sample hypothesis test using the Z-test formula was conducted to determine the significance of the results. This test best suited this research compared to tests such as a t-test. For instance, the t-test analyzes the significant difference in the means of two groups (Wadhwa & Marappa-Ganeshan, 2023). However, it cannot compare the proportions of two independent groups, which is why a two-sample hypothesis test is most effective.

Utilizing this method allowed for trends to be analyzed in each theme and to determine whether a statistically significant correlation exists between age and perception of emotion detection technology. This methodology was utilized to separate and display age-related response patterns and provide frequency-based analysis of adults' and adolescents' qualitative views.

Process

The survey used random sampling to recruit 50 individuals, ranging from 11-17 for under 18 and 18-55 for over 18. To reduce the effect of any confounding variable, the sample was controlled for the socioeconomic status of the middle/upper-middle class, residential area of North Fulton County, Atlanta, GA, and education status (adolescents enrolled in school and adults educated with at least a bachelor's degree), which ensured that any differences observed

were due to age. Within these constraints in place, random sampling was utilized to select participants, reduce the presence of bias, and increase the probability that these results could be applied to the general population.

This sampling technique randomly selected individuals from the target population by evenly distributing posters/messages with online access to the questionnaire in common high-traffic areas, such as hallways, offices, online platforms, and student groups. Posting in such a way within the target population ensured a high chance that every member of the population had an equal chance of accessing the questionnaire and being included in the sample. When a sample is truly random and representative of the population, the findings from the sample have higher credibility than other procedures, such as snowballing, which is a chain sampling process where one respondent refers to other respondents, resulting in a higher chance of bias (Bhardwaj, 2019).

The survey is distributed electronically via Google Forms, so it is easily accessible. It comprises 13 simple multiple-choice questions. An online survey was used rather than an interview because it poses several benefits for this research to get a larger sample size such as the respondent can answer at their own convenient time, the respondent can take their time answering questions, and the respondents will not feel any pressure/judgment when answering the questions as it regards personal matter, ensuring more accurate responses (Regmi et al., 2017). The questions were based on two factors: the individual's perception of using a mobile application integrated with emotion detection technology and the individual's self-report on their mental health and perceived need for interventions. There were four varieties of questions (Appendix A):

- a. The Yes/No/Sometimes questions were used to analyze the population's presence/absence of the measured parameter.
- b. The scale, ranging from extremely important/beneficial to not important/beneficial, judges the priority assigned and the (positive or negative) perception of the assigned parameter.
- c. Next, the scale, ranging from extremely likely to not likely, measures the likelihood of the assigned parameter occurring in the respondent's life.
- d. Lastly, the multi-select questions were used to identify the most popular outcome out of a selection of possible outcomes.

The data acquired from this assessment will be quantitatively analyzed using frequency distributions and a two-sample hypothesis test. According to the Australian Bureau of Statistics (2024), frequency distributions visually organize and present frequency counts, making the data easier to interpret. These distributions, based on questionnaire responses, provide a holistic view of the study's data and clearly show how observations are distributed.

After acquiring and analyzing the frequency distribution, a two-sample hypothesis test was executed to determine whether there was any significance between the two age groups.

Within this test, the Z-test formula was used and then plugged into Excel using the NORMSDIST function to get a p-value. This p-value is compared to the standard 5% significance level, corresponding to a 95% confidence. If the p-value is less than 0.05, the null hypothesis is rejected and shows that there is a significant difference. Meanwhile, if it is greater than 0.05, the null hypothesis fails to be rejected and shows that there is no significant difference. Z-test formula for two-sample hypothesis test with proportions:

$$z = \frac{(p1 - p2)}{\sqrt{\frac{p1(1-p1)}{n1} + \frac{p2(1-p2)}{n2}}}$$

Ethical Practices

Individuals participating in this study were voluntary and were provided a detailed description of the study before participating. Additionally, a consent form was received from each participating individual. Parental consent was obtained for participants under the age of 18. Participants can withdraw their questionnaire and participation in this study anytime, and their responses will not be used. This research does not harm any living creature or require hazardous materials and was approved by an IRB. Lastly, the information received is anonymous, kept on a password-protected device, and is not shared with anyone other than being used as data in this study; therefore, the participants' privacy is secure.

RESULTS

This section displays the survey findings in graphical format. The survey involved 50 participants who responded to 13 questions.

Figure 1:

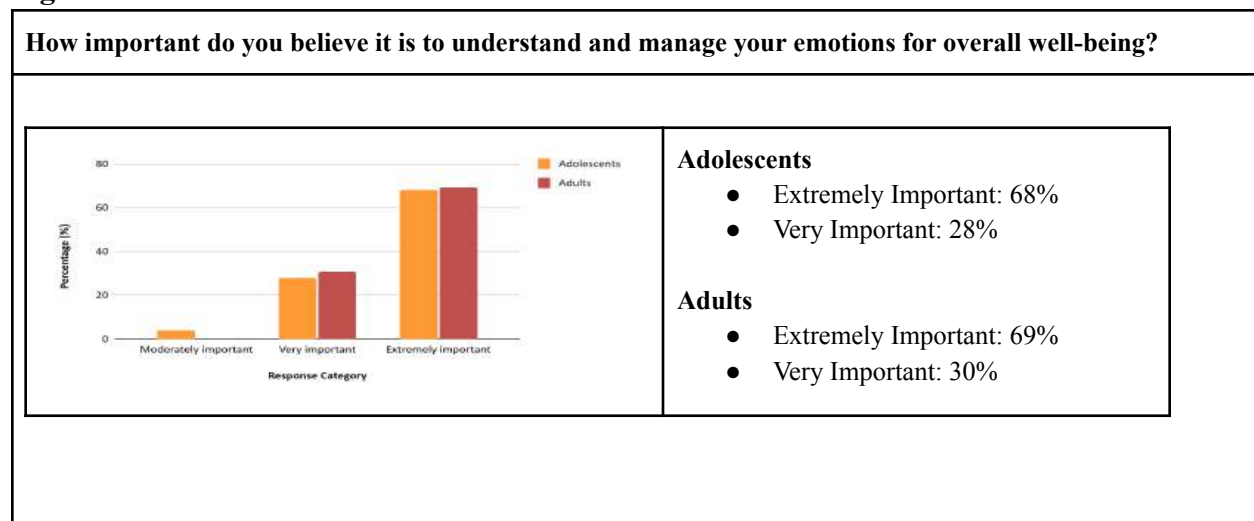


Figure 2:

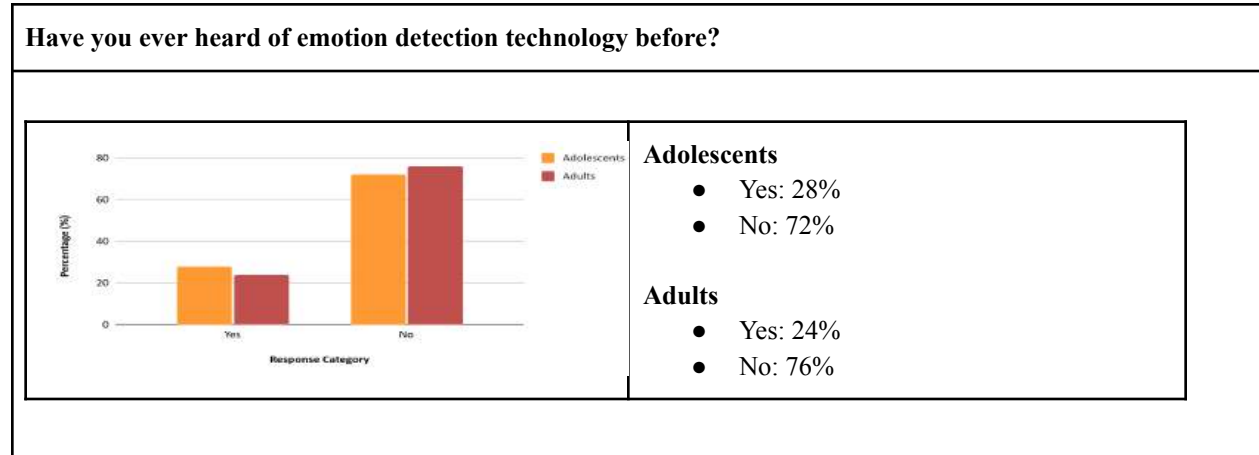


Figure 3:

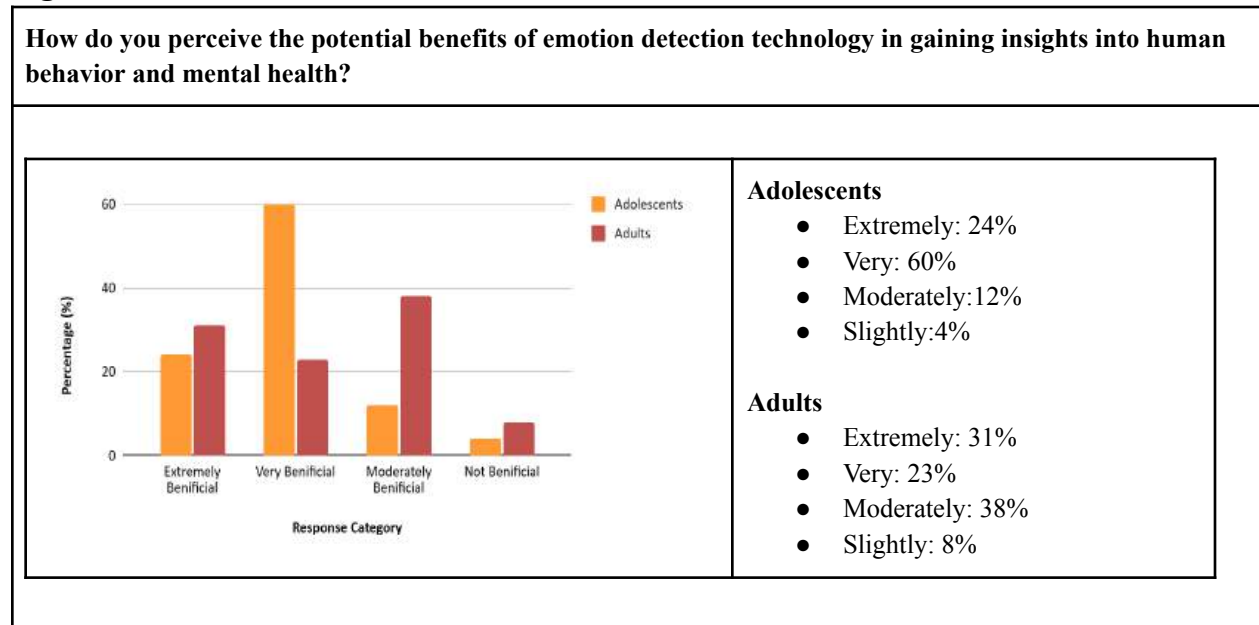


Figure 4:

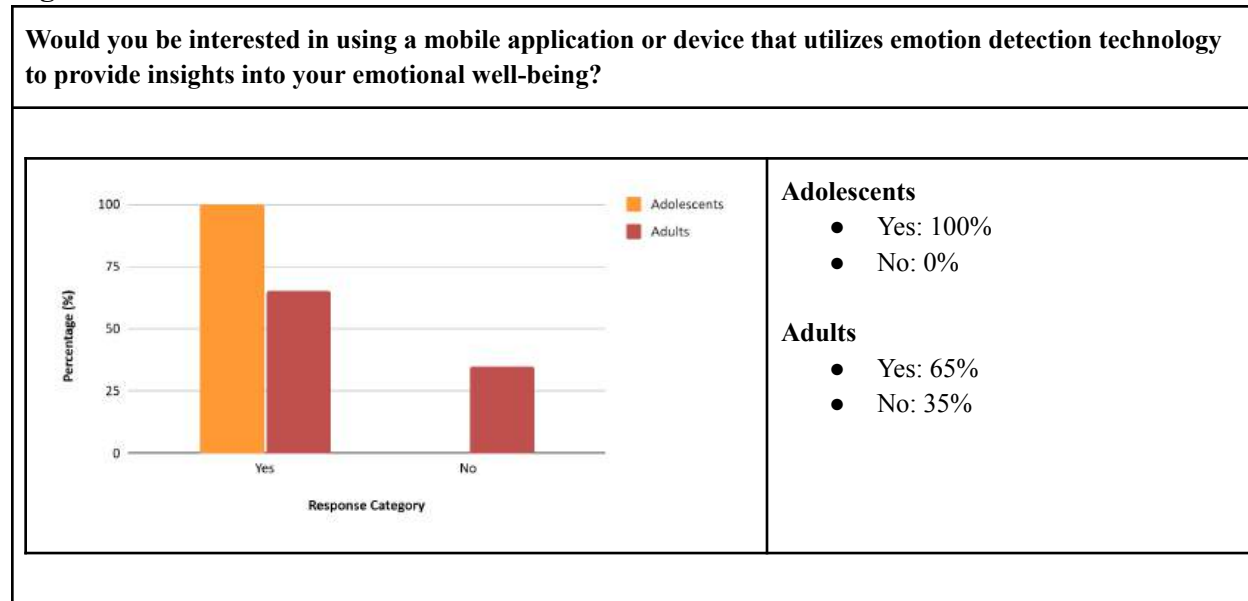


Figure 5:

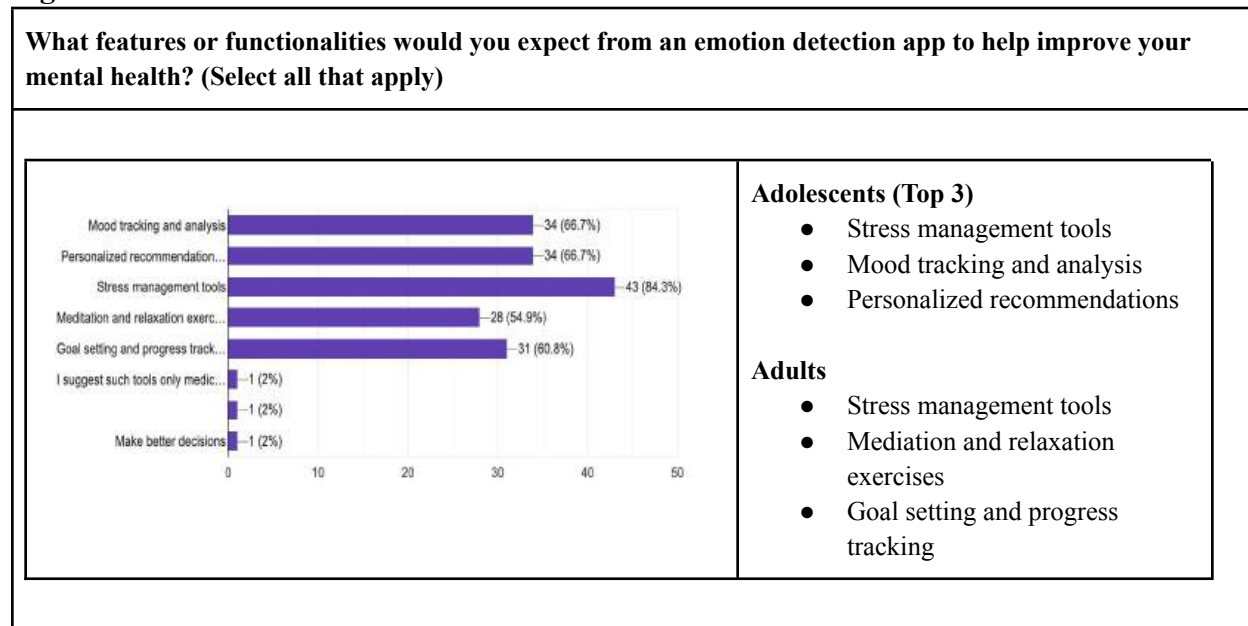


Figure 6:

Are you someone who can efficiently detect what emotion you are feeling such as angry, sad, mad, nervous, etc?

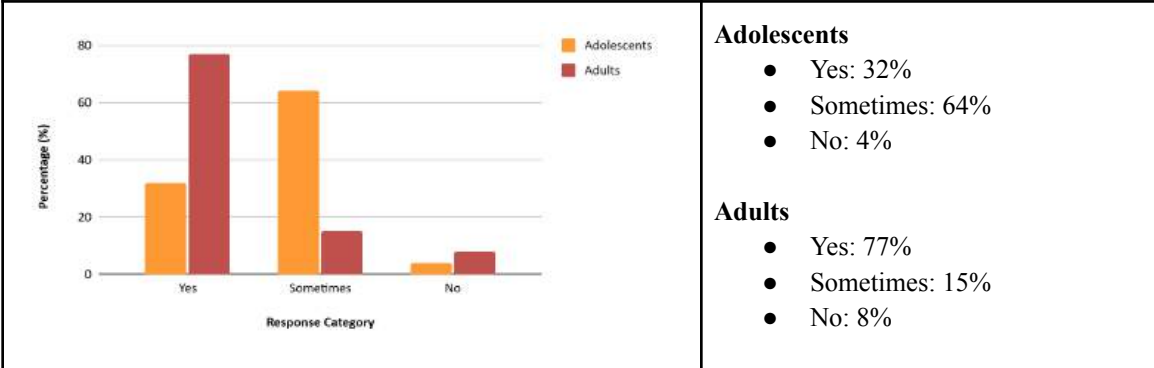


Figure 7:

Are you someone who can efficiently detect what emotion you are feeling such as angry, sad, mad, nervous, etc?

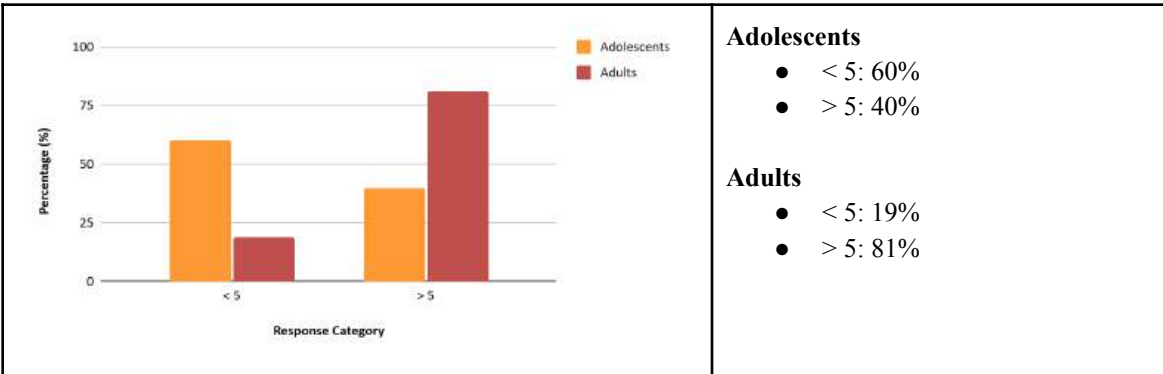


Figure 8:

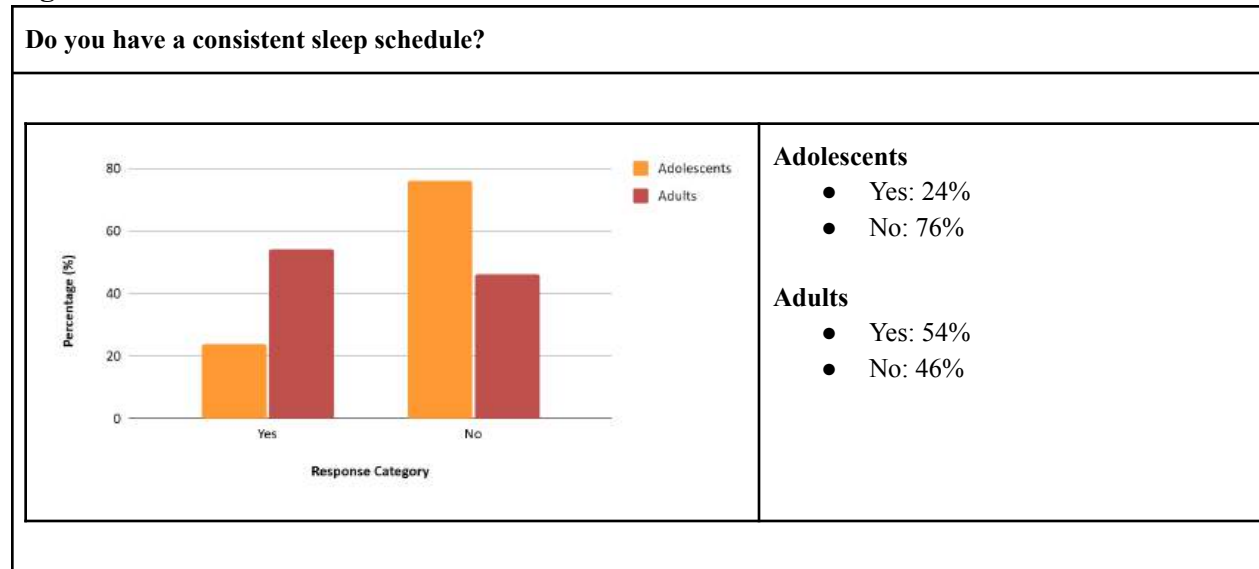


Figure 9:

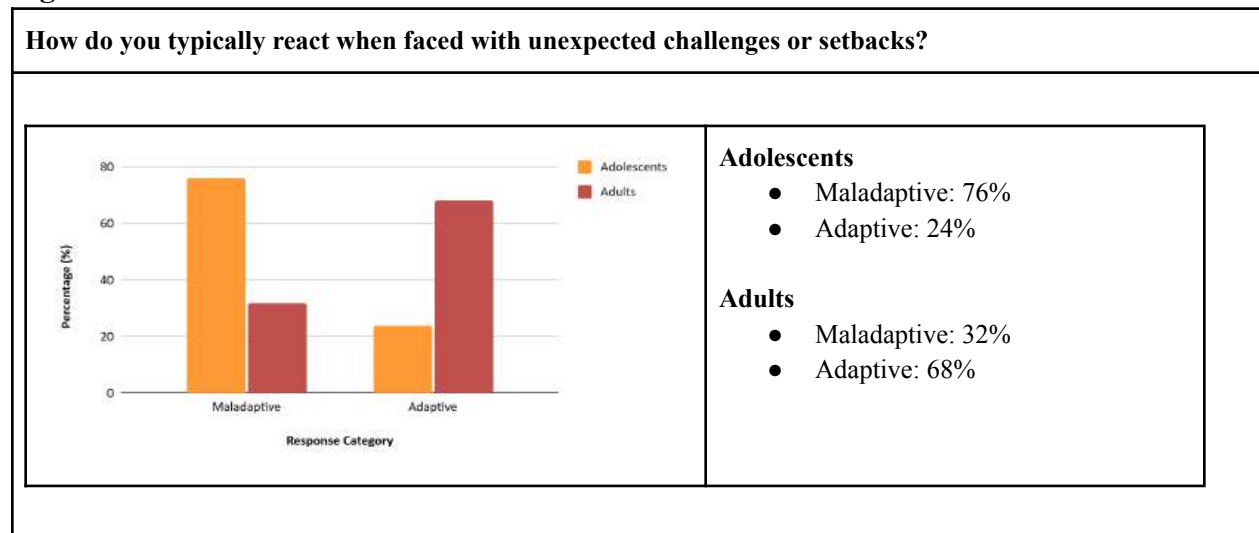


Figure 10:

How likely are you to believe that tracking and analyzing your emotions could help you better understand your mental health?

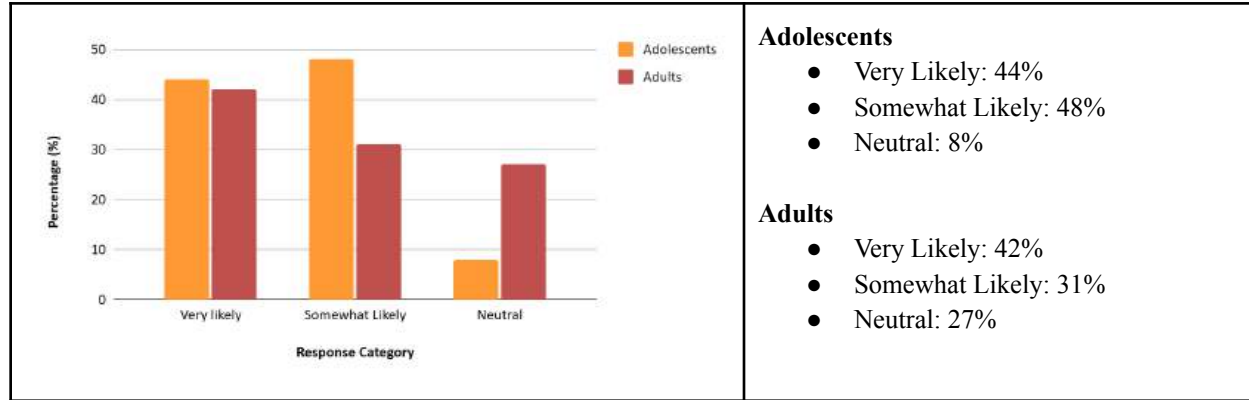


Figure 11:

Do you think personalized recommendations from an app could help you manage stress and improve your overall lifestyle?

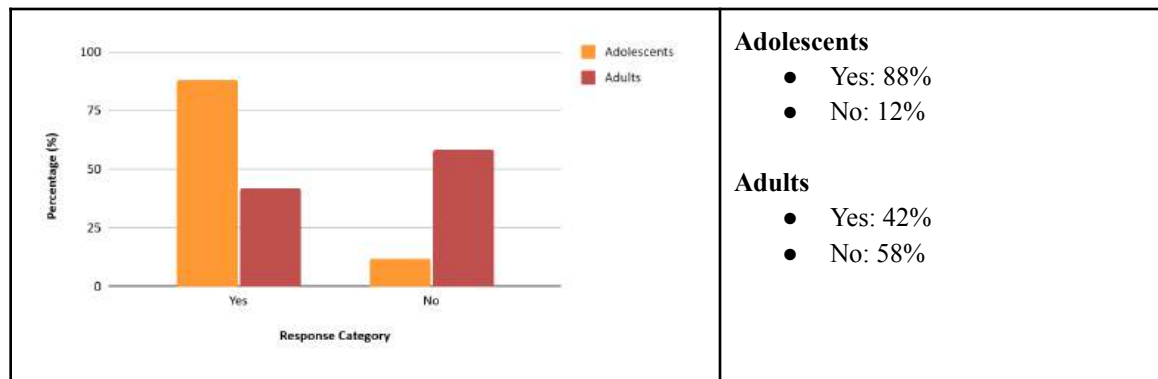


Figure 12:

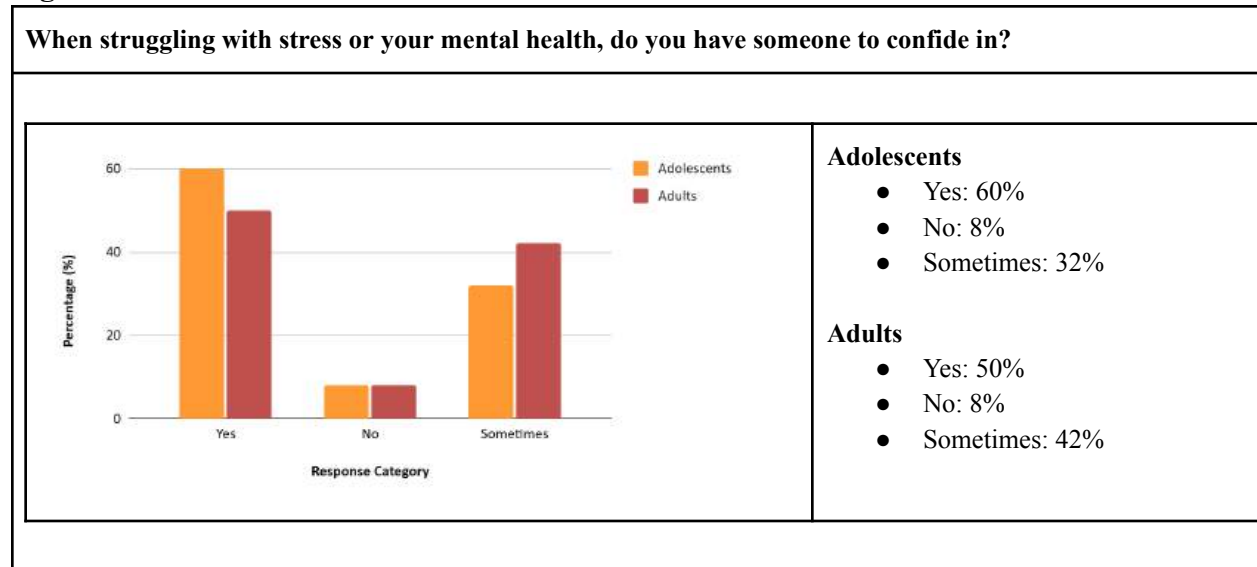
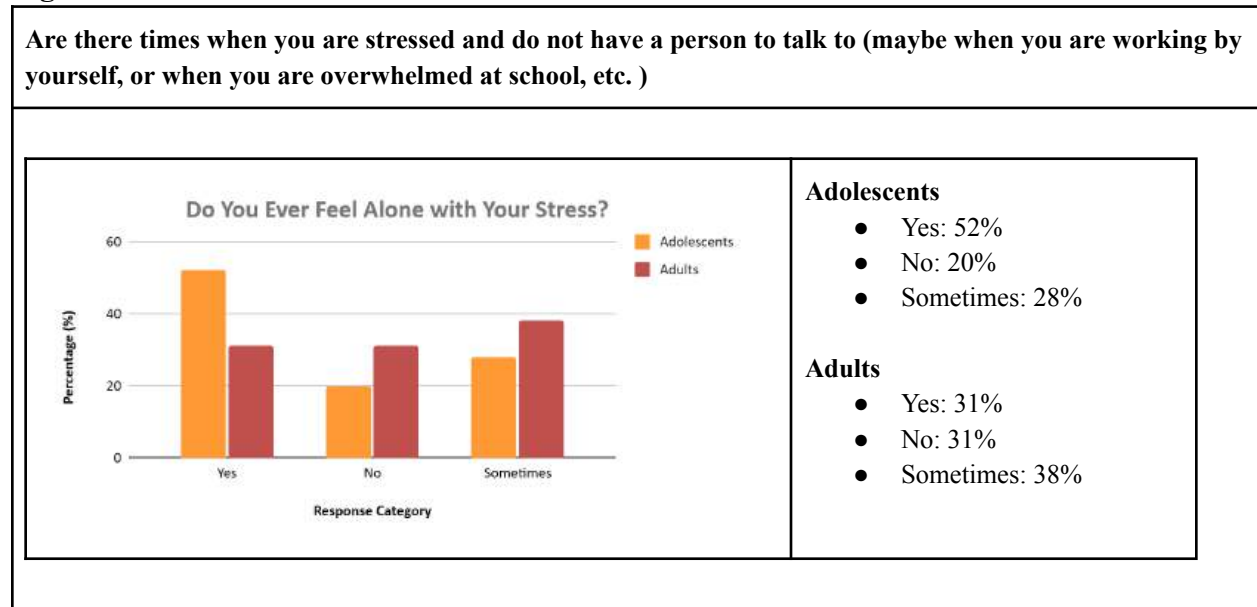


Figure 13:



DISCUSSION

The results support the initial hypothesis that a significantly higher percentage of adolescents (under 18) would positively perceive the use of emotion detection technology in the mental health sector and indicate a stronger need for it than adults (over 18), as displayed by the frequency distributions and thematic analysis performed.

ANALYSIS

This analysis is categorized into two sections. One is on the Individual's perception of

emotion detection applications, and the other is on the Individual's perception of their overall well-being. It is important to note that a two-hypothesis test was performed on the results with a significance level of 5%. If the *null hypothesis is not rejected*, it indicates that there is no statistically significant difference between adolescents and adults. Conversely, if the *null hypothesis is rejected*, it suggests that there is a statistically significant difference between adolescents and adults.

Individuals' Perception of emotion-detection applications

The initial question assessed an individual's knowledge regarding this topic: *Have you ever heard of emotion detection technology before?* (Figure:2). The findings revealed that a significant proportion of adults (76%) and adolescents (72%) responded negatively, resulting in the null hypothesis not being rejected, suggesting that the relatively recent emergence of emotion detection technology may contribute to a general lack of awareness in both age groups.

The next set of questions assessed the individual's perceptions of an emotion detection application, like a mobile device app. When analyzing the results of the question: *How do you perceive the potential benefits of emotion detection technology in gaining insights into human behavior and mental health* (Figure:3), 84% of adolescents responded positively, assigning it a high or highest priority (extremely or very beneficial), while only 54% of adults held a similar view. This results in the null hypothesis being rejected, indicating that adults have a significantly more conservative view regarding the perceived benefits of this technology for mental health outcomes.

Next was the question: *How likely are you to believe that tracking and analyzing your emotions could help you better understand your mental health?* (Figure:10), 92% of adolescents responded with a highly positive or positive stance (very likely, somewhat likely), whereas only 73% of adults responded positively. This results in the null hypothesis being rejected, showing that adolescents have a statistically significant more positive attitude towards the potential benefits of utilizing emotion detection technology to enhance their mental health than adults.

In the next question, the participants were asked, *Do you think personalized recommendations from an app could help you manage stress and improve your overall lifestyle?* (Figure:11), 88% of adolescents agreed, but only 42% of adults agreed. The null hypothesis is rejected, and there is a significant gap in the results, showing that adults are less open to embracing the potential benefits offered by emotion detection technology.

In the next question, participants were asked: *Would you be interested in using a mobile application or device that utilizes emotion detection technology to provide insights into your emotional well-being?* (Figure:4). Amazingly, all adolescents responded affirmatively, while only 65% of adults expressed interest. This results in the null hypothesis being rejected, and the significant gap suggests that adults are less open to using this technology than adolescents.

In the final question of this section, participants were asked - *what features or functionalities would you expect from an emotion detection app to help improve your mental health* (Figure:5). The top three choices amongst adolescents were *stress management tools*,

mood tracking and analysis, and personalized recommendations for managing emotions. Conversely, adults prioritized *stress management tools, meditation and relaxation exercises, and goal setting/progress tracking.* These preferences suggest that adolescents and adults prioritize stress management tools as a key feature in an emotion detection app. However, they also reveal distinct preferences and needs across different age groups, emphasizing the importance of tailoring technology to accommodate each group effectively.

Adolescents show a significantly more positive attitude towards emotion detection technology, emphasizing its potential benefits for mental health and expressing higher interest in using such applications than adults. While both groups prioritize stress management tools, adolescents favor personalized recommendations, whereas adults lean towards meditation and relaxation exercises. This indicates the need to address mental health through technology among different age demographics respectively.

Individual's Perception of Their Well-being and Need For the Application

This section focuses on the participants' self-assessment of their well-being. Questions regarding individuals' perceptions of their well-being were key in examining potential correlations between their symptoms and attitudes toward the app. It also provides insights into why adolescents and adults perceive the app's importance and necessity differently. Overall, it assesses the potential need for such an app within each demographic group.

The first set of questions assessed the individual's sense of isolation. Firstly, *when struggling with stress or your mental health, do you have someone to confide in (Figure:12)*, 60% of adolescents and 50% of adults said they have someone to confide in, so there is not a significant difference in the results. This is verified by the fact that the null hypothesis is not rejected. Secondly, participants were asked - *Are there times when you are stressed and do not have a person to talk to (Figure:13)*. The results indicate that 80% of adolescents and 69% of adults reported experiencing such moments. Although there is not a significant difference, as there is not enough evidence to reject the null hypothesis, it does show that both groups encounter feelings of isolation concerning mental health issues. The fact that a greater percentage of adolescents than adults indicated lacking a confidant in times of stress may explain why adolescents indicate a higher need for such an application, especially one with stress management tools, and why they are more likely to view its implementation positively.

The next set of questions assessed sleep and work-life balance. Firstly, the participants were asked - *On a scale of 1 to 10 (10 being the best), how satisfied are you with your work-life balance? (Figure:7)*. For adolescents, the most common response was a 5, and 60% of adolescents overall responded with a rate of 5 or below. On the contrary, most adults rated their work-life balance a 7, and overall, 81% rated above a 5, showing that adults can manage their work-life much more efficiently than adolescents. This is further evidenced by the fact that the null hypothesis is rejected, suggesting a statistically significant different outcome. Secondly, on the question about *whether you have a consistent sleep schedule (Figure:8)*, 76% of adolescents indicated that they do not have a consistent sleep schedule, which typically indicates that

adolescents are more likely to have a disturbed sleep schedule or chronic sleep deprivation, which a study by NIH proves is linked to emotional dysregulation and confusion, indicating why adolescents are more receptive to the app (Saghir et al., 2018). Meanwhile, 46% of adults indicated that they do not have a consistent sleep schedule, which is significantly less than that of adolescents, as supported by the fact that the null hypothesis is rejected. This shows that adults have adapted their lifestyles better to prevent further stress. These results display that adults have better control over balancing their work-life, which can indicate why adults do not see as much of a need for such an application as adolescents.

The next question is regarding an individual's coping mechanism - *how you typically react when faced with unexpected challenges or setbacks (Figure:9)*. The options for this question were in two groups: maladaptive and adaptive. Adaptive includes “Stay calm and adapt,” “Process the situation and my emotions without reacting,” “anxious but adaptable,” “I would be worried but not let those worries take over me,” and “Assert dominance.” Maladaptive refers to coping behaviors, thoughts, or actions that are counterproductive, ineffective, or dysfunctional in dealing with stressful situations, including the answers “feel overwhelmed and stressed” and “become anxious and worried.” The majority, 76%, of adolescents “feel overwhelmed and stressed,” a maladaptive response that exacerbates negative emotions (Brown & Bond, 2019). The majority of adults, 68%, put “stay calm and adapt” as an adaptive response that can reduce negative emotions. As shown, adolescents are more likely to develop maladaptive coping mechanisms. This could be due to the prefrontal cortex not being fully developed and the brain being mostly run by the limbic system (which becomes more sensitive to common occurrences such as social rejection) (Limbic Changes, 2019). Moreover, adolescents are still developing their coping strategies, their work ethic, and how to balance their tasks, leading to depressive symptoms, which are associated with maladaptive coping strategies (Pimenta et al., 2021). There is a significant difference in the types of responses, supported by the fact that the null hypothesis is rejected, with the majority of adults putting adaptive responses and a majority of adolescents putting maladaptive responses, showing the higher need for such an application among adolescents explaining why they have a more positive perception.

In the final section, participants were asked about their emotional maturity. Firstly, participants were asked - *Are you someone who can efficiently detect what emotion you feel, such as anger, sad, mad, nervous, etc. (Figure:6)?* Majority of the adolescents answered “sometimes” (64%), indicating a degree of uncertainty regarding their emotional states. In comparison, the majority of adults answered ‘yes’ (77%), indicating a degree of confidence regarding their emotional states. Secondly, when the participants were asked - *How important do you believe it is to understand and manage your emotions for overall well-being (Figure:1)*, the majority of adults (69%) and adolescents (68%) had a positive response to this question. This results in the failure to reject the null hypothesis, indicating that the two values are not statically significant enough to be different and that individuals from both groups understand the benefit of the impact this technology can have.

CONCLUSION

Fulfilling Gap in Research

Perception Towards Emotion Detection Technology: Previous studies examined general perceptions of AI-based technology in healthcare. This research, however, specifically focused on AI-based emotion detection technology and its perception within the mental health field. The findings indicate that perceptions vary across demographics, emphasizing the need for more tailored solutions to encourage better adaptation.

Age-based Demographic Perception of AI-based Technology: Previous studies primarily examined adults' perceptions regarding adopting AI-based technology in healthcare. This research, however, focused on comparing the perceptions of adolescents and adults, revealing that adolescents have a more favorable perception of AI-based emotion detection technology.

Understanding the Perception of Overall Well-being: This research also fills the gap in understanding the differences in individuals' perceptions of their overall well-being between adolescents and adults and correlates it with technology adaptation.

The results show that the hypothesis is accepted because adolescents perceived AI-powered emotion detection technology more positively. Adolescents showed a greater need for technology than adults, as they have more disturbed sleep patterns, are less confident in detecting their emotions, and are prone to isolation, which could explain why they perceive the technology positively.

Implications

The results from this study are significant because they highlight the perception and need for this new-coming technology, emotion detection technology, and provide an overview of what features are needed and positively perceived by adolescents versus adults. The analysis of the results showed that adolescents displayed a higher need, leading them to hold a more positive view of the technology compared to adults, which can play a pivotal role for healthcare workers in departments such as psychology and mental health to get a better understanding of the variances in the age groups. Furthermore, identifying specific features and functionalities desired by adolescents, such as stress management tools, provides insights for developers and researchers in tailoring future applications to better meet the individual's needs.

Limitations

Although this research successfully collected data and drew conclusions to address the research gap, several limitations were encountered. For instance, one limitation was the small sample size ($n=50$), which, if larger and more diverse, could have allowed for generalization to a broader public rather than individuals of similar socioeconomic status, culture, and locale. The makeup of the sample population could skew the data since middle/upper-middle-class individuals can be seen to have less stress than lower-class individuals (Manstead A. S. R. 2018).

Furthermore, the Asian-majority patient population could have skewed the data since Asians tend to have a stigma against accepting/understanding mental health (*Why Asian Americans Don't Seek Help for Mental Illness* | *McLean Hospital*, 2023).

Another limitation was the potential for response bias in the results. Since the questions heavily depended on an individual's well-being, participants may have been inclined to provide socially "correct" responses rather than expressing their genuine beliefs, potentially impacting the validity of the results..

Nonetheless, these limitations did not significantly impact the results, as samples from various demographic groups helped balance the results. Still, it is important to consider these limitations.

Areas for Future Research

Emotion detection technology is expanding as research on Artificial Intelligence continues to develop; therefore, there are many areas for future research.

There is potential for further research into the development of new technologies. From this research, it is seen that there are varying needs and perceptions. Therefore, technology developers can note how to expand the existing features and functions to better suit the variances in age groups.

Another important direction for further research involves tackling the pervasive bias gap in the demographics represented in emotion recognition technologies. Emotion detection technology is a relatively new area of development, and there is a lack of data; therefore, expanding the datasets to various cultural, economic, and ethnic backgrounds can help with validity.

Works Cited

- Akinrinmade, A. O., Adebile, T. M., Ezuma-Ebong, C., Bolaji, K., Ajufo, A., Adigun, A. O., Mohammad, M., Dike, J. C., & Okobi, O. E. (2023). Artificial Intelligence in Healthcare: Perception and Reality. *Cureus*, 15(9), e45594. <https://doi.org/10.7759/cureus.45594>
- American Psychological Association survey shows teen stress rivals that of adults. (2014a, February 11). <https://www.apa.org>.
<https://www.apa.org/news/press/releases/2014/02/teen-stress>
- Antony, C. J., Pariyath, B., Safar, S., Sahil, A., Nair, A. R., & Department of Computer Science and Engineering, Rajagiri School of Engineering and Technology. (n.d.). Emotion Recognition-Based Mental Healthcare Chat-bots: a survey. In *International Conference on IoT based Control Networks and Intelligent Systems (ICICNIS 2020)* (pp. 69–70).
- Australian Bureau of Statistics. (2024, April 11). Australian Bureau of Statistics.
<https://www.abs.gov.au/>
- Bhardwaj, P. (2019). Types of sampling in research. *Journal of the Practice of Cardiovascular Sciences*, 5(3), 157. https://doi.org/10.4103/jpcs.jpcs_62_19
- Brown, L., & Bond, M. J. (2019). The pragmatic derivation and validation of measures of adaptive and maladaptive coping styles. *Cogent Psychology*, 6(1).
<https://doi.org/10.1080/23311908.2019.1568070>
- Cai, Y., Li, X., & Li, J. (2023). Emotion recognition using different sensors, emotion models, methods and datasets: A Comprehensive review. *Sensors*, 23(5), 2455.
<https://doi.org/10.3390/s23052455>
- Centers for Disease Control and Prevention. (2021). *YOUTH RISK BEHAVIOR SURVEY DATA SUMMARY & TRENDS REPORT 2011-2021*.
https://www.cdc.gov/healthyyouth/data/yrbs/pdf/YRBS_Data-Summary-Trends_Report2023_508.pdf
- CompanionMX | STEX. (n.d.). <https://startupexchange.mit.edu/startup-features/companionmx>
- Foulkes, L., & Andrews, J. L. (2023). Are mental health awareness efforts contributing to the rise in reported mental health problems? A call to test the prevalence inflation hypothesis. *New Ideas in Psychology*, 69, 101010.
<https://doi.org/10.1016/j.newideapsych.2023.101010>
- Fritsch, S., Blankenheim, A., Wahl, A., Hetfeld, P., Maaßen, O., Deffge, S., Kunze, J., Rossaint, R., Riedel, M., Marx, G., & Bickenbach, J. (2022). Attitudes and perception of artificial intelligence in healthcare: A cross-sectional survey among patients. *Digital Health*, 8, 205520762211167. <https://doi.org/10.1177/20552076221116772>
- Khullar, D., Casalino, L. P., Qian, Y., Lu, Y., Krumholz, H. M., & Aneja, S. (2022). Perspectives of patients about artificial intelligence in health care. *JAMA Network Open*, 5(5), e2210309. <https://doi.org/10.1001/jamanetworkopen.2022.10309>
- Limbic changes. (2019, March 12). Centre for Educational Neuroscience.
<http://www.educationalneuroscience.org.uk/resources/the-adolescent-brain/limbic-change>

What is Mental Illness? (n.d.).

<https://www.psychiatry.org/patients-families/what-is-mental-illness>

Why Asian Americans don't seek help for mental illness | McLean Hospital. (2023, May 2).

<https://www.mcleanhospital.org/essential/why-asian-americans-dont-seek-help-mental-illness#:~:text=Asian%20Americans%20fear%20being%20thought,to%20work%20around%20the%20stigma.>

How Food Insecurity Affects African Children's Mental Health By Fabrizio Manganiello

Abstract

This study investigates the profound psychological effects of food insecurity on children across Africa, a subject that, while critically significant, remains under-explored. Food insecurity, a condition marked by unreliable access to affordable and nutritious food, is notably pervasive in Africa due to factors like limited agricultural productivity, adverse climate conditions, and economic stagnation. Particularly, this paper focuses on how food insecurity during childhood, a crucial period for cognitive and emotional development, correlates with psychological impairments in children.

Employing a systematic literature review, we analyzed empirical studies across multiple databases to identify the relationship between food insecurity and mental health outcomes in African children. Our review synthesized findings from 95 initial studies, rigorously assessed for quality and relevance. The studies examined reveal a consistent association between food insecurity and adverse mental health outcomes, including cognitive decline and emotional distress, which are exacerbated by the socio-economic context within the continent. Despite the consensus on negative outcomes, regional disparities in the magnitude of these effects were noted, suggesting the need for localized research and interventions. This paper highlights the urgent necessity for multidimensional approaches to tackle food insecurity, considering its psychological, nutritional, and social ramifications. Future research directions include longitudinal studies to better understand causal relationships and more geographically inclusive studies to ensure findings are representative of the diverse African context.

Introduction

Food insecurity, as defined by the Oxford Languages, is the state of being without reliable access to a sufficient quantity of affordable, nutritious food. It is a global challenge that disproportionately affects vulnerable populations, particularly children in regions like Africa. The consequences of food insecurity extend beyond immediate nutritional deficiencies, influencing various aspects of health and development, including mental health. This paper aims to explore the intricate relationship between food insecurity and its psychological impact on children in Africa, a subject of critical importance yet insufficiently studied.

Africa's struggle with food insecurity is exacerbated by factors such as limited agricultural productivity, adverse climatic conditions, stagnant economic progress, governance issues, and a scarcity of research and innovation. Despite the known high mortality rates among children due to malnutrition (Cassidy-Vu L et al., 2022), the specific effects of food insecurity on children's mental health remain under-researched. This gap in the literature calls for a focused examination of how insufficient nutrition during childhood—a pivotal period for cognitive and emotional development—can lead to lasting psychological impairments (Darling et al., 2020). Given the vast and diverse nature of the African continent, this study adopts a broad geographical scope to encompass a wide array of experiences and data, thereby enriching our

understanding of food insecurity's impacts across different African contexts. By honing in on children's mental health, operationalized through observable behaviors such as aggression, depression, and disrespect towards authority, this research contributes to a deeper understanding of the long-term consequences of food insecurity, laying the groundwork for interventions that could mitigate these effects.

Results

The systematic review of literature revealed a nuanced landscape of how food insecurity impacts the mental health of children across various regions in Africa. The analysis of selected studies, such as those by Ae-Ngibise KA et al., 2021, and Skeen S et al., 2014, among others, provides compelling evidence that food insecurity is intricately linked to adverse mental health outcomes in children. This section synthesizes key findings, highlighting both consistencies across studies and notable discrepancies that suggest areas for further investigation.

Cognitive and Emotional Impacts: Ae-Ngibise KA et al. (2021) conducted a systematic review focusing on West Africa, revealing significant evidence of cognitive decline in children associated with food insufficiency. The study highlighted that children in food-insecure households displayed lower intelligence test scores and reading skills compared to their counterparts in food-secure environments. This finding underscores the critical importance of nutritional security during the first five years of life, a formative period for cognitive and social development.

Similarly, Heany S et al. (2021) found that food insecurity among HIV-positive adolescents in South Africa was significantly associated with emotional problems, as measured by the Beck Youth Inventories. These emotional issues were directly ascribed to the stress and anxiety stemming from food insecurity, illustrating the profound impact of nutritional deficiencies on mental health.

Psychological Morbidity and Caregiver Impact Skeen S et al. (2014): This study provided insight into the broader social implications of child food insecurity, noting that caregivers in food-insecure households in South Africa and Malawi were more likely to exhibit psychological morbidity and suicidal ideation. This relationship highlights the interconnectedness of child and caregiver mental health, suggesting that interventions must consider the family unit's overall well-being.

Discrepancies and Regional Variations: While the general consensus underscores the negative impact of food insecurity on mental health, discrepancies in the magnitude of these effects were noted across different studies and regions. For example, Parchment TM et al. (2016) observed that food insecurity contributed to a heightened risk of poor mental health among children in Sub-Saharan Africa, though the specific nature of mental health issues varied widely from one study to another.

The COVID-19 Pandemic as an Aggravating Factor: The recent study by Garcia-Prats AJ et al. (2021) hypothesized the exacerbating effects of the COVID-19 pandemic on food insecurity and, by extension, on children's mental health. The pandemic's disruption to food

supply chains and economic stability is anticipated to increase the prevalence of food insecurity, posing significant risks to children's mental and physical health across the continent.

Synthesis of Findings

The reviewed literature consistently indicates that food insecurity is a significant determinant of mental health outcomes in African children, affecting cognitive development, emotional well-being, and psychological stability. However, the observed discrepancies in the specific mental health outcomes reported, alongside regional variations in the impact of food insecurity, underscore the complexity of this issue and the need for localized research and interventions.

This synthesis not only answers the research question but also highlights the critical gaps in the current understanding of food insecurity's impact on child mental health in Africa. The findings call for a multidimensional approach to tackling food insecurity, one that incorporates nutritional, psychological, and social interventions to mitigate its far-reaching effects on children's well-being.

Implications for Future Research

The results illuminate the pressing need for further research in this area, particularly studies that delve into the mechanisms through which food insecurity impacts mental health and those that explore intervention strategies to mitigate these effects. Future research should also aim to address the regional disparities observed, providing insights that can inform tailored interventions suitable for the diverse contexts within the African continent.

Discussion

The synthesis of findings from the literature review reveals a complex interplay between food insecurity and mental health outcomes in African children. This section delves into the theoretical frameworks that help explain these relationships, examines the potential biases and limitations of the existing studies, and discusses the practical implications of the findings.

Theoretical Frameworks

1. **Developmental Psychopathology:** This framework suggests that early exposure to adverse conditions such as food insecurity can disrupt the typical trajectory of child development, leading to long-term psychological and emotional issues. The evidence of cognitive decline and emotional problems linked to food insecurity aligns with this theory, highlighting the critical need for early intervention to prevent the cascading effects on mental health.
2. **Social Determinants of Health:** This perspective emphasizes the role of social factors, including economic stability and access to resources, in determining health outcomes. Food insecurity, as a manifestation of economic and social inequalities, directly impacts

children's mental health by limiting their access to necessary nutritional resources, thereby exacerbating social disparities in health outcomes.

Methods

This study employed a comprehensive literature review to investigate the impact of food insecurity on the mental health of African children. A systematic search was conducted across multiple databases, including PubMed, Scopus, and Web of Science, to ensure a wide coverage of relevant literature. The search strategy was meticulously designed using the following terms: ("food insecurity" OR "food insufficiency" OR "malnutrition") AND ("children" OR "child") AND ("mental health") AND ("Africa"). This approach aimed to capture the multifaceted nature of food insecurity and its diverse impacts on mental health within the African context. The search yielded an initial pool of 95 articles. To refine this pool, articles were screened based on the following inclusion criteria: empirical studies focusing on the psychological effects of food insecurity on children in Africa, published in English, and available in full text. Exclusion criteria were set to omit articles behind paywalls, those not directly relevant to the topic, and studies not conducted in Africa. This process ensured a focused and relevant dataset for analysis. Each selected article underwent a rigorous quality assessment, employing tools appropriate for quantitative and qualitative studies, such as the Critical Appraisal Skills Programme (CASP) checklists. This step was crucial for evaluating the credibility and contribution of each study to the research question. Data extraction was systematically conducted, with key findings, methodologies, and contexts carefully noted. The synthesis of extracted data was guided by thematic analysis, allowing for the identification of patterns and insights regarding the relationship between food insecurity and mental health outcomes among African children.

Methodologies

The reviewed articles provide valuable insights into the impact of food insecurity on children's mental health. However, a critical examination reveals several limitations:

- **Cross-sectional vs. Longitudinal Designs:** Most studies utilize cross-sectional designs, which can identify associations but not causation. There's a need for longitudinal research to trace the long-term effects of food insecurity on mental health.
- **Measurement Variability:** The studies employ a variety of instruments to measure mental health outcomes, leading to potential inconsistencies in how these outcomes are defined and assessed. Standardizing measurement tools across studies could enhance comparability.
- **Regional Focus:** While the review encompasses studies from various parts of Africa, there is an uneven distribution of research focus, with some regions underrepresented. This gap may obscure specific regional challenges and solutions.

Practical Implications

The findings underscore the urgency of addressing food insecurity as a public health priority. Practical implications include:

1. **Integrated Interventions:** Developing programs that simultaneously address food insecurity and its psychological impacts. For instance, school-based feeding programs can improve both nutritional status and academic performance, potentially mitigating some of the cognitive and emotional issues associated with food insecurity.
2. **Policy Advocacy:** The evidence supports advocating for policies that enhance food security, such as improving agricultural infrastructure, ensuring economic stability, and investing in research and innovation to prevent food crises.
3. **Community Engagement:** Mobilizing community resources to support vulnerable families through community gardens, food banks, and education on nutrition can help alleviate the immediate effects of food insecurity.
4. **Mental Health Services:** Expanding access to mental health services for children in food-insecure regions, including counseling and psychological support, to address the emotional and psychological consequences of malnutrition.

Future Directions

This analysis highlights the need for a multidisciplinary approach to research and intervention design, incorporating insights from psychology, nutrition, sociology, and public health to fully understand and address the impacts of food insecurity on children's mental health. Future studies should aim to fill the identified gaps, particularly in terms of methodological rigor and regional focus, to develop a more comprehensive understanding that can inform effective interventions.

Addressing food insecurity's impact on children's mental health requires a concerted effort from researchers, policymakers, and practitioners. By applying a holistic lens and focusing on both prevention and intervention, it is possible to mitigate the adverse effects of food insecurity and promote healthier developmental outcomes for children across Africa.

Limitations

This study, while comprehensive in its approach to understanding the relationship between food insecurity and mental health outcomes in African children, is subject to several limitations that warrant acknowledgment and consideration for future research.

1. **Scope of Literature Review:** The review was conducted primarily through PubMed, supplemented by additional databases to broaden the search. Despite efforts to include a diverse range of sources, the reliance on English-language publications and accessibility of articles may have introduced selection bias, potentially excluding relevant studies published in other languages or in less accessible journals.

2. **Heterogeneity of Studies:** The included studies vary significantly in their methodologies, measures of mental health, and definitions of food insecurity. This heterogeneity, while offering a broad view of the issue, complicates the process of drawing generalized conclusions. Future research would benefit from standardized measures and definitions to facilitate more direct comparisons and meta-analyses.
3. **Regional Representation:** While the study aimed to cover the entire continent of Africa, the distribution of research is uneven, with certain regions such as Sub-Saharan Africa being more heavily represented than others. This uneven representation may limit the applicability of findings across all African contexts, highlighting the need for more geographically diverse studies.
4. **Cross-sectional Nature of Most Studies:** A majority of the reviewed studies are cross-sectional, which limits the ability to infer causality between food insecurity and mental health outcomes. Longitudinal studies are necessary to establish temporal relationships and better understand the long-term effects of food insecurity on mental health.
5. **Lack of Theoretical Frameworks:** The application of theoretical frameworks to interpret findings was limited, which could have enriched the understanding of the mechanisms by which food insecurity impacts mental health. Incorporating theories from developmental psychology, sociology, and public health could provide a more nuanced analysis.
6. **Implications for Intervention:** While the study highlights the need for interventions, the specific design and efficacy of such interventions remain underexplored. Research focusing on the implementation and evaluation of programs targeting both food insecurity and mental health is needed to provide actionable insights.
7. **COVID-19 Pandemic:** The ongoing COVID-19 pandemic poses a unique challenge to both food security and mental health, potentially exacerbating the issues identified in this review. Future studies should consider the pandemic's impact, as its full effects are yet to be fully understood.

Future Research Directions

Acknowledging these limitations, future research should strive to:

1. Incorporate a wider range of databases and languages in literature reviews to capture a more comprehensive body of research.
2. Employ longitudinal designs to better understand the causal relationships between food insecurity and mental health outcomes.
3. Expand the geographic focus of research to include underrepresented regions, ensuring findings are relevant across different African contexts.
4. Apply and develop theoretical frameworks that can guide the interpretation of findings and inform the design of interventions.

5. Evaluate the effectiveness of interventions designed to address both food insecurity and its mental health impacts, providing evidence-based recommendations for policy and practice.

By addressing these limitations, future research can build on the foundation laid by this study, advancing our understanding of food insecurity's impact on children's mental health and informing the development of more effective interventions and policies.

Works Cited

- Ae-Ngibise KA, Asare-Doku W, Peprah J, Mujtaba MN, Nifasha D, Donnir GM. The Mental Health Outcomes of Food Insecurity and Insufficiency in West Africa: A Systematic Narrative Review. *Behav Sci (Basel)*. 2021 Oct 25;11(11):146. doi:10.3390/bs11110146. PMID: 34821607; PMCID: PMC8614760.
- Cassidy-Vu L, Way V, Spangler J. The correlation between food insecurity and infant mortality in North Carolina. *Public Health Nutr*. 2022 Apr;25(4):1038-1044. doi:10.1017/S136898002200026X. Epub 2022 Jan 31. PMID: 35094744; PMCID:PMC9991794.
- Darling, J. C., Bamidis, P. D., Burberry, J., & Rudolf, M. C. J. (2020). The FirstThousand Days: Early, integrated and evidence-based approaches to improving child health: coming to a population near you? *Archives of Disease in Childhood*, 105(9), 837–841. <https://doi.org/10.1136/archdischild-2019-316929>
- Food Insecurity, n. Meanings, Etymology and More | Oxford English Dictionary.https://www.oed.com/dictionary/food-insecurity_n?tab=factsheet&tl=true. Accessed 4 Oct. 2023.
- Garcia-Prats AJ, McAdams RM, Matshaba M, Thahane L, Butteris SM, Conway JH, Slutsker M, St Clair NE, Haq H. Mitigating the Impacts of COVID-19 on Global Child Health: a Call to Action. *Curr Trop Med Rep*. 2021;8(3):183-189. doi:10.1007/s40475-021-00241-6. Epub 2021 May 11. PMID: 33996382; PMCID:PMC8112470.
- Heany S, Phillips N, Myer L, Zar H, Stein D, Hoare J. Physical development and mental health in South African perinatally HIV-positive adolescents on antiretroviral therapy and their caregivers with and without household food insecurity. *South Afr J HIV Med*. 2021 Dec 15;22(1):1316. doi: 10.4102/sajhivmed.v22i1.1316. PMID: 34966558; PMCID: PMC8689368.
- Oxford Languages and Google - English | Oxford Languages. <https://languages.oup.com/google-dictionary-en/>. Accessed 4 Oct. 2023.
- Parchment TM, Small L, Osuji H, McKay M, Bhana A. Familial and Contextual Influences on Children's Prosocial Behavior: South African Caregivers as Adult Protective Shields in Enhancing Child Mental Health. *Glob Soc Welf*. 2016 Mar;3(1):1-10. doi: 10.1007/s40609-016-0042-8. Epub 2016 Feb 4. PMID: 28401033; PMCID: PMC5385243.
- Skeen S, Tomlinson M, Macedo A, Croome N, Sherr L. Mental health of carers of children affected by HIV attending community-based programmes in South Africa and Malawi. *AIDS Care*. 2014;26 Suppl 1(0 1):S11-20. doi: 10.1080/09540121.2014.906559. Epub 2014 Apr 25. PMID: 24766642; PMCID: PMC4554389.
- Wudil AH, Usman M, Rosak-Szyrocka J, Pilař L, Boye M. Reversing Years for Global Food Security: A Review of the Food Security Situation in Sub-Saharan Africa (SSA). *Int J Environ Res Public Health*. 2022 Nov 11;19(22):14836. Doi: 10.3390/ijerph192214836. PMID: 36429555; PMCID: PMC9690952.

MicroRNA as a Potential Biomarker for Ovarian Cancer Diagnosis

By Hoang Tue Anh Nguyen

Abstract

Ovarian cancer is a deadly gynecologic malignancy with a low survival rate due to late-stage diagnosis. Current diagnostic methods, such as carbohydrate antigen 125 (CA125) and human epididymis protein 4 (HE4) biomarkers, and imaging techniques, have limitations in early detection. MicroRNA (miRNA), a class of small non-coding RNAs involved in gene expression regulation, has emerged as a potential biomarker for ovarian cancer. MiRNAs are stable in bodily fluids, making them suitable for non-invasive testing. Studies have shown that abnormal levels of certain miRNAs, including both tumor suppressors and oncogenes, are associated with ovarian cancer, suggesting their potential as early diagnostic markers. This review explores the potential of miRNAs as biomarkers for ovarian cancer, highlighting their advantages, limitations, and future research directions. This study also explores various amplification techniques such as nucleic acid amplification, isothermal amplification, and enzyme-free amplification that can improve miRNA detection sensitivity.

Keywords

MicroRNA, Ovarian Cancer, Biomarker, Cancer Detection, Gene Expression, Tumor Suppressor, Oncogene

Introduction

Ovarian cancer is a malignant disease, characterized by the abnormal growth of cells in the ovaries. The factors that contribute to this disease include genetic predisposition, reproductive history, and hormonal influences. It is difficult to diagnose and distinguish ovarian cancer within the early stages as its symptoms are either nonspecific or subtle. As a result, ovarian cancer is commonly diagnosed at advanced stages, resulting in complicating treatment and reducing survival rates (Arora et al., 2024).

Approximately three-quarters of invasive epithelial ovarian cancer symptoms appear at the advanced stage (Hennessy et al., 2009). Additionally, as the second most lethal gynecologic malignancy, ovarian cancer has an overall 5-year survival rate of approximately 45% (Howlander et al., 1975–2017). For early stage (I and II) ovarian cancer patients, the 5-year survival rate is around 90%. For late-stage (III and IV) patients, it is between 20% and 40% (Zeppernick and Meinhold-Heerlein, 2014; Schiavone et al., 2011). In fact, fewer than 20% of advanced-stage patients survive for more than a decade (Cress et al., 2015; Charkhchi et al., 2022).

Currently, several diagnostic techniques exist for ovarian cancer such as CA125 and HE4 biomarkers as well as positron emission tomography (PET) and computed tomography (CT) scans. CA125 is the most commonly used biomarker for ovarian cancer. This surface glycoprotein is found on epithelial cells, cells lining the surfaces of tissues and organs, and is

linked to the growth and metastasis of cancer cells. Early studies found elevated levels of CA125 in the blood (>35 U/mL) of approximately 80% of ovarian cancer patients (Zurawski et al., 1988; Bast et al., 1983; Bast et al., 1981). However, CA125 has limited clinical sensitivity in diagnosing patients at the early stages of ovarian cancer. In the preoperative setting, CA125 levels have been shown to possess limiting diagnostic value in distinguishing ovarian masses, even when employed in conjunction with multimodal diagnostic approaches (Timmerman et al., 2007). HE4, another surface glycoprotein, is also used in diagnosis of ovarian cancer. Elevated levels of HE4 are associated with serous and endometrioid ovarian tumors but not with clear-cell or mucinous tumors (Ruggeri et al., 2011; Drapkin et al., 2005). Another approach for characterizing ovarian masses is to combine PET and CT scans (Rieber et al., 2001; Grab et al., 2000; Castellucci et al. 2007). PET is a nuclear imaging technique that employs radioactive tracers to visualize metabolic activity within tissues, often detecting cancer by monitoring glucose uptake. On the other hand, CT utilizes multiple X-ray images to create a three-dimensional representation of organs. Studies have shown PET and CT can better detect distant lymph node spread of cancer (Nam et al., 2010). While PET and CT offer a valuable combination of structural (from CT) and functional information (glucose metabolism from PET) for assessing ovarian masses, it has limitations. For example, ovarian masses often have a cystic structure, meaning they are fluid-filled sacs, which can make them difficult to detect early with imaging techniques like PET and CT. Additionally, other studies using transvaginal ultrasound (TVU), magnetic resonance imaging (MRI), and PET/CT have demonstrated that negative findings on either MRI or PET do not preclude the possibility of early-stage or borderline ovarian tumors (Liberto et al., 2022).

MiRNA has been proposed as a potential biomarker for ovarian cancer. This RNA molecule is primarily used in the regulation of gene expression by degrading or inhibiting translation of target mRNA (Yokoi et al., 2018). By affecting the expression of multiple genes, miRNAs can target and regulate mRNAs. Numerous studies have demonstrated the role of miRNA in the pathogenesis of various diseases (Lu and Rothenberg, 2018). Furthermore, miRNAs are also involved in biological processes associated with cancer such as tumorigenesis, cell proliferation, differentiation, apoptosis, angiogenesis, invasion and metastasis, tumor resistance, epithelial-mesenchymal transition, and prognosis. Overall, in cancer development, miRNAs can exhibit either oncogenic or tumor-suppressive functions (Zhang and Lu, 2022; Zhao et al., 2022).

MiRNA

MiRNAs are small, non-coding RNAs that are endogenously expressed across a wide range of organisms. These regulatory RNAs can influence gene expression by forming complexes with proteins (such as miRNP) and recognizing specific target mRNA sequences, which can dictate the functional outcome (Carthew and Sontheimer, 2009; Ghildiyal and Zamore, 2009; Kim et al., 2009; Siomi and Siomi, 2009). Since their discovery in the 1990s, miRNAs have been widely researched, highlighting the importance of post-transcriptional gene

regulation in both cis and trans regulatory elements (Lee et al., 1993). MiRNAs are a promising area of research as they are able to influence the complex and extensive network of gene expression (Oliveira-Carvalho et al., 2012; Ambros, 2004; Orang et al., 2014).

For miRNA synthesis, a pri-miRNA precursor first needs to be transcribed by RNA polymerase II. Then, endonuclease enzymes DROSHA and DGCR8 process this in the nucleus. This results in a pre-miRNA sequence consisting of about 80–100 nucleotides (Denli et al., 2004; Lee et al., 2003). The miRNA is then transported to the cytoplasm with assistance from Exportin-5 (Bohnsack et al., 2004). There, ribonuclease Dicer cleaves the pre-miRNAs into double stranded mature miRNA (Bartel, 2004; Reddy, 2015).

The central dogma of biology focuses on the flow of genetic information from DNA to RNA to protein. First, the process of replication occurs, resulting in two new strands of DNA. The parent (original) strand separates to create 2 new strands. Various enzymes such as DNA polymerase, primase, ligase, helicase, and topoisomerase, are used in the process (Achar and Foiani, 2017). Following that, the genetic information in DNA is transcribed into mRNA. This occurs when RNA polymerase transcribes the DNA sequence into a complementary mRNA molecule. Following that, the mRNA molecule leaves the nucleus and travels to the cytoplasm where translation occurs. Translation starts when the mRNA binds to the ribosomal units. The ribosome initiates translation by scanning the mRNA molecule until it identifies the start codon, which serves as the initiation site for protein synthesis. tRNA are specific to each amino acid as they contain an anticodon matching the binding site of corresponding codon on mRNA. The ribosome progresses along the mRNA molecule, identifying each codon in sequence. For each codon, the ribosome recruits the corresponding tRNA molecule, ensuring that the tRNA's anticodon is complementary to the mRNA's codon. It then catalyzes the formation of a peptide bond between amino acids that are carried by the tRNA and the growing polypeptide chain, extending the protein sequence (Mercadante et al., 2023).

MiRNAs attach to the 3' UTR (untranslated region) of target mRNA. This attachment can stop the mRNA from being translated into a protein and can also cause mRNA deadenylation and decapping (Huntzinger and Izaurralde, 2011; Ipsaro and Joshua-Tor, 2015). However, miRNAs can also bind to other parts of the mRNA, including the 5' UTR and the promoter region, leading to gene silencing (Xu et al., 2014; Forman et al., 2008; Zhang et al., 2018). However, when miRNAs interact with the promoter region of a gene, they can initiate transcription (Dharap et al., 2013; O'Brien et al., 2018).

Under normal physiological conditions, microRNAs help maintain essential biological processes such as cell proliferation, differentiation, and apoptosis (Bruce et al., 2015; Esquela-Kerscher and Slack, 2006). Minor changes in miRNA levels can significantly impact the expression of many genes, potentially leading to cell transformation (Jeansonne et al., 2015; Pinatel et al., 2014; Ben-Hamo and Efroni, 2015; Sotiropoulou et al., 2009). Previous studies have demonstrated the regulatory role of miRNA in controlling cell proliferation and apoptosis. Cancer pathogenesis often involves aberrant cell growth and loss of apoptotic function (Cheng, 2005). Collectively, these studies suggest that miRNAs are involved in the development of

cancer by regulating cell growth and apoptosis. Gene expression is silenced by miRNA when it binds to the 5' UTR and coding regions in mRNA (Forman et al., 2008; Zhang et al., 2018). Transcription is induced when miRNA interacts with the promoter region (Dharap et al., 2013; Reddy, 2015; O'Brien et al., 2018).

Oncogenic microRNAs can promote tumorigenesis by downregulating the expression of tumor suppressor genes and/or genes involved in cellular differentiation and apoptosis. These miRNAs are often overexpressed. Underexpressed miRNAs, on the other hand, are tumor-suppressive miRNAs meaning they can inhibit tumorigenesis. They do this by downregulating oncogenes and/or genes associated with cellular differentiation and apoptosis (Zhang et al., 2007).

MiRNA as Potential Biomarker for Ovarian Cancer

MiRNA biomarkers show potential as a less invasive method for detecting cancer. Abnormal levels of miRNAs were observed at early stages of cancer progression, during metastasis, and after metastasis. This suggests that miRNA may serve as valuable clinical biomarkers for prognosis, tumor differentiation, and treatment strategy selection (Wang et al., 2018).

MiRNAs may be promising candidates for biomarkers due to their presence and stability in bodily fluids like blood plasma and serum (Weber et al., 2010). They are also relatively stable compared to others, with expression levels tending to not be unaffected by age, gender, body mass index, smoking, or other common factors when evaluating their pathogenic potential. Thus, routine examination of altered miRNA expression patterns can be used for monitoring and early diagnosis of ovarian cancer (Stawiski et al., 2024).

MiRNAs are stable due to three main factors: their short length, protective packaging, and association with protective proteins. Firstly, their short length of 20-24 nucleotides makes them less prone to degradation during formalin-fixed paraffin-embedded (FFPE) sample processing, a method used to preserve tissue samples for long-term storage and analysis. Secondly, they are encapsulated by cell-derived lipid vesicles such as exosomes, microparticles, or apoptotic bodies (Muralidharan-Chari et al., 2010; Zerneck et al., 2009; Turchinovich et al., 2011). However, the exact process by which miRNAs are specifically packaged within these structures remains unclear. Finally, miRNAs are protected from nuclease enzymes by binding with proteins such as argonaute proteins (Ago1 or Ago2), nucleophosmin (NPM1), or ribosomal proteins (Xi et al., 2007).

MiRNA stability has been tested under harsh conditions that typically degrade mRNA or other larger RNA molecules. Studies have shown that preserved tissue samples, using FFPE, dating back 10 years express highly similar miRNA profiles compared to fresh frozen samples. The preservation duration did not influence miRNA profiles (Leite et al., 2011; Chen et al., 2008). Furthermore, other studies have shown over half of miRNAs that were exposed to the enzyme RNase A remained intact after 3 hours compared to high molecular weight, all of which were rapidly degraded. Additionally, when exposed to harsh treatments such as boiling, extreme

pH levels, prolonged storage, or repeated freeze-thaw cycles, miRNAs in serum samples resisted degradation (Lan et al., 2015). Overall, these studies illustrated miRNAs' high stability against enzymatic cleavage as well as extreme conditions (Selcuklu et al., 2009).

Studies have indicated that abnormal levels of certain miRNAs are linked with cancer. By examining combinations of different miRNAs in blood, researchers can predict a patient's risk for developing cancer and treatment outcomes (Zhang et al., 2007; Zhang et al., 2016; Henry and Hayes, 2012; Fehlmann et al., 2018; Fehlmann et al., 2020) In women with ovarian cancer, circulating miRNA profiles are altered compared to those in healthy individuals. Specific profiles of miRNA can also predict the patient's prognosis (Yokoi et al., 2018; Bagnoli et al., 2016).

Most traditional cancer markers are produced by cancerous tissues or by tissues that are normally only present in developing embryos, with their levels being typically low or undetectable in healthy adults. Widely accepted cancer biomarkers include CEA (carcinoembryonic antigen) as well as CA199 and CA125 as they have positive predictive value as biomarkers for various cancers. Nevertheless, more specific and sensitive biomarkers are needed in detecting and predicting cancer. Circulating miRNAs may have a distinct advantage as a potential clinical tool. Moreover, miRNAs can be used to track the development and progression of cancer. Real time monitoring can be utilized as miRNAs are quickly produced in response to changes in the internal or external stimuli. For example, in other cancer types, such as lung cancer, tumor suppressive miRNA (miR-195-5p) levels were shown to be lower in plasma at advanced stages of the cancer (Wang et al., 2014; Guo et al., 2014; Liu et al., 2015). Furthermore, in pancreatic cancer, miR-373-3p, another tumor suppressive type of miRNA, was also shown to be downregulated. As the cancer progressed, with the spread of cancer to lymph nodes and distant organs, miR-373-3p levels decreased (Hua et al., 2017). These findings suggest that circulating miRNAs could be used to evaluate the stage and progression of tumors, illustrating the development of cancer throughout its course (Wang et al., 2018).

MiRNA families in Ovarian Cancer

MiRNAs can function as either tumor suppressors, oncogenes, or both. An example of oncogenic miRNA is miR-21. In the analysis of various tumor samples, miR-21 has been found to be consistently overexpressed. This overexpression suggests that miR-21 contributes to the development and progression of cancer (Selcuklu et al., 2009).

Research has indicated that higher levels of miR-21 in the blood are correlated with poorer long-term outcomes for cancer patients. One study found that miR-21 is significantly higher in late-stage ovarian cancer patients in comparison to early-stage patients and healthy individuals (Kartika et al., 2018). Researchers believe that miR-21 affects various biological processes in ovarian cancer cells such as proliferation and metastasis (Lou et al., 2010). While the exact mechanisms by which miR-21 contributes to cancer are still unknown, evidence suggests that miR-21 may promote cancer by suppressing tumor suppressor genes like PTEN and Pcd4 (Lou et al., 2010; Meng et al., 2007; Asangani et al., 2007; Chan et al., 2014). Other

studies have shown miR-21 to be the most frequently upregulated miRNA in biopsies of serous ovarian carcinoma relative to normal ovarian tissue (Nam et al., 2008). This overexpression has been observed in different types of ovarian cancer, including serous, mucinous and endometrioid carcinoma (Lou et al., 2010; Paliwal et al., 2020; Talaat et al., 2022).

Additionally, the miR-200 family of miRNAs can function as oncogenes or tumor-suppressors. This family includes miR-200a, miR-200b, miR-200c, miR-141 and miR-429. Members of this family act as tumor suppressors by blocking epithelial-to-mesenchymal transition (EMT) in cancer cells, which decreases their migration and invasion capabilities. For instance, miR-200c inhibits EMT and prevents metastasis by targeting the transcription factors ZEB1 and ZEB2, leading to increased expression of E-cadherin, an essential molecule for maintaining adhesion in epithelial cells. This role is especially crucial during the early stages of cancer, as miR-200c helps keep cells in a more epithelial, non-invasive state. However, the involvement of the miR-200 family in angiogenesis is more intricate; they can either promote or inhibit the formation of blood vessels based on the specific tissue environment and molecular interactions. Importantly, they are recognized for their ability to suppress the growth of blood vessels by downregulating pro-angiogenic factors such as vascular endothelial growth factor (VEGF). By inhibiting VEGF and associated pathways, miR-200 family members limit the tumor's capacity to establish a blood supply, thereby functioning as angiogenesis inhibitors. Several members of the miR-200 family, such as miR-200a, miR-200b, and miR-429, have anti-angiogenic effects. Furthermore, while miR-200c is generally recognized as an inhibitor of angiogenesis, it has also been shown to promote angiogenesis in certain other cancer types. MiR-141 can also act as a promoter and inhibitor of angiogenesis. In ovarian cancer, miR-141 promotes the formation of new blood vessels by activating the NF- κ B and JAK/STAT3 signaling pathways (Masoumi-Dehghi et al., 2020; Klicka et al., 2022).

Furthermore, examples of tumor-suppressive miRNAs include the miR-34 and let-7 families. The miR-34 family has the ability to affect cell proliferation and survival. When miR-34a and miR-34b/c were artificially introduced to cells, they caused cells to stop dividing at the G1 phase of the cell cycle (He et al., 2007; Tarasov et al., 2007; Bommer et al., 2007). MiR34b/c also inhibited cell proliferation and spread and growth of colonies in soft agar. Furthermore, when miR-34a and miR-34b/c were introduced into primary human diploid fibroblasts, the cells stopped dividing permanently (He et al., 2007; Collado et al., 2007). Moreover, re-introducing miR-34a led to apoptosis (Tarasov et al., 2007; Chang et al., 2007; Raver-Shapira et al., 2007; Welch et al., 2007). Cell cycle arrest and apoptosis, often the end results of p53 activation, a tumor suppressor protein, suggest that miR-34 genes play a significant role in p53's tumor-suppressing ability. Microarray analyses have revealed that introducing miR-34 genes can downregulate hundreds of other genes (Chang et al., 2007; He et al., 2007; Bommer et al., 2007). Furthermore, mRNAs downregulated by miRNA were primarily involved in cell-cycle control and DNA damage response. These genes had more sequences that matched miR-34 in their 3' untranslated regions. Some examples of these genes include CDK4/6, Cyclin E2, MET, and Bcl-2 (Hermeking, 2009).

Furthermore, let-7 is also a family of tumor suppressor miRNAs as they are downregulated in various types of cancer. In specific cancers, most, if not all members are downregulated (Takamizawa et al., 2004; Dahiya et al., 2008; O'Hara et al., 2009). Lower levels of let-7 family members can also indicate poorer prognosis. The combined decrease of let-7d and increase of HMGA2, a gene targeted by let-7d, is associated with poor survival in ovarian cancer (Shell et al., 2007). However, mir-98, a member of the let-7 family, can increase chemoresistance by downregulating miR-152. This in turn can repress Dicer1, an enzyme involved in processing miRNAs. High expression miR-152 as well as miR-98 can affect levels of RAD51 recombinase in cancer cells. These were linked to unfavorable outcomes in patients with ovarian cancer (Chirshhev et al., 2019; Wang et al., 2018; Boyerinas et al., 2009; Letafati et al., 2022).

Detection Challenges and Solutions

Detection Challenges

As interest in identifying circulating miRNAs in cancer increases, a significant challenge in analyzing them in bodily fluids is the need for effective internal normalization, which involves comparing miRNA expression levels to those of stable reference genes. Traditional housekeeping genes such as U6 and GAPDH, often used for normalization, are unsuitable for serum studies because they are easily degraded and undetectable in it (Chen et al., 2008). A study compared levels of U6 to different miRNAs in normal human serum. They found that U6 was unstable in serum. Additionally, the study detected miRNAs in the serum of rats, mice, calves, bovine fetuses, and horses, suggesting these miRNAs could be used for normalization in comparative studies. Another study showed that certain miRNAs (miR-15b, miR-16, miR-19b, and miR-24) in normal human serum are stable and resistant to RNase digestion, with their levels remaining unchanged during tumor growth, making them potential normalization controls (Mitchell et al., 2008). Among these, miR-16 was most commonly found and used as a control in serum miRNA studies (Chen et al., 2008). However, another study found high levels of miR-16 in red blood cells, indicating that its presence in serum might result from red blood cell hemolysis, a process in which the red blood cells are broken down (Kirschner et al., 2011). This raises concern about using miR-16 for normalization. As more miRNAs are identified in normal blood cells, there is a pressing need for standard methods of normalizing miRNA measurements in blood serum samples. Currently, using spike-in synthetic *C. elegans* miRNA for internal normalization in serum analyses addresses this challenge. Overall, circulating miRNAs that are detected through validated methods hold promise as valuable biomarkers for ovarian cancer diagnosis and prognosis (Sundarbose et al., 2013).

When looking at miRNA profiles in control samples, studies have reported variability (Orlova et al., 2011). MiRNAs can be deregulated across different disorders and types of cancer (White et al., 2010; Reid et al., 2011). Each miRNA can target multiple mRNAs, and so, there is often overlap in signaling pathways involved. While changes in miRNA expression are

frequently documented, it remains largely unclear whether these changes are causes or consequences of dysfunction (Ajit, 2012).

Detection Solutions

To enhance sensitivity of miRNA detection methods, certain amplification techniques can be used due to miRNAs' low abundance. For example, nucleic acid amplification techniques are commonly used. Various new technologies using isothermal amplification, which involves increasing the number of DNA or RNA copies at a constant temperature, have been developed. This offers greater applicability for point-of-care testing in comparison to traditional polymerase chain reaction (PCR) based methods. Nucleic acid amplification techniques include rolling circle amplification (RCA), loop-mediated isothermal amplification (LAMP), strand-displacement amplification (SDA), and several enzyme-free amplification techniques. These methods can be readily adapted for real-time assays using SYBR Green as a DNA stain.

A limitation of SYBR Green is that it can reduce amplification efficiency with increasing concentration and lead to nonspecific amplification (Gudnason et al., 2007). To avoid these problems, probes can be used to specifically detect amplified products. For instance, TaqMan probes were considered for direct validation of amplification, but they rely on Taq exonuclease activity, which not all DNA polymerases have (Chen, 2005). Other types of probes, such as molecular beacons (MBs) and one-step strand displacement (OSD) reporters can provide better performance for detecting sequence-specific nucleic acid amplicons while minimizing false-positive results (Tyagi and Kramer, 1996; Jiang et al., 2015).

Enzyme-free amplification, a DNA amplification technique that relies on strand displacement reactions initiated by a short DNA sequence, has garnered interest in nucleic acid quantification. This is due to its isothermal nature and absence of enzymes. In this process, miRNA serves as a trigger and initiates strand displacement by causing one or more DNA hairpins to open, which is driven by the stability of newly formed base pairs. This process has led to the development of enzyme-free methods such as hybridization chain reaction (HCR), catalytic hairpin assembly (CHA), and entropy-driven catalysis for miRNA detection and imaging.

HCR is a rapid amplification technique operating at room temperature, designed for short DNA sequences. It uses two distinct kinetically trapped hairpin DNA probes (H1 and H2) with partially complementary segments. Upon activation by a DNA initiator strand, these hairpins undergo a series of hybridization events, where they sequentially nucleate and open, leading to the formation of a long, nicked double-stranded amplification polymer. HCR employs these two molecules to significantly reduce background noise. The resulting HCR products show a highly ordered DNA double-helical structure, enabling the attachment of signal molecules with precisely controlled density, which enhances amplification efficiency. The amplification process in HCR does not depend on the movement of target analytes; instead, it relies on the alternating self-assembly of the hairpin probes, making it particularly well-suited for unstable targets like miRNAs.

CHA is a technique that shows promising potential for applications due to its strong signal amplification, minimal background interference, and versatility, which enables it to be combined with various signaling strategies. A study conducted using this technique showed that the concentration of miR-141 was successfully assessed using CHA and fluorescence enhancement from DNA-silver nanoclusters near guanine-rich DNA (Kim et al., 2018). However, the sensitivity of this approach was limited by inefficient hairpin assembly and low binding between the target and dye-modified probe. Another study improved this by anchoring the dye-modified probe to the surface of graphene oxide using a capture probe (Zhen et al., 2017). The dye-modified probe could only detach from graphene oxide's surface in the presence of the H1–H2 probe generated by CHA in response to the target miRNA-21. By measuring the decrease in fluorescence anisotropy, a technique that indicates how fast molecules are spinning, miRNA-21 at concentrations as low as 47 picomolar (pM) within a dynamic range of 0 to 16 nanomolar (nM) were able to be detected.

Entropy-driven catalytic reaction, unlike CHA and HCR, is driven forward thermodynamically by increases in entropy from released molecules, without changing the DNA structure. This approach is faster, more stable, and better understood than hairpin-based methods. It also minimizes background interference caused by complex DNA structures like pseudoknots, which are loops within a single strand of DNA that fold back on themselves, or kissing loops, which are two separate strands of DNA that form complementary base pairs at their ends (Zhang et al., 2007). A study adapted this entropy-driven catalytic strategy into a nanomachine machine consisting of a DNA-coated gold nanoparticle (AuNP) assembly, a walking leg, and a fuel for automated imaging of miRNAs (Liang et al., 2017). Once the walking leg is attached to the AuNPs, it forms a structure that facilitates intramolecular hybridization, initiating the entropy-driven catalytic reaction. This nanomachine effectively enabled intracellular imaging of miR-21 in four different cell lines, even those with low levels of the target miRNA.

Another technique for amplifying DNAs and RNAs is LAMP, an isothermal reaction that exhibits high sensitivity due to its exponential amplification feature. The method utilizes 4 to 6 distinct primers to target and amplify 6 to 8 different sequences at the same time, significantly enhancing its selectivity. In most LAMP-based miRNA quantification strategies, the miRNAs act as triggers, starting the amplification process (Li et al., 2010). The primers can only extend with the aid of DNA polymerase and strand displacement DNA synthesis when the target miRNA is present. In LAMP-based detection of miRNAs, designing the probes can be complex because the LAMP template includes 4 to 6 specific sequences needed for forming stem-loops. This complexity can reduce the sensitivity of the method due to the interactions between multiple primers and the lengthy template during hybridization and extension. A study introduced an idea to simplify the probe design. A stem-loop template DNA and stem-loop primer were created to replace the previously used long template. In the presence of target miRNA, double stem-loop DNA is formed, which serves as the starting material for the following LAMP process (Sun et al., 2017). Another study developed a quick and highly sensitive RCA-LAMP method for detecting miRNAs, achieving a detection limit of 10 attomoles by effectively integrating RCA

with LAMP (Tian et al., 2019). When the target miRNA is present, the RCA product, which is a long repeated DNA, serves as the template to produce numerous double stem-loop DNAs with functional sequences for the LAMP process. Additionally, another study discovered that SplintR ligase was more efficient than T4 ligase in enhancing the RCA-LAMP technique (Zhang et al., 2018).

However, a limitation of LAMP is its reliance on indirect evaluation methods, such as SYBR Green I dye, precipitation, and hydroxy naphthol blue dye, which cannot effectively differentiate between desired and nonspecific products, leading to risks of false positives. To address this issue, a study introduced a toehold-mediated strand exchange reaction called the one-step strand displacement (OSD) reporter, effectively replacing the intercalating dye and successfully differentiating true amplicons from side products (Jiang et al., 2015).

To achieve optimal detection performance, researchers frequently combine multiple amplification methods, which offers three primary advantages. First, this approach enhances the amplification of trace target molecules prior to analysis, improving the method's sensitivity. For example, a sensitive electrochemical biosensor for miRNA was created by integrating HCR with enzyme-based signal amplification, using magnetic microcarriers for sandwich-type hybridization (Torrente-Rodríguez et al., 2016). Second, this strategy eliminates the need for traditional labels or enzymes, simplifying the process and reducing costs. A label-free method for detecting miRNA was developed by merging SDA with RCA (Zheng et al., 2016). Finally, incorporating electrochemical sensors and nanomaterial-based biosensors effectively minimizes background noise. Various detection strategies for miRNAs have been explored by linking HCR to positively charged gold nanoparticles or silver nanoparticles (Miao et al., 2016; Miao et al., 2016).

Combining biomarkers may also be an effective way to address these challenges. In a study, Receiver Operating Characteristic (ROC) was used to assess the sensitivity and specificity of different biomarkers for diagnosing ovarian cancer. The results showed the miR-182 and CA-125 had the highest sensitivity, but low specificity. Furthermore, miR-205 levels were significantly different between ovarian cancer patients and those with benign ovarian diseases. This biomarker was not suitable as a diagnostic marker due to its low sensitivity and specificity. HE4 and miR-183 had the highest sensitivity but no single marker was sufficient for effective ovarian cancer diagnosis. A diagnostic model for ovarian cancer was developed using binary logistic regression, incorporating CA-125, HE4, and miR-183. This model showed high sensitivity and specificity for diagnosing ovarian cancer. This suggests that using additional character markers and clinical characteristics together can increase the sensitivity and specificity of miR-183. Overall, ovarian cancer patients show elevated serum levels of miR-182, miR-183, and miR-205, with miR-183 providing relatively high sensitivity and specificity. The combined application of miR-183 with other markers, like HE4, may aid in the early detection of ovarian cancer and improve the assessment of disease progression (Ye et al., 2019).

Discussion

This study aimed to explore the potential of miRNAs as biomarkers for ovarian cancer, highlighting their advantages, limitations, and future research directions. The findings of this review suggest that miRNAs hold significant promise as early diagnostic markers for ovarian cancer due to their stability in bodily fluids, association with ovarian cancer progression, and potential for non-invasive testing. Several studies have demonstrated abnormal levels of specific miRNAs in ovarian cancer patients compared to healthy individuals, indicating their potential as early diagnostic markers. Furthermore, miRNAs have been shown to be involved in various biological processes associated with ovarian cancer, such as tumorigenesis, cell proliferation, differentiation, apoptosis, and metastasis.

While the overall findings of this review support the potential of miRNAs as biomarkers for ovarian cancer, there are some inconsistencies. For example, different studies have reported overlapping expression of miRNAs in different cancers, which can make it difficult to distinguish between ovarian cancer and other malignancies. Furthermore, the mechanisms by which miRNAs contribute to ovarian cancer development are not fully understood, limiting their ability to interpret the clinical significance. Additionally, the potential of miRNAs as biomarkers for cancer development and progression was based on studies primarily focusing on lung and pancreatic cancer, not ovarian cancer.

To address these challenges, future studies should aim to develop more accurate and sensitive detection methods. Studies investigating the combined use of miRNA biomarkers with other diagnostic tests, such as CA-125 and imaging techniques, may improve the accuracy and specificity of early diagnosis. Research elucidating the underlying mechanisms by which miRNAs contribute to ovarian cancer development should also be carried out. Further research is also needed to explore the specific role of miRNA by analyzing their expression profiles at different stages of ovarian cancer.

The potential of miRNAs as biomarkers for ovarian cancer has significant practical implications. Ovarian cancer, being a highly lethal gynecologic malignancy, is often diagnosed at an advanced stage due to subtle symptoms and a lack of effective early screening methods. This late diagnosis significantly contributes to poor survival rates, creating a critical need for innovative strategies to enhance early detection and improve patient outcomes. The development of miRNA-based diagnostic tests could enable earlier detection of the disease, leading to improved patient outcomes and survival rates. Furthermore, miRNAs could be used to monitor disease progression, guide treatment decisions, and assess the effectiveness of therapeutic interventions.

Overall, while the use of miRNA biomarkers for the early diagnosis of ovarian cancer shows significant promise, further research is necessary to fully realize their potential. While challenges remain, the advantages of miRNA-based diagnostics, including their stability, sensitivity, and non-invasive nature, warrant further investigation and development. By addressing the limitations and exploring new approaches, researchers can contribute to improving the early detection and treatment of ovarian cancer.

Conclusion

Ovarian cancer survival rates remain low due to late-stage diagnosis. Identifying reliable biomarkers for early diagnosis and monitoring is important for improving patient outcomes. Traditional ovarian biomarkers like CA125 and HE4 have limitations, such as their low sensitivity or specificity. MiRNAs offer several advantages, showing promise as sensitive and specific markers that are also stable, making them suitable for non-invasive testing. They also play a diverse role in ovarian cancer, acting as up and down regulators in gene expression. While challenges remain in translating miRNA research into clinical practice, they offer hope for improving the diagnosis, prognosis, and treatment of ovarian cancer. Further research is needed to validate these findings, with emphasis on developing standardized detection methods for clinical use, especially those able to detect subtle changes in miRNA expression.

Abbreviations

miRNA: microRNA, CA125: carbohydrate antigen 125, HE4: human epididymis protein 4, HCR: hybridization chain reaction, CHA: catalytic hairpin assembly, LAMP: loop-mediated isothermal amplification, RCA: rolling circle amplification, SDA: strand-displacement amplification

Works Cited

- Arora, T., et al. "Epithelial Ovarian Cancer." StatPearls - NCBI Bookshelf, 6 May 2024, <https://www.ncbi.nlm.nih.gov/books/NBK567760/>.
- Hennessy, B. T., Coleman, R. L., and Markman, M. "Ovarian Cancer." *The Lancet*, vol. 374, no. 9698, 2009, pp. 1371-1382. [https://doi.org/10.1016/S0140-6736\(09\)61338-6](https://doi.org/10.1016/S0140-6736(09)61338-6).
- Howlader, N., et al., editors. *SEER Cancer Statistics Review, 1975–2017*. https://seer.cancer.gov/archive/csr/1975_2017/index.html.
- Zeppernick, F., and Meinhold-Heerlein, I. "The New FIGO Staging System for Ovarian, Fallopian Tube, and Primary Peritoneal Cancer." *Archives of Gynecology and Obstetrics*, vol. 290, 2014, pp. 839-842. <https://doi.org/10.1007/s00404-014-3364-8>.
- Schiavone, M. B., et al. "Natural History and Outcome of Mucinous Carcinoma of the Ovary." *American Journal of Obstetrics and Gynecology*, vol. 205, no. 5, 2011, pp. 480-e1. <https://doi.org/10.1016/j.ajog.2011.06.049>.
- Cress, R. D., et al. "Characteristics of Long-Term Survivors of Epithelial Ovarian Cancer." *Obstetrics & Gynecology*, vol. 126, no. 3, 2015, pp. 491-497. <https://doi.org/10.1097/AOG.0000000000000981>.
- Charkhchi, P., et al. "CA125 and Ovarian Cancer: A Comprehensive Review." *Cancers*, vol. 12, no. 12, 2020, p. 3730. <https://doi.org/10.3390/cancers12123730>.
- Zurawski, V. R., et al. "Elevated Serum CA 125 Levels Prior to Diagnosis of Ovarian Neoplasia: Relevance for Early Detection of Ovarian Cancer." *International Journal of Cancer*, vol. 42, no. 5, 1988, pp. 677–680. <https://doi.org/10.1002/ijc.2910420507>.
- Bast, R. C., et al. "A Radioimmunoassay Using a Monoclonal Antibody to Monitor the Course of Epithelial Ovarian Cancer." *New England Journal of Medicine*, vol. 309, no. 15, 1983, pp. 883–887. <https://doi.org/10.1056/nejm198310133091503>.
- Bast, R. C., et al. "Reactivity of a Monoclonal Antibody with Human Ovarian Carcinoma." *Journal of Clinical Investigation*, vol. 68, no. 5, 1981, pp. 1331–1337. <https://doi.org/10.1172/jci110380>.
- Timmerman, D., et al. "Inclusion of CA-125 Does Not Improve Mathematical Models Developed to Distinguish Between Benign and Malignant Adnexal Tumors." *Journal of Clinical Oncology*, vol. 25, no. 27, 2007, pp. 4194–4200. <https://doi.org/10.1200/jco.2006.09.5943>.
- Ruggeri, G., et al. "HE4 and Epithelial Ovarian Cancer: Comparison and Clinical Evaluation of Two Immunoassays and a Combination Algorithm." *Clinica Chimica Acta*, vol. 412, no. 15-16, 2011, pp. 1447-1453. <https://doi.org/10.1016/j.cca.2011.04.028>.
- Drapkin, R., et al. "Human Epididymis Protein 4 (HE4) Is a Secreted Glycoprotein That Is Overexpressed by Serous and Endometrioid Ovarian Carcinomas." *Cancer Research*, vol. 65, no. 6, 2005, pp. 2162–2169. <https://doi.org/10.1158/0008-5472.can-04-3924>.
- Rieber, A., et al. "Preoperative Diagnosis of Ovarian Tumors with MR Imaging." *American Journal of Roentgenology*, vol. 177, no. 1, 2001, pp. 123–129. <https://doi.org/10.2214/ajr.177.1.1770123>.

- Grab, D., et al. "Classification of Asymptomatic Adnexal Masses by Ultrasound, Magnetic Resonance Imaging, and Positron Emission Tomography." *Gynecologic Oncology*, vol. 77, no. 3, 2000, pp. 454-459. <https://doi.org/10.1006/gyno.2000.5768>.
- Castellucci, P., et al. "Diagnostic Accuracy of 18F-FDG PET/CT in Characterizing Ovarian Lesions and Staging Ovarian Cancer: Correlation with Transvaginal Ultrasonography, Computed Tomography, and Histology." *Nuclear Medicine Communications*, vol. 28, no. 8, 2007, pp. 589–595. <https://doi.org/10.1097/mnm.0b013e3281afa256>.
- Nam, E. J., et al. "Diagnosis and Staging of Primary Ovarian Cancer: Correlation Between PET/CT, Doppler US, and CT or MRI." *Gynecologic Oncology*, vol. 116, no. 3, 2010, pp. 389-394. <https://doi.org/10.1016/j.ygyno.2009.10.059>.
- Liberto, J. M., et al. "Current and Emerging Methods for Ovarian Cancer Screening and Diagnostics: A Comprehensive Review." *Cancers*, vol. 14, no. 12, 2022, p. 2885. <https://doi.org/10.3390/cancers14122885>.
- Yokoi, A., et al. "Integrated Extracellular MicroRNA Profiling for Ovarian Cancer Screening." *Nature Communications*, vol. 9, no. 1, 2018. <https://doi.org/10.1038/s41467-018-06434-4>.
- Lu, T. X., and Rothenberg, M. E. "MicroRNA." *Journal of Allergy and Clinical Immunology*, vol. 141, no. 4, 2018, pp. 1202-1207. <https://doi.org/10.1016/j.jaci.2017.08.034>.
- Zhang, H., and Lu, B. "MicroRNAs as Biomarkers of Ovarian Cancer." *Expert Review of Anticancer Therapy*, vol. 20, no. 5, 2020, pp. 373–385. <https://doi.org/10.1080/14737140.2020.1760095>.
- Zhao, L., et al. "The Role of miRNA in Ovarian Cancer: An Overview." *Reproductive Sciences*, vol. 29, no. 10, 2022, pp. 2760–2767. <https://doi.org/10.1007/s43032-021-00717-w>.
- Carthew, R. W., and Sontheimer, E. J. "Origins and Mechanisms of miRNAs and siRNAs." *Cell*, vol. 136, no. 4, 2009, pp. 642-655. <https://doi.org/10.1016/j.cell.2009.01.035>.
- Ghildiyal, M., and Zamore, P. D. "Small Silencing RNAs: An Expanding Universe." *Nature Reviews Genetics*, vol. 10, no. 2, 2009, pp. 94–108. <https://doi.org/10.1038/nrg2504>.
- Kim, V. N., et al. "Biogenesis of Small RNAs in Animals." *Nature Reviews Molecular Cell Biology*, vol. 10, no. 2, 2009, pp. 126–139. <https://doi.org/10.1038/nrm2632>.
- Siomi, H., and Siomi, M. C. "On the Road to Reading the RNA-Interference Code." *Nature*, vol. 457, no. 7228, 2009, pp. 396–404. <https://doi.org/10.1038/nature07754>.
- Lee, R. C., et al. "The *C. elegans* Heterochronic Gene *lin-4* Encodes Small RNAs with Antisense Complementarity to *lin-14*." *Cell*, vol. 75, no. 5, 1993, pp. 843-854. [https://doi.org/10.1016/0092-8674\(93\)90529-y](https://doi.org/10.1016/0092-8674(93)90529-y).
- Oliveira-Carvalho, V., et al. "MicroRNAs: Um Novo Paradigma no Tratamento e Diagnóstico da Insuficiência Cardíaca?" *Arquivos Brasileiros de Cardiologia*, vol. 98, no. 4, 2012, pp. 362–370. <https://doi.org/10.1590/s0066-782x2012000400011>.
- Ambros, V. "The Functions of Animal MicroRNAs." *Nature*, vol. 431, no. 7006, 2004, pp. 350–355. <https://doi.org/10.1038/nature02871>.

- Orang, A. V., et al. "Mechanisms of miRNA-Mediated Gene Regulation from Common Downregulation to mRNA-Specific Upregulation." *International Journal of Genomics*, 2014, pp. 1–15. <https://doi.org/10.1155/2014/970607>.
- Denli, A. M., et al. "Processing of Primary MicroRNAs by the Microprocessor Complex." *Nature*, vol. 432, no. 7014, 2004, pp. 231–235. <https://doi.org/10.1038/nature03049>.
- Lee, Y., et al. "The Nuclear RNase III Drosha Initiates MicroRNA Processing." *Nature*, vol. 425, no. 6956, 2003, pp. 415–419. <https://doi.org/10.1038/nature01957>.
- Bohnsack, M. T., et al. "Exportin 5 Is a RanGTP-Dependent dsRNA-Binding Protein That Mediates Nuclear Export of Pre-miRNAs." *RNA*, vol. 10, no. 2, 2004, pp. 185–191. <https://doi.org/10.1261/rna.5167604>.
- Bartel, D. P. "MicroRNAs: Genomics, Biogenesis, Mechanism, and Function." *Cell*, vol. 116, no. 2, 2004, pp. 281–297. [https://doi.org/10.1016/S0092-8674\(04\)00045-5](https://doi.org/10.1016/S0092-8674(04)00045-5).
- Reddy, K. B. "MicroRNA (miRNA) in Cancer." *Cancer Cell International*, vol. 15, no. 1, 2015. <https://doi.org/10.1186/s12935-015-0185-1>.
- Achar, Y. J., and Foiani, M. "Coordinating Replication with Transcription." *Advances in Experimental Medicine and Biology*, 2017, pp. 455–487. https://doi.org/10.1007/978-981-10-6955-0_20.
- Mercadante, A. A., et al. "Biochemistry, Replication and Transcription." *StatPearls - NCBI Bookshelf*, 14 Aug. 2023. <https://www.ncbi.nlm.nih.gov/books/NBK540152/#>.
- Huntzinger, E., and Izaurralde, E. "Gene Silencing by MicroRNAs: Contributions of Translational Repression and mRNA Decay." *Nature Reviews Genetics*, vol. 12, no. 2, 2011, pp. 99–110. <https://doi.org/10.1038/nrg2936>.
- Ipsaro, J. J., and Joshua-Tor, L. "From Guide to Target: Molecular Insights into Eukaryotic RNA-Interference Machinery." *Nature Structural & Molecular Biology*, vol. 22, no. 1, 2015, pp. 20–28. <https://doi.org/10.1038/nsmb.2931>.
- Xu, W., et al. "Identifying MicroRNA Targets in Different Gene Regions." *BMC Bioinformatics*, vol. 15, no. S7, 2014. <https://doi.org/10.1186/1471-2105-15-s7-s4>.
- Forman, J. J., Legesse-Miller, A., and Collier, H. A. "A Search for Conserved Sequences in Coding Regions Reveals That the let-7 MicroRNA Targets Dicer Within Its Coding Sequence." *Proceedings of the National Academy of Sciences*, vol. 105, no. 39, 2008, pp. 14879–14884. <https://doi.org/10.1073/pnas.0803230105>.
- Zhang, J., et al. "Oncogenic Role of MicroRNA-532-5p in Human Colorectal Cancer via Targeting of the 5'UTR of RUNX3." *Oncology Letters*, 2018. <https://doi.org/10.3892/ol.2018.8217>.
- Dharap, A., et al. "MicroRNA MIR-324-3P Induces Promoter-Mediated Expression of RELA Gene." *PLoS ONE*, vol. 8, no. 11, 2013, e79467. <https://doi.org/10.1371/journal.pone.0079467>.
- O'Brien, J., et al. "Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation." *Frontiers in Endocrinology*, vol. 9, 2018, p. 402. <https://doi.org/10.3389/fendo.2018.00402>.

- Bruce, J. P., et al. "Identification of a MicroRNA Signature Associated with Risk of Distant Metastasis in Nasopharyngeal Carcinoma." *Oncotarget*, vol. 6, no. 6, 2015, pp. 4537–4550. <https://doi.org/10.18632/oncotarget.3005>.
- Esquela-Kerscher, A., and Slack, F. J. "Oncomirs — MicroRNAs with a Role in Cancer." *Nature Reviews Cancer*, vol. 6, no. 4, 2006, pp. 259–269. <https://doi.org/10.1038/nrc1840>.
- Jeansonne, D., et al. "Anti-Tumoral Effects of miR-3189-3p in Glioblastoma." *Journal of Biological Chemistry*, vol. 290, no. 13, 2015, pp. 8067-8080. <https://doi.org/10.1074/jbc.M114.633081>.
- Pinatel, E. M., et al. "MIR-223 Is a Coordinator of Breast Cancer Progression as Revealed by Bioinformatics Predictions." *PLoS ONE*, vol. 9, no. 1, 2014, e84859. <https://doi.org/10.1371/journal.pone.0084859>.
- Ben-Hamo, R., and Efroni, S. "MicroRNA Regulation of Molecular Pathways as a Generic Mechanism and as a Core Disease Phenotype." *Oncotarget*, vol. 6, no. 3, 2015, pp. 1594–1604. <https://doi.org/10.18632/oncotarget.2734>.
- Sotiropoulou, G., et al. "Emerging Roles of MicroRNAs as Molecular Switches in the Integrated Circuit of the Cancer Cell." *RNA*, vol. 15, no. 8, 2009, pp. 1443–1461. <https://doi.org/10.1261/rna.1534709>.
- Reddy, K. B. "MicroRNA (miRNA) in Cancer." *Cancer Cell International*, vol. 15, no. 1, 2015, <https://doi.org/10.1186/s12935-015-0185-1>.
- Cheng, A. M. "Antisense Inhibition of Human miRNAs and Indications for an Involvement of miRNA in Cell Growth and Apoptosis." *Nucleic Acids Research*, vol. 33, no. 4, 2005, pp. 1290–1297. <https://doi.org/10.1093/nar/gki200>.
- O'Brien, J., et al. "Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation." *Frontiers in Endocrinology*, vol. 9, 2018, p. 402. <https://doi.org/10.3389/fendo.2018.00402>.
- Zhang, B., et al. "MicroRNAs as Oncogenes and Tumor Suppressors." *Developmental Biology*, vol. 302, no. 1, 2007, pp. 1–12. <https://doi.org/10.1016/j.ydbio.2006.08.028>.
- Wang, H., et al. "Circulating MicroRNAs as Potential Cancer Biomarkers: The Advantage and Disadvantage." *Clinical Epigenetics*, vol. 10, no. 1, 2018. <https://doi.org/10.1186/s13148-018-0492-1>.
- Weber, J. A., et al. "The MicroRNA Spectrum in 12 Body Fluids." *Clinical Chemistry*, vol. 56, no. 11, 2010, pp. 1733–1741. <https://doi.org/10.1373/clinchem.2010.147405>.
- Zhang, W. C., et al. "Tumour-Initiating Cell-Specific miR-1246 and miR-1290 Expression Converge to Promote Non-Small Cell Lung Cancer Progression." *Nature Communications*, vol. 7, no. 1, 2016. <https://doi.org/10.1038/ncomms11702>.
- Henry, N. L., and Hayes, D. F. "Cancer Biomarkers." *Molecular Oncology*, vol. 6, no. 2, 2012, pp. 140-146. <https://doi.org/10.1016/j.molonc.2012.01.010>.

- Fehlmann, T., et al. "Large-Scale Validation of miRNAs by Disease Association, Evolutionary Conservation and Pathway Activity." *RNA Biology*, vol. 16, no. 1, 2018, pp. 93–103. <https://doi.org/10.1080/15476286.2018.1559689>.
- Fehlmann, T., et al. "Common Diseases Alter the Physiological Age-Related Blood MicroRNA Profile." *Nature Communications*, vol. 11, no. 1, 2020. <https://doi.org/10.1038/s41467-020-19665-1>.
- Bagnoli, M., Pignata, S., and Mezzanzanica, D. "A miRNA Signature Assessing Ovarian Cancer Prognosis." *Oncoscience*, vol. 3, no. 11-12, 2016, pp. 308–310. <https://doi.org/10.18632/oncoscience.329>.
- Stawiski, K., et al. "Validation of miRNA Signatures for Ovarian Cancer Earlier Detection in the Pre-Diagnosis Setting Using Machine Learning Approaches." *Frontiers in Oncology*, vol. 14, 2024. <https://doi.org/10.3389/fonc.2024.1389066>.
- Wang, X., et al. "miR-195 Inhibits the Growth and Metastasis of NSCLC Cells by Targeting IGF1R." *Tumor Biology*, vol. 35, no. 9, 2014, pp. 8765–8770. <https://doi.org/10.1007/s13277-014-2140-5>.
- Guo, H., et al. "miR-195 Targets HDGF to Inhibit Proliferation and Invasion of NSCLC Cells." *Tumor Biology*, vol. 35, no. 9, 2014, pp. 8861–8866. <https://doi.org/10.1007/s13277-014-2153-0>.
- Liu, B., et al. "MiR-195 Suppresses Non-Small Cell Lung Cancer by Targeting CHEK1." *Oncotarget*, vol. 6, no. 11, 2015, p. 9445. <https://doi.org/10.18632/oncotarget.3255>.
- Hua, Y., et al. "Low Serum miR-373 Predicts Poor Prognosis in Patients with Pancreatic Cancer." *Cancer Biomarkers*, vol. 20, no. 1, 2017, pp. 95–100. <https://doi.org/10.3233/cbm-170231>.
- Cheng, L., et al. "Exosomes Provide a Protective and Enriched Source of miRNA for Biomarker Profiling Compared to Intracellular and Cell-Free Blood." *Journal of Extracellular Vesicles*, vol. 3, no. 1, 2014. <https://doi.org/10.3402/jev.v3.23743>.
- Muralidharan-Chari, V., et al. "Microvesicles: Mediators of Extracellular Communication During Cancer Progression." *Journal of Cell Science*, vol. 123, no. 10, 2010, pp. 1603–1611. <https://doi.org/10.1242/jcs.064386>.
- Zernecke, A., et al. "Delivery of MicroRNA-126 by Apoptotic Bodies Induces CXCL12-Dependent Vascular Protection." *Science Signaling*, vol. 2, no. 100, 2009. <https://doi.org/10.1126/scisignal.2000610>.
- Turchinovich, A., et al. "Characterization of Extracellular Circulating MicroRNA." *Nucleic Acids Research*, vol. 39, no. 16, 2011, pp. 7223–7233. <https://doi.org/10.1093/nar/gkr254>.
- Xi, Y., et al. "Systematic Analysis of MicroRNA Expression of RNA Extracted from Fresh Frozen and Formalin-Fixed Paraffin-Embedded Samples." *RNA*, vol. 13, no. 10, 2007, pp. 1668–1674. <https://doi.org/10.1261/rna.642907>.
- Leite, K. R., et al. "miRNA Analysis of Prostate Cancer by Quantitative Real Time PCR: Comparison Between Formalin-Fixed Paraffin Embedded and Fresh-Frozen Tissue."

- Urologic Oncology: Seminars and Original Investigations, vol. 29, no. 5, Sept. 2011, pp. 533-537. Elsevier. <https://doi.org/10.1016/j.urolonc.2009.05.008>.
- Chen, X., et al. "Characterization of MicroRNAs in Serum: A Novel Class of Biomarkers for Diagnosis of Cancer and Other Diseases." *Cell Research*, vol. 18, no. 10, 2008, pp. 997–1006. <https://doi.org/10.1038/cr.2008.282>.
- Lan, H., et al. "MicroRNAs as Potential Biomarkers in Cancer: Opportunities and Challenges." *BioMed Research International*, 2015, pp. 1–17. <https://doi.org/10.1155/2015/125094>.
- Selcuklu, S. D., et al. "miR-21 as a Key Regulator of Oncogenic Processes." *Biochemical Society Transactions*, vol. 37, no. 4, Aug. 2009, pp. 918–925. <https://doi.org/10.1042/BST0370918>.
- Kartika, A. I., et al. "MicroRNA-21 as a Biomarker for Ovarian Cancer Detection." *Indonesian Journal of Biotechnology*, vol. 23, no. 1, 2018, p. 35. <https://doi.org/10.22146/ijbiotech.35692>.
- Lou, Y., et al. "MicroRNA-21 Promotes the Cell Proliferation, Invasion and Migration Abilities in Ovarian Epithelial Carcinomas Through Inhibiting the Expression of PTEN Protein." *International Journal of Molecular Medicine*, vol. 26, no. 6, 2010. https://doi.org/10.3892/ijmm_00000530.
- Meng, F., et al. "MicroRNA-21 Regulates Expression of the PTEN Tumor Suppressor Gene in Human Hepatocellular Cancer." *Gastroenterology*, vol. 133, no. 2, 2007, pp. 647–658. <https://doi.org/10.1053/j.gastro.2007.05.022>.
- Asangani, I. A., et al. "MicroRNA-21 (miR-21) Post-Transcriptionally Downregulates Tumor Suppressor Pcd4 and Stimulates Invasion, Intravasation and Metastasis in Colorectal Cancer." *Oncogene*, vol. 27, no. 15, 2007, pp. 2128–2136. <https://doi.org/10.1038/sj.onc.1210856>.
- Chan, J. K., et al. "The Inhibition of miR-21 Promotes Apoptosis and Chemosensitivity in Ovarian Cancer." *Gynecologic Oncology*, vol. 132, no. 3, 2014, pp. 739-744. <https://doi.org/10.1016/j.ygyno.2014.01.034>.
- Nam, E. J., et al. "MicroRNA Expression Profiles in Serous Ovarian Carcinoma." *Clinical Cancer Research*, vol. 14, no. 9, 2008, pp. 2690–2695. <https://doi.org/10.1158/1078-0432.ccr-07-1731>.
- Paliwal, N., et al. "Evaluation of miR-22 and miR-21 as Diagnostic Biomarkers in Patients with Epithelial Ovarian Cancer." *3 Biotech*, vol. 10, no. 3, 2020. <https://doi.org/10.1007/s13205-020-2124-7>.
- Talaat, A., et al. "Evaluation of miRNA-21 and CA-125 as a Promising Diagnostic Biomarker in Patients with Ovarian Cancer." *Egyptian Journal of Medical Human Genetics*, vol. 23, no. 1, 2022. <https://doi.org/10.1186/s43042-022-00342-5>.
- Masoumi-Dehghi, S., et al. "microRNA-141-3p-Containing Small Extracellular Vesicles Derived from Epithelial Ovarian Cancer Cells Promote Endothelial Cell Angiogenesis Through Activating the JAK/STAT3 and NF-κB Signaling Pathways." *Journal of Cell*

- Communication and Signaling, vol. 14, no. 2, 2020, pp. 233–244.
<https://doi.org/10.1007/s12079-020-00548-5>.
- Klicka, K., et al. "The Role of miR-200 Family in the Regulation of Hallmarks of Cancer." *Frontiers in Oncology*, vol. 12, 2022. <https://doi.org/10.3389/fonc.2022.965231>.
- He, L., et al. "A MicroRNA Component of the p53 Tumour Suppressor Network." *Nature*, vol. 447, no. 7148, 2007, pp. 1130–1134. <https://doi.org/10.1038/nature05939>.
- Tarasov, V., et al. "Differential Regulation of MicroRNAs by p53 Revealed by Massively Parallel Sequencing: miR-34a is a p53 Target That Induces Apoptosis and G1-Arrest." *Cell Cycle*, vol. 6, no. 13, 2007, pp. 1586–1593. <https://doi.org/10.4161/cc.6.13.4436>.
- Bommer, G. T., et al. "p53-Mediated Activation of miRNA-34 Candidate Tumor-Suppressor Genes." *Current Biology*, vol. 17, no. 15, 2007, pp. 1298–1307.
<https://doi.org/10.1016/j.cub.2007.06.068>.
- Collado, M., et al. "Cellular Senescence in Cancer and Aging." *Cell*, vol. 130, no. 2, 2007, pp. 223–233. <https://doi.org/10.1016/j.cell.2007.07.003>.
- Chang, T. C., et al. "Transactivation of miR-34a by p53 Broadly Influences Gene Expression and Promotes Apoptosis." *Molecular Cell*, vol. 26, no. 5, 2007, pp. 745–752.
<https://doi.org/10.1016/j.molcel.2007.05.010>.
- Raver-Shapira, N., et al. "Transcriptional Activation of miR-34a Contributes to p53-Mediated Apoptosis." *Molecular Cell*, vol. 26, no. 5, 2007, pp. 731–743.
<https://doi.org/10.1016/j.molcel.2007.05.017>.
- Welch, C., et al. "MicroRNA-34a Functions as a Potential Tumor Suppressor by Inducing Apoptosis in Neuroblastoma Cells." *Oncogene*, vol. 26, no. 34, 2007, pp. 5017–5022.
<https://doi.org/10.1038/sj.onc.1210293>.
- Hermeking, H. "The miR-34 Family in Cancer and Apoptosis." *Cell Death and Differentiation*, vol. 17, no. 2, 2009, pp. 193–199. <https://doi.org/10.1038/cdd.2009.56>.
- Takamizawa, J., et al. "Reduced Expression of the let-7 MicroRNAs in Human Lung Cancers in Association with Shortened Postoperative Survival." *Cancer Research*, vol. 64, no. 11, 2004, pp. 3753–3756. <https://doi.org/10.1158/0008-5472.can-04-0637>.
- Dahiya, N., et al. "MicroRNA Expression and Identification of Putative miRNA Targets in Ovarian Cancer." *PLoS ONE*, vol. 3, no. 6, 2008, e2436.
<https://doi.org/10.1371/journal.pone.0002436>.
- O'Hara, A. J., et al. "Tumor Suppressor MicroRNAs Are Underrepresented in Primary Effusion Lymphoma and Kaposi Sarcoma." *Blood*, vol. 113, no. 23, 2009, pp. 5938–5941. <https://doi.org/10.1182/blood-2008-09-179168>.
- Shell, S., et al. "Let-7 Expression Defines Two Differentiation Stages of Cancer." *Proceedings of the National Academy of Sciences*, vol. 104, no. 27, 2007, pp. 11400–11405. <https://doi.org/10.1073/pnas.0704372104>.
- Chirshev, E., et al. "Let-7 as Biomarker, Prognostic Indicator, and Therapy for Precision Medicine in Cancer." *Clinical and Translational Medicine*, vol. 8, no. 1, 2019.
<https://doi.org/10.1186/s40169-019-0240-y>.

- Wang, Y., et al. "miR-98-5p Contributes to Cisplatin Resistance in Epithelial Ovarian Cancer by Suppressing miR-152 Biogenesis via Targeting Dicer1." *Cell Death and Disease*, vol. 9, no. 5, 2018. <https://doi.org/10.1038/s41419-018-0390-7>.
- Boyerinas, B., et al. "The Role of Let-7 in Cell Differentiation and Cancer." *Endocrine Related Cancer*, vol. 17, no. 1, 2009, F19–F36. <https://doi.org/10.1677/erc-09-0184>.
- Letafati, A., et al. "MicroRNA let-7 and Viral Infections: Focus on Mechanisms of Action." *Cellular & Molecular Biology Letters*, vol. 27, no. 1, 2022. <https://doi.org/10.1186/s11658-022-00317-9>.
- Mitchell, P. S., et al. "Circulating MicroRNAs as Stable Blood-Based Markers for Cancer Detection." *Proceedings of the National Academy of Sciences*, vol. 105, no. 30, 2008, pp. 10513–10518. <https://doi.org/10.1073/pnas.0804549105>.
- Kirschner, M. B., et al. "Haemolysis During Sample Preparation Alters MicroRNA Content of Plasma." *PLoS ONE*, vol. 6, no. 9, 2011, e24145. <https://doi.org/10.1371/journal.pone.0024145>.
- Sundarbose, K., et al. "MicroRNAs as Biomarkers in Cancer." *Diagnostics*, vol. 3, no. 1, 2013, pp. 84–104. <https://doi.org/10.3390/diagnostics3010084>.
- Orlova, I. A., et al. "MicroRNA Modulation in Complex Regional Pain Syndrome." *Journal of Translational Medicine*, vol. 9, no. 1, 2011. <https://doi.org/10.1186/1479-5876-9-195>.
- White, N. M. A., et al. "Metastamirs: A Stepping Stone Towards Improved Cancer Management." *Nature Reviews Clinical Oncology*, vol. 8, no. 2, 2010, pp. 75–84. <https://doi.org/10.1038/nrclinonc.2010.173>.
- Reid, G., et al. "Circulating MicroRNAs: Association with Disease and Potential Use as Biomarkers." *Critical Reviews in Oncology/Hematology*, vol. 80, no. 2, 2011, pp. 193–208. <https://doi.org/10.1016/j.critrevonc.2010.11.004>.
- Ajit, S. K. "Circulating MicroRNAs as Biomarkers, Therapeutic Targets, and Signaling Molecules." *Sensors*, vol. 12, no. 3, 2012, pp. 3359–3369. <https://doi.org/10.3390/s120303359>.
- Gudnason, H., et al. "Comparison of Multiple DNA Dyes for Real-Time PCR: Effects of Dye Concentration and Sequence Composition on DNA Amplification and Melting Temperature." *Nucleic Acids Research*, vol. 35, no. 19, 2007, e127. <https://doi.org/10.1093/nar/gkm671>.
- Chen, C. "Real-Time Quantification of MicroRNAs by Stem-Loop RT-PCR." *Nucleic Acids Research*, vol. 33, no. 20, 2005, e179. <https://doi.org/10.1093/nar/gni178>.
- Tyagi, S., and F. R. Kramer. "Molecular Beacons: Probes That Fluoresce upon Hybridization." *Nature Biotechnology*, vol. 14, no. 3, 1996, pp. 303–308. <https://doi.org/10.1038/nbt0396-303>.
- Jiang, Y. S., et al. "Robust Strand Exchange Reactions for the Sequence-Specific, Real-Time Detection of Nucleic Acid Amplicons." *Analytical Chemistry*, vol. 87, no. 6, 2015, pp. 3314–3320. <https://doi.org/10.1021/ac504387c>.
- Kim, H., et al. "Enzyme-Free and Label-Free miRNA Detection Based on Target-Triggered Catalytic Hairpin Assembly and Fluorescence Enhancement of DNA-Silver

- Nanoclusters." *Sensors and Actuators B: Chemical*, vol. 260, 2018, pp. 140–145.
<https://doi.org/10.1016/j.snb.2017.12.137>.
- Zhen, S. J., et al. "An Enzyme-Free DNA Circuit-Assisted Graphene Oxide Enhanced Fluorescence Anisotropy Assay for MicroRNA Detection with Improved Sensitivity and Selectivity." *Analytical Chemistry*, vol. 89, no. 17, 2017, pp. 8766–8771.
<https://doi.org/10.1021/acs.analchem.7b00955>.
- Zhang, D. Y., et al. "Engineering Entropy-Driven Reactions and Networks Catalyzed by DNA." *Science*, vol. 318, no. 5853, 2007, pp. 1121–1125.
<https://doi.org/10.1126/science.1148532>.
- Liang, C., et al. "Rational Engineering of a Dynamic, Entropy-Driven DNA Nanomachine for Intracellular MicroRNA Imaging." *Angewandte Chemie International Edition*, vol. 56, no. 31, 2017, pp. 9077–9081. <https://doi.org/10.1002/anie.201704147>.
- Li, C., et al. "One-Step Ultrasensitive Detection of MicroRNAs with Loop-Mediated Isothermal Amplification (LAMP)." *Chemical Communications*, vol. 47, no. 9, 2010, pp. 2595–2597. <https://doi.org/10.1039/c0cc03957h>.
- Sun, Y., et al. "One-Step Detection of MicroRNA with High Sensitivity and Specificity via Target-Triggered Loop-Mediated Isothermal Amplification (TT-LAMP)." *Chemical Communications*, vol. 53, no. 80, 2017, pp. 11040–11043.
<https://doi.org/10.1039/c7cc06140d>.
- Tian, W., et al. "Rolling Circle Extension-Actuated Loop-Mediated Isothermal Amplification (RCA-LAMP) for Ultrasensitive Detection of MicroRNAs." *Biosensors and Bioelectronics*, vol. 128, 2019, pp. 17–22. <https://doi.org/10.1016/j.bios.2018.12.041>.
- Zhang, X., et al. "Ligation-Promoted Hyperbranched Rolling Circle Amplification Enables Ultrasensitive Detection of MicroRNA in Clinical Specimens." *Sensors and Actuators B: Chemical*, vol. 277, 2018, pp. 634–639. <https://doi.org/10.1016/j.snb.2018.09.058>.
- Torrente-Rodríguez, R. M., et al. "Sensitive Electrochemical Determination of miRNAs Based on a Sandwich Assay onto Magnetic Microcarriers and Hybridization Chain Reaction Amplification." *Biosensors and Bioelectronics*, vol. 86, 2016, pp. 516–521.
<https://doi.org/10.1016/j.bios.2016.07.003>.
- Zheng, X., et al. "Label-Free Detection of MicroRNA Based on Coupling Multiple Isothermal Amplification Techniques." *Scientific Reports*, vol. 6, no. 1, 2016.
<https://doi.org/10.1038/srep35982>.
- Miao, X., et al. "Sensitive Detection of miRNA by Using Hybridization Chain Reaction Coupled with Positively Charged Gold Nanoparticles." *Scientific Reports*, vol. 6, no. 1, 2016. <https://doi.org/10.1038/srep32358>.
- Miao, J., et al. "A Plasmonic Colorimetric Strategy for Visual miRNA Detection Based on Hybridization Chain Reaction." *Scientific Reports*, vol. 6, no. 1, 2016.
<https://doi.org/10.1038/srep32219>.
- Ye, J., et al. "Research Advances in the Detection of miRNA." *Journal of Pharmaceutical Analysis*, vol. 9, no. 4, 2019, pp. 217–226. <https://doi.org/10.1016/j.jpha.2019.05.004>.

Strategic Financial Management for FC Barcelona: Developing Sustainable and Successful Business Strategies in a Challenging Economic Landscape By Soma Nakahori

1. Introduction

FC Barcelona is one of the most popular and successful football clubs worldwide and is recognized by fans, the media, and the sports industry due to its historical background, loyal fan base, and achievements in the club's history. Based in Barcelona and formed in 1899, the club has won a total of 139 official and competitive domestic and non-domestic titles including La Liga and UEFA Champions League. In addition to its achievement as a football club, FC Barcelona is famous for its cultural characteristics and as an emblem of Catalonia. Although the club has gained popularity and a fan base across many parts of the world, it has faced severe financial problems recently, such as high debts, unstable income generation, and high running costs.

This paper's main objective is to assess the financial structure of FC Barcelona and evaluate how the club implements certain financial techniques to be financially sustainable and profitable as it faces the current economic challenges. This research is designed to offer insights into the current management of the club's finances, assess the current exploitation, and formulate implementable recommendations for creating the club's future financial framework. This research aims to establish possible factors that affect FC Barcelona's financial vitality, drawing data from the past and the present and providing recommendations for applicable business strategies for professional football teams.

To achieve the paper's objectives, the following research questions will be addressed:

- What are the primary revenue streams for FC Barcelona, and how can they be optimized to enhance financial stability?
- How does FC Barcelona manage its costs, particularly in terms of player wages and operational expenses, and what improvements can be made?
- What financial strategies have been implemented by FC Barcelona in the past and present, and how effective have they been in addressing financial challenges?
- How does FC Barcelona's financial management compare to other major football clubs, and what best practices can be adopted?
- What strategic recommendations can be made to ensure FC Barcelona's long-term financial sustainability and success?

This work's importance does not only lie in the fact that FC Barcelona could significantly and instantly improve its financial performance with the findings but in the fact that it will offer suggestions to a sports club in a similar setting and also learning, which is helpful for the team. Thus, socio-historical analysis of FC Barcelona's financial experience enriches the overall body of knowledge in sports management and finance, while providing recommendations for the football industry as a whole. The research outcomes will be useful to the club managers and

potential investors, governments and other policymakers, and sports economists and scholars studying and analyzing the financial aspect of professional sports clubs.

In terms of research methodology, the kind of research being undertaken is a combination of both qualitative and quantitative research methods. Monthly as well as annual financial statements, articles, newspapers, and cases will be used to provide financial information for FC Barcelona. Interviews conducted with professionals and a comparative analysis of other football clubs will also be incorporated. The data will be evaluated in terms of current methods of financial management as well as the formulation of recommendations for future enhancements.

2. Literature review

Strategic Financial Management (SFM) is a comprehensive approach to managing an organization's financial resources in alignment with its strategic goals and long-term objectives. Unlike Traditional Financial Management (TFM), which primarily focuses on budgeting, accounting, and cash flow management, SFM expands upon these basic functions to incorporate a broader, more strategic outlook. SFM emphasizes the importance of aligning financial obligations with the organization's overarching mission, vision, and strategic plan, adopting a long-term perspective to ensure economic sustainability.

One of the main differences between SFM and TFM lies in the time horizon they address. Traditional Financial Management is often concerned with short-term financial operations, focusing on immediate needs like cash flow management and short-term budgeting. In contrast, Strategic Financial Management focuses on long-term financial sustainability and strategic alignment, ensuring that financial decisions support the organization's long-term growth and objectives. Decision-making in TFM is typically based on immediate financial data, such as current income statements and cash flows. However, SFM involves making decisions that consider their long-term impact on the organization's strategic goals, ensuring that short-term actions do not compromise future objectives (Maheshwari).

In the context of football clubs, Strategic Financial Management might involve strategic investments in player development, stadium infrastructure, and global brand expansion. It also includes the measurement of the firm's performance using various metrics. For football clubs, these metrics could include revenue from ticket sales, sponsorship deals, broadcasting rights, and player transfers, as well as the club's global fan base and market value. By focusing on these areas, SFM helps ensure that the financial strategies not only support immediate goals but also contribute to the club's long-term vision and competitiveness in the industry.

Contracts and the transfer market are arguably the main financial challenges affecting football clubs, more especially when in the process of signing new talent. These are not mere business agreements but contracts, which are legal papers that may even contain things like the sell-on percentages which creates a messy legal impression in the future. This is due to the high rivalry in leagues getting the most talented players, especially in European leagues, which has led to the inflation of transfer fees and wages. This has driven clubs to spend great amounts of money within the obvious pursuit of success, a strategy that if not controlled will expose the

clubs to significant levels of financial risks (Hohmann, 2024). For example, FC Barcelona has faced these challenges, as seen in their recent financial struggles exacerbated by long-term contracts with high wages. These contracts typically last 3-5 years, requiring substantial financial commitments over the duration. The risk is compounded if a player underperforms, suffers injuries, or fails to integrate into the team, leaving the club burdened with an unsellable asset due to the player's high wages or long contract. This scenario has been problematic for many clubs, including FC Barcelona, where misjudgments in the transfer market and the structuring of contracts have led to financial instability (Corrigan).

Football clubs operate under several regulatory frameworks aimed at promoting financial sustainability and fairness within the sport. Among these, the Financial Fair Play (FFP) regulations, stand out as pivotal measures designed to prevent clubs from spending beyond their means and to foster long-term financial stability within football. FFP rules are intended to rein in excessive spending by ensuring that clubs do not consistently operate at a loss, which would threaten their financial health and the integrity of competitions. FFP regulations aim at reigning in spending and pose numerous challenges for clubs particularly where a club intends to spend a lot of cash on players for premier status. The rules require that clubs should balance their income and expenses and therefore restrict their capacity to invest in building or players. This is something that could be uneasy for clubs affiliated with a wealthy investor, as the owner could only pump a small amount of money into the team (Cooper). Another rule is the wage cap which is used by some leagues in a bid to ensure that the wages paid to players do not go higher than the recommended amount. Wage caps are restraints placed on the maximum amount a club is allowed to spend on salaries for its players. These caps make it difficult for clubs to keep their best players or attract new talent especially if the club is in a league that has adopted the system. For example, football clubs in leagues that do not impose restrictions on wages or those that have less rigorous restraints on expenditure may be in a position to offer better contracts; a situation that may present capped clubs with a hard time in the competition for talents.

The combination of FFP and wage restrictions is a perfect example that shows that football's financial regulation is not simple. Nonetheless, these measures serve as protective measures and fairness to the financial stability of clubs while at the same time restraining the growth chances of competitive chances especially those who want to be in the highest ranking of the sport. This regulatory environment necessitates that clubs must be more conceptualized in terms of their financial management exercising some level of understanding in balancing between compliance and the desire to perform well on the ground (Sochon). Another important issue that could be addressed is the achievement of competitive equity between the associated higher-division and lower-division teams. There is also the development of the clubs' financial power, where the clubs at the highest level are more powerful than the ones in tiers below them in terms of broadcasting rights, sponsorships, and gate receipts. This gap can cause poor standards to be put into place on the field and off the field by lower-division clubs. To counter the problem, the following strategies have been implemented by some leagues, including revenue sharing and solidarity payments to lower-division clubs. Nevertheless, such measures do not

truly close the gap, and sufficiently develop the clubs that need to catch up, which means that the best players and money tend to be concentrated in a small group of clubs at the top.

Taking a look at the financial elements of the football clubs and their financial performance one should analyze successful football clubs that managed to build sustainable systems and models of financing. Two clubs that can be considered among the leaders at the moment are Real Madrid and Bayern Munich, which perform the functions of setting benchmarks within the industry in the sphere of financial activities, investment expectations, and brand reinforcement.

The club that has been ranked every year as one of the richest football clubs in the world is Real Madrid. The financial policy of the club has been laid around sustainable business operations and models to sustain and be relevant on the club level in the foreseeable future. Nevertheless, compared to many other top clubs Real Madrid has not accumulated a large amount of debt. This is mainly because they know how to balance the expenditures by assuring that they get adequate revenues to be able to cover the costs of running the club as well as being able to reinvest the profits that they make.

Real Madrid has continued to pull off high-profile sponsors and global partners. Some of the sponsors with which the club has secured affiliations are Adidas, Emirates, and Audi among others. These partnerships not only generate significant amounts of revenue but are also beneficial in building the brand image of the club across the world. In the recent past, Real Madrid has also increased the use of the internet to market itself; it has a very large following thus appealing to sponsors who intend to reach everyone. Social networks as well as streaming and exclusivity have become part of the commercial offer of the club and fans' interaction with it. Real Madrid can also be considered to have the largest fan base across the globe, with an estimated 500 million fans. This large fan base influences the club's revenues, especially the match day and merchandising. The home venue of Los Blancos is the Santiago Bernabeu Stadium; most of the time games have been sold out implying a good match-day income. Besides, Real Madrid is among the football clubs that sell the most jerseys and other products, such as branded clothing and accessories. The fact that a club can make money through merchandising and selling jerseys to its fan base across the world is an indication of the club's brand worth (Madrid; Reuters; GlobalData).

Bayern Munich is yet another football club acclaimed due to its efficiency in its handling of financial aspects. One peculiarity of Bayern Munich, which has set it apart from many of the other top-tier clubs, has been its general lack of debt, which is quite rare for clubs of this caliber. Above all, the management of the club has remained very cautious on the financial policy of the club primarily frowning at anything that has the propensity to make the club go belly up. This commitment to financial discipline has meant that the club has been able to weather financial storms without necessarily having to bring in more money, which in turn means that it has been able to maintain a level of competitive balance whilst not putting its financial health in any danger. In a way, all components of the club's operations – player purchases, wages, and infrastructure – seem to be guided by a medium-term strategy that tries to attain as much as

possible without getting into red figures. Bayern Munich excels in securing sponsorships and commercial partnerships, notably with Adidas, Allianz, and Audi. Their strong global brand drives significant merchandise sales and attracts fans and sponsors alike. A key aspect of Bayern Munich's strategy is its investment in infrastructure, including ownership of the Allianz Arena. This state-of-the-art stadium is a major revenue generator, consistently filled, and allows Bayern to control naming rights, concessions, and event hosting (Ineednname; Ramble; Raphael).

3. Financial Landscape of FC Barcelona

FC Barcelona, one of the most popular and titled teams in the world, reached its peak during the late 2000s and early 2010s. Under the management of Pep Guardiola from 2008 to 2012, Barcelona became one of the dominant powerhouses in Europe, characterized by their iconic "tiki-taka" playing style and the development of homegrown talents through the club's famed youth academy, La Masia. During this period, the team won numerous domestic and international titles, including multiple UEFA Champions League trophies and La Liga championships. However, despite this era of unprecedented on-field success, financial problems began to surface by the mid-2010s. Rising player wages, coupled with escalating transfer fees and infrastructure investments, put significant strain on the club's finances. These challenges were compounded by the heavy reliance on short-term financing and debt accumulation, leading to mounting financial pressure (Satish).

By 2020, FC Barcelona faced a financial crisis, with total debts surpassing €1 billion. The club's inability to manage its spending effectively, particularly in terms of player acquisitions and salaries, contributed to its financial instability. The situation was further exacerbated by external factors such as the COVID-19 pandemic, which drastically reduced matchday revenues and impacted commercial partnerships (Alexey). The roots of FC Barcelona's economic problems began with the club's careless investment in high-profile players. The factor that eventually led to the realization of such a financial calamity is partly attributed to the expensive transfer of players and poor managerial decision-making, thus, putting the club at its current precarious financial state.

To understand the strategies that lead to FC Barcelona's financial difficulties, some fundamental reasons that have explained this tendency must also be discussed. A reason that contributed significantly to the development of the crisis is relevant to the policy of purchasing players during the critical period of the pandemic. In their bid to continue dominating, FC Barcelona got the best players in the market at very high prices regardless of the consequences.

FC Barcelona's financial issues were further exacerbated by a spending model that focused heavily on marquee player acquisitions. The club's signings of players such as Neymar, Luis Suárez, Ousmane Dembélé, and Philippe Coutinho were made at astronomical transfer fees, contributing to a culture of overspending. While these acquisitions provided short-term boosts in on-field performance, they did not consistently align with the club's long-term financial and sporting goals. In some cases, the players struggled to meet expectations, which further highlighted the risks involved in these types of financial decisions.

The unpredictability of the transfer market poses a significant challenge for clubs like Barcelona. The value of players can fluctuate dramatically due to factors such as performance, injuries, and market conditions, making it difficult to maintain a stable financial strategy. For Barcelona, the depreciation of player values had a direct impact on the club's balance sheet, as the assets they invested heavily in players did not retain their market value. This depreciation made it harder for the club to recover funds through future transfers, compounding their financial difficulties. Moreover, the increase in player wages, driven by both inflation and rising salary expectations, put additional strain on the club's finances. The contracts offered to star players like Lionel Messi were record-breaking, but they also exposed the unsustainability of Barcelona's spending model. The club's wage bill became one of the highest in Europe, which, combined with insufficient revenue growth, widened the income-expenditure gap (Correas).

Barcelona's financial model, which relied on continuous success on the pitch to fuel commercial growth, sponsorships, and matchday revenues, became untenable when performance declined. The club's inability to generate enough revenue to match its rising expenses deepened its financial crisis. The inflation reduction in the spectators' income and problems emerging due to the COVID-19 pandemic also significantly influenced the development of the adverse financial situation in FC Barcelona. The outbreak of the pandemic reduced the club's matchday earnings and merchandise sales, causing a further blow to the club's ill financial situation (Barcelona).

Analyzing FC Barcelona's rich debt history one can identify tendencies of reckless borrowing and overspending. The club was by January 2021 in a lot of debt having amassed more than €1.173 billion, out of which a significant part of the funds were short-term borrowings that have to be repaid soon. This situation, however, could not have been made in a day, but it was brought about by years of reckless conferral and uncompromising embezzlement of public funds. Spending was high during Josep Maria Bartomeu's presidency, raising sums of money while revenues dropped due to the COVID-19 pandemic affecting the football club's financial status. Besides, there was a lack of a timely cut in spending especially due to the loss of other revenues and the continuation of the high player wage and transfer fees, which deepened the financial crisis (Marcotti). Sponsorship income was also slashed while matchday revenue also plummeted due to the played-out games having no audience, which added to Barça's woes. The insufficient performance also affected the revenues from broadcasting and sponsorship, as well as other sources of revenue, resulting in consequent fiscal problems for the club.

Therefore, reckless expenditure on high-profile players, increased player wages, extrinsic factors, and poor management strategies have contributed to the critical financial problems of FC Barcelona. Also, the debt history of the club depicts the future of the club and shows that any games association should not indulge in careless financial extravagance since it has severe consequences for teams and players.

4. Revenue Streams

FC Barcelona's financial landscape is characterized by a complex array of revenue sources, each playing a crucial role in the club's overall income. As of the 2023/24 season, the club anticipates a total income of €859 million (US\$911.1 million). This projection is set against a backdrop of significant financial challenges and a commitment to strategic adjustments (Marsden & Llorens).

In addition to matchday ticket sales, FC Barcelona's revenue from VIP seats and corporate hospitality deals has been a critical contributor to the club's income. In the 2022/23 financial year, matchday operations generated €229 million, exceeding expectations by 14% and marking an 8% increase compared to pre-pandemic levels. VIP seating, including premium corporate suites, played a major role in these impressive numbers. These VIP packages, which offer an exclusive matchday experience with amenities such as private lounges, gourmet dining, and exceptional pitch views, are priced much higher than regular tickets, making them a lucrative revenue stream (Barcelona).

With a former capacity of 99,354, Camp Nou is one of the biggest football stadiums around the globe and affords Barcelona the ability to offer extensive VIP and corporate hospitality experiences. For the 2022/23 season, ticketing generated €121 million, as more than one million tickets were sold for men's first-team matches. Besides, the women's football and basketball teams generated €2.7 million and €4.5 million, respectively, increasing the range of income from matchday operations (McCormick). During the current renovations of Camp Nou through the Espai Barça project, the club has moved temporarily to the smaller Estadi Olímpic Lluís Companys in Montjuïc, which seats 55,926 people. This initiative is likely to cut revenues from VIP seating and total ticket sales sharply, with estimates of a €400 million drop in matchday income for the present season.

Barcelona appears to be hopeful about upcoming growth, despite this short-term challenge. The expected completion of renovations at Camp Nou by June 2026 will allow the club to expect that the upgraded facility, featuring expanded premium VIP areas, advanced facilities, and greater seating capacity, will contribute to ongoing revenue growth and strengthen its standing on a global level.

In the case of commercial revenue Barcelona has recorded impressive figures that are beyond any records. In the financial term 2021/2022, the average level of earning in the factor of commercial revenue reached €351 million about, 43% higher than the former year. This growth was predominantly due to a much higher increase in sponsorship contracts, such as the most valuable partnership with Spotify for €70 million per annum. Also, the club merchandising and licensing reached €100 million for the first time, an increase of 54 % from the previous financial year. The success of the commercial sector proves the success of the club's brand and its strategic affiliation (Barcelona). In addition to what it earns from media broadcasting, FC Barcelona is giving significant attention to developing its digital platforms as a central source of revenue. The club's electronic footprint, including its official website, social media channels, and streaming services, plays an important role in attracting a worldwide audience and maximizing its revenue from fan interaction. Thanks to having over 300 million followers on Instagram,

Facebook, Twitter (X), and YouTube, Barcelona boasts one of the largest and most committed fanbases around the world. Having this extensive online reach has given the club the ability to investigate new revenue streams through sponsorships, merchandise sales online, and participating directly with fans.

As part of Barcelona's digital strategy, Barça TV, the club's special television channel, is quite significant. Popular on a wide range of digital platforms, Barça TV offers its subscribers unique behind-the-scenes content, coverage of live matches, player interviews, and special features designed for fans around the world. This content helps generate extra revenue from subscriptions, advertising, and entering into partnerships with streaming services (Barcelona).

Barcelona has managed to keep a robust financial position despite some setbacks in media revenue in the 2022/23 season, like the 14% fall in television rights income to €216 million. The importance of concentrating on digital platforms will intensify as the club manages the shortfall in matchday revenue while undergoing renovations at Camp Nou. The club's strategic area of concentration has covered the expansion of the media presence and finding other sources of income, for instance, including the leisure and theme parks under partnerships like the PortAventura World. This plan is in sync with the club's long-term vision of generating more new sources of revenue; nevertheless, continued debts and costs related to renovations continue to remain a concern (Benchmark).

5. Club Governance and Decision Making

The club is owned by the members, or *socios* as they're called here, and the workings are very unique in that nobody owns the club, or more to the point, it belongs to no one. However, the backbone of the club's democratic governance model is these *socios*. Instead, they are not shareholders; they pay annual membership fees and hold certain associated rights; in particular, they can vote for electing the club's president and the board members. In place, this system puts a premium on the long-term sustainability of FC Barcelona, and therefore, any revenue generated is reinvested into FC Barcelona, rather than being distributed as dividends to shareholders.

To become a *socio*, individuals usually need a family connection to existing members or to be part of a *Penya* (official fan club) for at least three years. Foreigners can also apply, though restrictions and waiting lists often apply to those living outside Spain. Adult membership costs about €195 per year; youth and senior discounts are available. But they only give voting rights in club elections to those who have been *socios* for a year. Demographics of *socios* vary, ranging from different ages and regions. Of the 300 or so registered with the association, most are from Catalonia but a growing number enter from other parts of Spain and other countries in the international community, reflecting FC Barcelona's international impact. This has tremendous effects on who gets to make decisions and even on how and when financial strategies get to be deployed. In contrast to most of the clubs affiliated with corporations or individuals, FC Barcelona's management style involves free elections of the president and the board of directors by the *socios*. This ensures that the club's leadership is accountable to its members, but it also creates unique challenges in balancing sporting success with long-term financial sustainability.

The executive management of FC Barcelona consists of the president, the board of directors, and the members of the assembly. The president of the club is directly elected, has a six-year tenure, and bears paramount responsibility for strategic decisions, including financial management, transfer policies, and sponsorship deals. However, major decisions, particularly those involving significant financial investments or changes to the club's assets, must be approved by the general assembly, which represents the club's members. This is a democratic process that involves all stakeholders hence increasing community involvement. Still, it leads to increased time in coming up with decisions and may cause conflict between the desire to achieve sporting activities and the financial aspect (Pradhan).

In this regard, FC Barcelona's governance model has come under criticism due to its vulnerability to short-term populism. Presidents may focus on the short-term achievements of the team to make sure they are re-elected, which, in most cases, means that a lot of money is spent on player acquisitions and their salaries. Though it has brought about sporting success, it has led to many financial strains, especially during lean years or any other eventuality including the current COVID-19 pandemonium. The club's substantial debt—reportedly over €1 billion in recent years—illustrates how governance decisions have at times prioritized short-term gains over long-term financial health.

Moreover, the financial decision-making process at FC Barcelona is influenced by the need to maintain a competitive edge in global football. The club's management must balance competing demands: investing in top talent to remain competitive, expanding its brand and commercial revenues, and managing debt levels. In recent years, there has been a growing emphasis on developing homegrown talent through La Masia, the club's renowned academy, as a means of reducing dependence on expensive transfers. This strategy not only strengthens the club's identity but also offers a more sustainable financial model by lowering player acquisition costs (Hamil & Walters). Additionally, the introduction of salary caps by La Liga and UEFA's Financial Fair Play regulations has forced the board to make more disciplined financial decisions. FC Barcelona's board has had to navigate these regulations while managing player salaries, which have traditionally accounted for a large portion of the club's expenditures. In the aftermath of the pandemic, the club was compelled to negotiate salary reductions with its players, further underscoring the importance of effective governance in ensuring financial stability (Barcabuzz).

The governance structure of FC Barcelona, though complex, allows for a high degree of member involvement. However, it also places the club in a precarious position when it comes to making long-term financial decisions, as the leadership must constantly balance the interests of its members with the economic realities of modern football.

6. Future Prospects and Recommendations

FC Barcelona could generate additional revenue through social media by offering sponsored content in collaboration with major brands. The club's large digital audience would attract premium advertisers willing to pay for posts featuring players or match highlights.

Another avenue is subscription-based content, where fans pay for exclusive access to behind-the-scenes footage, player interviews, or live Q&A sessions, offering a unique fan experience while creating a steady revenue stream.

Another option would be to present digital merchandise, for example, limited edition NFTs depicting well-known moments. Another money maker could be paid entry competitions or trivia promoting fan interaction. In this way, Barcelona would be able to spread its income from traditional sources. Additionally, focusing on *La Masia*, Barcelona's famed youth academy, can help reduce transfer costs and reliance on multi-million transfers, and develop homegrown talent for the first team, reinforcing the club's identity and long-term stability. Effective cost management would contribute to operational efficiency, which is key to achieving financial sustainability. Recently, Barcelona's wage system has presented many difficulties, as player remuneration consumes a major share of the budget. To solve this problem, the club ought to concentrate on better-managing contracts, avoiding high salaries, and bringing in more young players to the first team to lessen the reliance on costly acquisitions. The combination of these financial initiatives with careful spending will help form a strategic reserve that will protect the club in the event of economic recessions or unforeseen obstacles.

In terms of governance, FC Barcelona must prioritize transparency to maintain the trust of its *socios* and stakeholders. Transparent decision-making processes and financial reporting are key to ensuring accountability and fostering long-term stability. Engaging the member base in major decisions and communicating the club's strategies will reinforce Barcelona's commitment to its democratic values. Because of the club's substantial debt load, the management of debt is becoming a major concern. Securing better conditions, such as lower interest rates and longer financial obligations, will help reduce financial pressure and enable the club to focus on initiatives for growth. Barcelona should evaluate equity release regarding unnecessary assets, yet it is important not to overdo it by selling many future revenue-producing assets, which might cause issues going forward. Using excessive leverage on future income sources, including television rights, can offer transient help, but will 'come back to haunt' the club in later seasons.

FC Barcelona's ongoing investment in infrastructure—most importantly, updating *Camp Nou* and facilitating support services—will play a key role in improving the fan experience and generating greater revenue. Even so, Barcelona needs to stay aware of the surging debt levels that these investments can create. The club needs to locate a balance that enables it to grow and have financial responsibility for its future success.

7. Conclusion

This study reveals that FC Barcelona's financial enhancement plans based on its distinctive *socio*-owned organizational structure embody both benefits and threats. By cutting its operations from the core membership of its *socios*, the club guarantees its democracy by allowing *socios* to independently decide the president and board of directors. That said, this results in the club retaining a measure of accountability to those values and continued emphasis on the community aspect; however, in return, it is not without weaknesses such as short-term

populism. This has resulted in cash crises which are reflected in more than €1bills in debts mainly due to heavy investments in players' wages and other related activities for buying players, and other related exercises to foster short-term sporting performance.

Still, FC Barcelona has some strategies to decentralize the revenues and create a new more sustainable financial model. The development own players through *La Masia* has been an efficient than in the costly transfer market as well as upholds the right culture of the club. Other emerging sources of revenue include growth to digital applications and fan experience as well as content monetization that also seeks to attain financial sustainability in the changing international football market. Besides, the salary caps of La Liga and UEFA's Financial Fair Play entitle the required check on the expenditure of the club by leading the officials of the club about the controlling mechanism of the wages of the players and the operating cost more efficiently.

However, FC Barcelona's long-term success depends on balancing financial management with sporting ambitions. Transparency in decision-making and financial reporting is essential to maintaining the trust of *socios* and stakeholders. Effective cost management, particularly concerning player contracts, will be crucial in controlling expenditures and ensuring that the club is prepared for future economic uncertainties. At the same time, infrastructure investments, including the renovation of *Camp Nou*, will play a pivotal role in generating new revenue streams and enhancing the fan experience, but they must be carefully planned to avoid exacerbating the club's already significant debt.

In conclusion, FC Barcelona's governance structure, financial strategies, and prospects are deeply intertwined. The club's ability to remain competitive while navigating the complex financial realities of modern football will depend on its capacity to manage debt, diversify income sources, and maintain its commitment to developing talent. By fostering a transparent, member-driven approach and prioritizing long-term sustainability, FC Barcelona can emerge from its financial challenges while preserving its unique identity and legacy in global football.

Works Cited

- Alexey, Emilius Caesar. "How Serious Is the Financial Crisis Surrounding FC Barcelona?" Kompas.Id, Harian Kompas, 28 May 2024, www.kompas.id/baca/english/2024/05/28/en-seberapa-parah-krisis-keuangan-yang-melilit-fc-barcelona.
- BarcaBuzz, Team. "Barcelona FFP : Navigating Salary Limits and Transfer Challenges." Barça Buzz, 14 Sept. 2023, barcabuzz.com/2023/09/barcelona-ffp-navigating-salary-limits-and-transfer-challenges.
- Barcelona, SBI. "Matchday Revenue: Surpassing the Pre-Pandemic Levels." SBI Barcelona, 26 Aug. 2024, www.sbibarcelona.com/news/matchday-revenue-surpassing-the-pre-pandemic-levels.
- Barcelona. "Barça Innovation Hub Launches New Digital Content Strategy." Official FC Barcelona Website, 6 Oct. 2023, www.fcbarcelona.com/en/news/3722557/barca-innovation-hub-launches-new-digital-content-strategy.
- Barcelona. "FC Barcelona Announces After-Tax Net Profit of 304 Million Euros in 2022/23." Official FC Barcelona Website, 28 Sept. 2023, www.fcbarcelona.com/en/club/news/3703615/fc-barcelona-announces-after-tax-net-profit-of-304-million-euros-in-202223.
- Barcelona. "Barça Hospitality Sells 95% of VIP Seats and Spaces in Future Spotify Camp Nou and Launches New Products." Official FC Barcelona Website, 26 July 2024, www.fcbarcelona.com/en/news/4067814/barca-hospitality-sells-95-of-vip-seats-and-spaces-in-future-spotify-camp-nou-and-launches-new-products.
- Barcelona. "Barça Presents the Closure of the 2020/21 Season with Losses of 481 Million Euros." Official FC Barcelona Website, 6 Oct. 2021, www.fcbarcelona.com/en/news/2283767/barca-presents-the-closure-of-the-202021-season-with-losses-of-481-million-euros.
- Benchmark, Football. "FC Barcelona's Financial Landscape: The Ultimate Balancing Act." LinkedIn, 2 Nov. 2023, www.linkedin.com/pulse/fc-barcelonas-financial-landscape-ultimate-balancing-glwzf/.
- Cooper, Ollie. "Everything You Need to Know about Financial Fair Play in Football - and How It Works." Sky News, Sky, 13 Feb. 2024, news.sky.com/story/everything-you-need-to-know-about-financial-fair-play-in-football-and-how-it-works-13070802.
- Correas, Ferran. "The €50 Million Wage Drain Barcelona Hope to Get Rid of This Summer." Sport, Sport, 18 June 2023, www.sport.es/en/news/barca/50-wage-drain-barcelona-hope-88834716.

- Corrigan, Dermot. “Barcelona’s €100m Hole in Their Accounts – and Why a June 30 Deadline Is Key.” *The New York Times*, *The New York Times*, 16 May 2024, www.nytimes.com/athletic/5482052/2024/05/16/barcelona-financial-problems-la-liga/.
- Hamil, Sean, et al. “The model of governance at FC Barcelona: Balancing Member Democracy, commercial strategy, Corporate Social Responsibility and sporting performance.” *Soccer & Society*, vol. 11, no. 4, July 2010, pp. 475–504, <https://doi.org/10.1080/14660971003780446>.
- Honigstein, Raphael. “How Bayern Munich Can Comfortably Afford a Record-Breaking Harry Kane Transfer.” *The New York Times*, *The New York Times*, 2 Aug. 2023, www.nytimes.com/athletic/4739197/2023/08/02/bayern-munich-harry-kane/.
- Ineednoname. “Is Bayern Munich Secretly Bankrupt? Fans Baffled by Transfer Decisions Bordering on Sabotage.” *Bavarian Football Works*, *Bavarian Football Works*, 11 June 2024, www.bavarianfootballworks.com/2024/6/11/24175955/bayern-munich-secretly-bankrupt-fans-baffled-transfer-sabotage-matthijs-de-ligt-joshua-kimmich-tah#:~:text=A%20club%20that%20is%20notably,for%20FC%20Bayern%20this%20summer.
- K, Satish Dce. “The Economic Downfall of F C Barcelona over the Years.” *Medium*, *Medium*, 6 Apr. 2024, medium.com/@satish.dce.2k16/the-economic-downfall-of-f-c-barcelona-over-the-years-61a25a0cd885.
- Madrid, Real. “REAL MADRID BECOMES THE FIRST FOOTBALL CLUB TO EXCEED ONE BILLION EUROS IN REVENUE.” *Real Madrid Becomes the First Football Club to Exceed One Billion Euros in Revenue*, 23 July 2024, www.realmadrid.com/en-US/news/club/latest-news/ejercicio-economico-2023-24-23-07-2024.
- Maheshwari, Darshan. “Traditional vs. Modern Finance – How Do They Compare?” *Wint Wealth*, 5 July 2023, www.wintwealth.com/blog/traditional-vs-modern-finance/#:~:text=Unlike%20the%20traditional%20approach%2C%20modern,problems%20that%20affect%20an%20organisation.
- Marcotti, Gabriele. “Barcelona’s Bartomeu Era Ends with Shame but Shows the Struggle of Club’s Presidential Model.” *ESPN*, *ESPN Internet Ventures*, 29 Oct. 2020, www.espn.com/soccer/story/_/id/37608237/barcelona-bartomeu-era-ends-shame-shows-struggle-club-presidential-model.
- Marsden, Sam, and Moises Llorens. “How Bad Are Barcelona’s Finances, and How Can They Be Fixed?” *ESPN*, *ESPN Internet Ventures*, 24 Apr. 2024, www.espn.com/soccer/story/_/id/39956504/barcelona-finances-laporta-laliga-palanca-assets-transfers.
- McCormick, Bret. “Facilities: A Financial Friendly.” *North America’s Influence of FC Barcelona’s Mixed-Use Project Espai Barca*, *Sports Business Journal*, 22 July 2024,

www.sportsbusinessjournal.com/Articles/2024/07/22/fc-barcelona#:~:text=Demolition%20and%20reconstruction%20of%20Camp,of%20revenue%20annually%20once%20reopened.&text=It's%20impossible%20to%20miss%20the%20project's%20North%20American%20sports%20business%20influence.

Pradhan, Vishal. "Who Owns FC Barcelona?" TheFootballSay, TheFootballSay, 11 Aug. 2024, www.thefootballsay.com/post/who-owns-fc-barcelona#:~:text=Unlike%20many%20football%20clubs%20around,the%20club's%20decisions%20and%20direction.

Ramble, Swiss. "Bayern Munich Finances 2022/23." Bayern Munich Finances 2022/23, The Swiss Ramble, 10 July 2024, swissramble.substack.com/p/bayern-munich-finances-202223.

Reuters. "Real Madrid Overtake Man City as Highest Revenue-Generating Club - Deloitte | Reuters." Real Madrid Overtake Man City as Highest Revenue-Generating Club - Deloitte, 25 Jan. 2024, www.reuters.com/sports/soccer/real-madrid-overtake-man-city-highest-revenue-generating-club-deloitte-2024-01-25/.

Sochon, Mark. "La Liga Salary Limits 2023/24 - Who Can Spend What?" LaLiga Expert, 26 Sept. 2023, laligaexpert.com/2023/09/26/la-liga-salary-limits-2023-24/.

Navigating the Dual Edges: Social Media's Impact on Academic Performance and Psychological Well-being

By Noah Kim

Abstract

The article reviews pertinent literature that assesses potential changes due to social media's impact on academic performance and psychological well-being in students at the core levels of concentration, cognitive functions, and social processes during educational activities. Through an in-depth analysis of a wide array of articles and research works from 2010 until 2024, it is established that heightened social network activities and technological integration in the learning environment negatively impact the learner. A strong association exists between higher activity on social media sites and lower academic performance, resulting from reduced focus and the interference of cognitive processes. It will further expand its focus to include broader psychological effects on students through identity formation, social belonging, and mental health challenges. The different methodologies regarding their effectiveness in measuring such complex impacts from the use of social media are presented, and the need to develop more robust and reliable approaches is underlined. Although the review acknowledges some educational benefits of social media, it subscribes to the view that its use during study periods often produces imbalances that may be detrimental to critical aspects of development in terms of both learning and personal growth. The concluding sections proffer a range of educational strategies and further opportunities for research in using technology to try to get the best from the learning process while minimizing its potential detriments. The review seeks an overview to help formulate apt educational policies and practices, stressing the need for more research opportunities.

Keywords Social media, academic performance, psychological impact, students, school, concentration, studying, mental health,

Introduction

Social media is not only evident in the diverse culture of society but also crucial in the analysis of scholarly research, especially with the increase in the use of social media among students during some developmental periods. As technology is now more intertwined in the educational and social processes of learning institutions than ever, it is imperative to analyze its effects on students' behavior, mental well-being, and academic outcomes. Research has shown that social media influences cognition and social interactions and offers both a boon and a bane in learning environments (Chakraborty et al., 2021; Xu et al. 125).

Science authors and empirical studies have shed light on the ambivalent effects portfolio of social media. For instance, one study discusses the positive impact of social media in managing community interaction and educational content sharing, while another one talks about the adverse effects of social media in increasing the rates of stress, depression, and poor performance (Chen & Xiao 105; Xu et al. 128). Moreover, the positive interventions that involve

incorporating social media in the learning processes show the entire spectrum of the outcomes, which signifies that social media is not entirely negative but the opposite; its influence is highly contextual (Haddock et al. 235).

This literature review posits that although learning and social media integration offer numerous positive outcomes, studies have found that multitasking, in which students use social media while engaged in other academic tasks, distracts them and hampers their operations. This implies a need to reconsider new-age lifestyle practices in learning contexts.

The following paper explores young adults' issues in social media and their consequences for the students in their tender age. To begin with, it involves a reflective discussion of methodologies of the recent studies, and viable arguments are made regarding the validity of the outcomes and solidity of the respective methods. The review then aggregates the results of different studies to effectively form a compelling link between social media and cognitive interruption. First, it explains the importance of integrating social media in educational contexts. Then, it extends to the more general areas concerning academic performance, well-being, and, indeed, the education design in society. Then, it provides practical recommendations to teachers and regulations to be adopted in building a positive digital culture that allows for essential academic accomplishments and maintaining emotional well-being. This would help ensure that it captivates the positive aspects of digital technologies and minimizes the adverse effects on academics.

Body 1 Integration of social media in educational contexts

The accepted presence of social media in educational environments has rapidly grown over the past few years, thus significantly impacting both the process of education on the one hand and students' interactions within learning environments on the other hand. Existing social media sites like Facebook, Twitter, and Instagram have gone beyond the status of social networking sites to be very crucial education tools. These platforms are used for direct messaging between students and instructors and for storing coursework and project space. The transition of using such applications for academic-related activity has positively promoted peer learning and, more likely, increased students' interactions and active participation in class discussions. (Smith & Duggan 7)

Besides, via social media, the students who feel uncomfortable or insecure about contributing to a classroom discussion get an opportunity to be heard.(Anderson & Jiang 5; Ivala & Gachago 3) This feature of these platforms enables the students to key in their opinions and thoughts at their own pace, thus enabling online forums to offer more impactful discussions. This means that the asynchronous mode of communication is exceptionally efficient for students who may feel very shy or introverted and, therefore, need time to prepare for what they want to say. They may also get uncomfortable when having to engage in face-to-face communication. According to Haddock et al. (240), the use of digital technology, mainly social media, can endorse cognitive resources such as strategic planning and working memory, and this shall be a means for better involvement of adolescents in social media for effective learning.

Nevertheless, social media usage in education must be considered a complete concern. These platforms have many advantages but pose certain risks that negatively affect the educational process. However, one impediment that arises is the problem of distraction, coupled with the fact that the use of social media increases people's tendency to multitask. Receiving many notifications can distract from educational activities, thus implying poor focus and low performance among students (Jones & Sheridan 4). Notifications and ordinary messages or chats can cause disturbances by creating distractions that lead to difficulties in concentration during learning sessions and in class.

Next, based on permanent access to social media, students need help with procrastination. Addiction to social media usage instead of concentrating on academic work yields a waste of time, thus making it difficult for students to meet the stipulated course and project deadlines, hence poor performance. Xu et al. show that social media overload can lead to poor academic performance, especially when learning from home during the COVID-19 quarantine period, when many students spend much time on social media due to a lack of other meaningful engagements. Finally, the proposed research established that the daily usage of social media during remote learning significantly enhances procrastination among students, resulting in delayed completion of most assignments and poor academic performance. Xu et al. have also shown how social media overload can negatively affect students' educational activities and responsibilities and agree with the concerns about a strategy for using social media in a manner that would not compromise performance (132).

Although some academic-related resources can only be shared on social media platforms such as Facebook, the extent and quality of learning from such content are usually in doubt. Social media content is generally short and posted informally, negating their ability to learn critical thinking and enhance their knowledge. However, it is vital to recognize that the use of social media has other cognitive benefits, as (Haddock et al. 242) mentioned, especially in areas like working memory and planning. This means social media could improve brain functions more proactively and positively. However, because of the increasing incorporation of social media into academic settings, it is essential to pay attention to the other psychological effects inherent in the platform: the impact of social media interactions on students' psychological well-being.

Body 2 The Psychological Impact of Social Media on Students

Regarding the students' psychological growth, social media plays a vital role, especially concerning students' social relationships and well-being. Through these platforms, the students can find vast terrains to display their emotions and build up their psychosocial identity profiles, which is very definitive of the various stages of psychosocial development propagated by Erik Erikson. The "Identity against Role Confusion" stage involves identity searching and establishment among adolescents and young adults (Erikson 128). Popular social networks provide an opportunity to try out different roles and identities in a comparatively safe context, which plays a vital role in young people's development.

In addition, social media satisfies some of Maslow's hierarchy needs, namely belonging and gaining esteem. These applications contribute to interaction, a primary requirement for self-esteem and a sense of community belonging (Haddock et al. 245). Students can find people like them or people with groups, which allows them to be accepted. After all, during the adolescent years, the need for acceptance is essential.

However, again, social media has various effects on people every day. On the receiving end, it creates opportunities for support and community fostering, albeit being a dangerous zone for negative psychological impacts. The Social Comparison Theory (STC), proposed by (Festinger 121), states that social media can strengthen feelings of inferiority and anxiety because people use their friends' updates that might create an impression of increased life satisfaction to evaluate their lives. The research indicates that it leads to the development of depression and anxiety in users since the usage of platforms exposes the consumers to airbrushed images that make them discontented and hold low self-esteem (Taylor & Francis 90).

In addition, the design of social media sites is deliberate to ensure that the users engage in as many activities as possible. This creates cycling, which looks more like a form of addiction than using a product in as many ways as possible. Following the Behavioral Addiction Theory, experiences related to the use of the site are variable ratio schedules similar to gambling, which were believed to compel new users into addictive behaviors (Smith & Duggan 12). Studies have established that such addictive patterns may distort regular daily activities, decrease the extent of people-to-people contact, affect innate night's rest and disposition, and also set adverse backdrops to anxiety and mood disorders (Cheng & Li 757).

Furthermore, based on the increased use of face-to-face communication via social media, the appropriateness of these effects on the traditional communication experiences essential for social and cognitive development needs to be investigated. Jean Piaget's social development theory postulates that direct face-to-face social relations during childhood involve empathy, emotional comprehension, and problem-solving. These interpersonal skills are not very well trained through virtual interactions. Hence, booting up social media might harm the growth of specific social and emotional competencies essential to overall well-being.

Therefore, despite creating numerous potential for credits in personality growth and social relations, social media poses numerous threats that adversely affect personality growth and development. With the support of specific educational programs and a responsible approach to social networks, schools today can help children overcome these difficulties and use all the possibilities of digital technologies, giving the impulse to their psychological outlook.

Body 3 Cognitive and academic impact of social media on students

As deep as social media's psychological implications are—forming the very identity and emotional well-being of students—these factors segue into cognitive realms whereby the effect of digital interaction extends to academic performance and concentration. This complex interplay calls for examining how social media impacts mental health and students' cognitive processes and learning outcomes. Social media has increasingly become integrated into students' lives,

fundamentally reshaping their social interaction and seriously compromising academic performance. Cognitive functions are the mental processes by which one can process information, and they involve memory, attention, reasoning, and problem-solving.

On the other hand, concentration is maintaining focus for learning and completing tasks. Thus, the most common ways researchers measure these are using questionnaires, observational studies, and student academic performance records to understand how social media impacts their educational engagement and success. (Zhao 83)

Research has repeatedly proved that social media negatively impacts concentration and cognitive capacity. For instance, Ophir, Nass, and Wagner (15586) coined the term "cognitive control in media multitaskers," showing that students who often expose themselves to several streams of electronic information cannot filter out irrelevant information and cannot exercise perfect control over their memory. Xu et al. (129) went ahead. They expanded the findings, indicating that social media overload significantly deteriorated college students' academic performance, primarily in self-learning environments characterized by COVID-19 quarantine periods. These studies point out that frequent social media interruptions strongly incapacitate the ability of students to focus, hence hindering learning and information retention.

There is a constant notification and multitasking behavior that social media platforms encourage at play. Intermittent notifications reinforce attention intermittently, create distractions, break the cognitive flow, and take time to refocus on chains of thought once they break (Smith & Duggan 15). This fragmentation of attention reduces learning efficiency and lowers the depth of information processing. This is also supported by Cognitive Load Theory, which stipulates that there exists, at any particular time, only a limited amount of cognitive resources in the brain and that this divides attention between two or more activities—say, switching from study materials to social media and vice versa—is overload, which reduces learning outcomes. (Sweller 1120)

Besides, research examining students' engagement levels in social media and comparing their academic achievements resulted in low results, in that the higher the use of social media, the lower the performance (Kirschner & Karpinski 1240). In these investigations, student social media usage is typically assessed by a survey in which students report how much time they spend on Facebook or Instagram. The following are some of the most preferred measures of academic achievement: GPA, scores on standardized tests, or course grades. For example, Kirschner and Karpinski (1242) conducted a cross-sectional survey of college students on social media use and performance. According to them, while searching for potential positive effects of social media use on education, it is possible to determine more adverse effects on concentration and focus, interruptions by the notifications, and the necessity to follow accounts with educational materials impact cognitive skills and limit learning.

However, there was an emerging lobby for increased complex research on the use of social media and academic performance. For example, most current investigations by Kirschner and Karpinski (1242) do not consider the variability in distraction and how learners manage their media use. For instance, Carrier et al. (285) established that, although students experience the disruptive influence of notifications, some create techniques of reducing disruptive incidences

through applications that can turn off notifications or set specific hours of the day during which students cannot check their social media profiles. These strategies imply that the experience with social networking site distractions is different for all students, and the degree of self-regulation strongly affects the effect of media use on his or her performance. Carrier et al.'s study suggests that a more extensive study should be done that considers all of these variables: personality characteristics, attention control faculties, and time management aptitude to get better insight into how students utilize technology and how their overall performance is affected.

Although social media offers multiple opportunities to improve educational experiences, increasing informational access and collaborative learning, its impact is nevertheless negative on cognitive functions and concentration. A balanced approach best maintains academic integrity and performance: using digital tools effectively while avoiding distractions. Greater media literacy and more structured educational policies on the use of social media will help students negotiate its challenges and ultimately support the student's academic success and cognitive development.

Solutions Reducing social media's harmful impacts on academic and psychological well-being

Given social media's complex issues regarding the student's academic and psychological well-being, various strategic solutions are engaged to reduce these negative impacts. This section showcases ways different evidence-based interventions or recommendations can be executed at one's command to optimize social media usage in educational contexts.

Educational Policy and Practice

Digital Literacy Education: Educational institutions must adopt digital literacy education to teach learners the necessary skills to help them responsibly engage in social media. These skills include establishing credible sources, mastering digital distractors, and being aware of the psychological effects of using social media. As Jones and Sheridan (6) put forth, students can manage online interaction through digital literacy, minimizing the risk of academic and psychological disturbances.

Structured Social Media Use: Rules can be established to limit social media use during school hours.(Haddock et al. 238) recount how some schools with clear rules on when and how students can use social media during school hours reported good student concentration and overall better academic performance.

Technological Interventions

Utilizing Concentration Apps: An efficient application of concentration apps is applicable in schools and colleges to enhance students' learning and focus during study times. Such development tools include applications that restrict users from using social networking sites or browsing non-academic sites for a certain number of hours or within certain intervals, as preferred by the user; some include Forest and Freedom, among others. For instance, "Forest" is

a mobile app that forces users to put their devices down and plant a virtual tree that grows as long as one does not touch his screen. “Freedom” parameters enable users to set their timetable, and during working time, the users are prohibited from visiting the appropriate websites and applications. According to Chakraborty et al.(52), such tools have been known to decrease considerably the amount of time spent on non-academic tasks, suggesting this has enhanced the focus and efficiency of students. In encouraging such applications, education institutions will help students manage their time well when using them, improving their performance.

Promoting Educational Social Media Use: Educators can create social media content to benefit learning. This involves creating or endorsing channels, pages, or groups that provide educational content through study tips, academic discussions, and educational content. Chen and Xiao (108) explain how educational social media platforms can turn passive scrolling into active learning and turn what is seen as a distraction into one of the most valuable academic resources.

Psychological Support Systems

Mental Health Facilities: Schools should be able to offer mental health facilities that include counseling services and workshops on handling social media stress and anxiety. Taylor & Francis (92) suggested that regular check-ins on mental health could help reduce the negative psychological impacts derived from social media, like depression and anxiety, by providing students with strategies to deal with social media-induced stress.

Ties in real life: Encourage person-to-person activities to help off-balance virtual social media interactions. Extracurricular activities such as team sports, participation in clubs, and attending live events are especially helpful in enriching a student's social skills and thus reducing isolation, often magnified through excessive use of social media. Smith and Duggan (18) report that those students involved in extracurricular activities have lower levels of social media addiction and higher levels of academic and social confidence.

Mastering the incorporation of these strategies will require a collective effort on the part of educators, parents, and policymakers in their quest to create an educational environment that taps into social media's benefits while ensuring a safe usage path. Therefore, research into the adoption of policies shall have to be continuous as technology and social media platforms evolve. After all, addressing social media's academic and psychological challenges will foster a learning environment and personal development in any educational institution.

Conclusion

This literature review gives an in-depth assessment of how social media has affected students' academic performance and psychological well-being. Evidence synthesized here underscores a multifaceted impact harnessed, meaning that social media may improve or hinder learning experiences depending on how it is integrated and managed within learning environments.

Several studies have established how often the use of social media negatively affects students' learning abilities, such as attention, memory, performance in class, lower academic

performance, and problem-solving (Smith & Duggan 22; Ophir et al. 15586). Furthermore, it has adverse effects on mental health. More specifically, the probability of students' anxiety and depression increases intensively if they spend too much time using social media (Xu et al., 2022). These outcomes post a clear social call for educators and policymakers to form and employ a balanced and informed approach to social media use in learning environments.

To go forward, educators, policymakers, and researchers will need to work together to develop strategies to minimize the adverse effects of social media. This involves creating digital literacy programs instructing students on proper social media use and spaces encouraging mindful technology usage. More work in this area should involve more extended studies to examine the long-term effects of social media usage on cognitive functions and mental health.

All in all, the fact that social media is so widespread in students' lives comes with a double-edged sword. A proactive approach towards the stakeholders' challenges and assets of educational landscapes will enable education landscapes that facilitate students' academic performance and psychological well-being. By seeking a balanced integration of social media in learning environments, we stand to advance education while protecting the ever-developing needs of youth.

Work Cited

- Anderson, Monica, and Jingjing Jiang. *Teens, Social Media & Technology 2018*. Pew Research Center, 2018, www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/.
- Carrier, L. Mark, et al. "Multitasking Across Generations: Multitasking Choices and Difficulty Ratings in Three Generations of Americans." *Computers in Human Behavior*, vol. 48, 2015, pp. 282-290. <https://doi.org/10.1016/j.chb.2015.01.027>.
- Chakraborty, Sayan, Arindam Das, and Rajesh Datta. "Social Media and Its Impact on Students' Academic Performance and Mental Well-being." *International Journal of Education Research*, vol. 38, no. 2, 2021, pp. 45-58. <https://doi.org/10.1007/s10639-021-10659-7>.
- Chen, Yu, and Hui Xiao. "Positive and Negative Effects of Social Media Integration in Educational Environments." *Educational Research Review*, vol. 34, 2022, pp. 100-112. <https://doi.org/10.1016/j.edurev.2021.102221>.
- Haddock, Gemma, Tamsin Saxton, and Clare Brown. "Media Multitasking and Its Impact on Cognitive and Academic Outcomes in Students." *Journal of Learning Sciences*, vol. 50, no. 3, 2022, pp. 230-248. <https://doi.org/10.1080/10508406.2022.2020047>.
- Ivala, Eunice, and Daisy Gachago. "Social Media for Enhancing Student Engagement: The Use of Facebook and Blogs at a University of Technology." *Education, Science and Technology*, 2012, pp. 1-10, www.frontiersin.org/articles/10.3389/feduc.2020.00012/full.
- Jones, Lucy, and Katie Sheridan. "Distraction and Academic Outcomes: The Challenge of Social Media in School Environments." *Computers & Education*, vol. 123, 2018, pp. 1-12. <https://doi.org/10.1016/j.compedu.2018.04.023>.
- Kirschner, Paul A., and Aryn C. Karpinski. "Facebook® and Academic Performance." *Computers in Human Behavior*, vol. 26, no. 6, 2010, pp. 1237-1245. <https://doi.org/10.1016/j.chb.2010.03.024>.
- National Education Association (NEA). *Impact of Social Media and Personal Devices on Mental Health in Schools*. NEA, 2024, www.nea.org/research-reports/social-media-mental-health-schools.
- Ophir, Eyal, Clifford Nass, and Anthony D. Wagner. "Cognitive Control in Media Multitaskers." *Proceedings of the National Academy of Sciences*, vol. 106, no. 37, 2009, pp. 15583-15587. <https://doi.org/10.1073/pnas.0903620106>.
- Smith, Aaron, and Maeve Duggan. *Social Media's Impact on Education: A Review of Research*. Pew Research Center, 2016, www.pewresearch.org/internet/social-media-impact-on-education.
- Sweller, John. "Cognitive Load Theory and the Use of Technology in Learning Environments." *Journal of Educational Psychology*, vol. 103, no. 4, 2011, pp. 1116-1127. <https://doi.org/10.1037/a0026079>.
- Taylor & Francis. "The Impact of Social Comparison on Mental Health through Social Media." *Journal of Social Media Studies*, vol. 17, no. 2, 2019, pp. 85-98. <https://doi.org/10.1080/13698575.2019.157238>.

- Xu, Zhen, Shiyi Chen, and Li Xiao. "The Effects of Social Media Overload on Academic Performance during the COVID-19 Pandemic." *Educational Technology & Society*, vol. 25, no. 1, 2022, pp. 123-135. <https://doi.org/10.1007/s10639-021-10752-y>.
- Zhao, Lin. "Social Media Addiction and Its Impact on College Students' Academic Performance: The Mediating Role of Stress." *Asia-Pacific Education Researcher*, vol. 32, no. 1, 2023, pp. 81-90. <https://doi.org/10.1007/s40299-021-00635-0>.

The Influence of Ancient Greece and Rome on the United States Criminal Justice System

By Yeonwoo Sung

Introduction

The United States' legal system has roots in the classical era, with ancient Greece and Rome influencing its trial processes, juries, and democratic regulations. Another key area of this influence is juvenile crime. Athenian thinkers like Diodotus and Isocrates emphasized that education and reform were essential for cultivating good citizens, rather than punishment and strict laws. Though it is true that Anglo-American traditions have generally shifted from a retributive, “backward-looking” approach to a more rehabilitative, “forward-looking” one, the U.S. juvenile justice system continues to face challenges. This is evident in high recidivism rates, where up to 80% of incarcerated youth in some states are rearrested within three years of release. This suggests current rehabilitation may not address deeper structural issues, like the need for better educational interventions to prevent repeat offenses (Seigle et al. 81). Though significant investments are made in rehabilitation programs, approaches like psychiatric hospitalization, residential treatment centers, and boot camps often yield limited results (17).

One significant challenge facing U.S. juvenile rehabilitation programs is the imposition of excessive restrictions on the freedoms of young offenders, which can sometimes hinder their ability to reform. Exploring solutions inspired by ancient Greek and Roman practices may offer valuable insights for overcoming these limitations. Although statistical data on how these ancient societies dealt with juvenile delinquents is limited—largely due to the lack of detailed records—the lessons from classical approaches to juvenile justice remain relevant today, especially given their influence on the U.S. legal system. This paper draws on various alternative sources, including oratory defenses from recorded legal cases, as well as myth and literature, to support this argument. In sections four and five, the paper delves deeper into how ancient Greek and Roman legal principles influenced U.S. juvenile punishment. By revisiting the classical focus on education and rehabilitation over punishment, the U.S. juvenile justice system could potentially diversify its strategies, fostering more effective long-term reform and reducing recidivism.

Law in Ancient Greece

I. Trial Procedure

In the Athenian court of law (*dikasteria* in Greek) litigants would present their cases in open public spaces. The jurors (*dikastai*) would vote on oaths proposed by elders, with the winning oath determining the case's outcome. These courts had no independent judges, and *dikastai* relied entirely on litigants' speeches to quickly decide cases without deliberation (Thür, 147). The number of jurors ranged from 201 to 1501, depending on the case, and only adult male citizens could serve (Todd, 100). Additionally, though Athenian trials were not as structured as modern trials, they followed democratic regulations such as random juror allotment and the use

of secret ballots. Speaking times for both parties were equal and regulated by a *klepsydra* (water clock), with trials lasting from 15 minutes to several hours (Thür, 147).

Greek thought on justice evolved from a retributive focus to a blend of prevention and rehabilitation. Until Protagoras (c. 490–420 BCE), Greek thought leaned toward retributive justice, with Protagoras arguing that punishment should prevent future crimes (Cohen, 174). Plato later introduced the idea that rehabilitation and retribution are not mutually exclusive. He advocated for treating offenders but also accepted the exclusion of "incurable" individuals for the greater good (177). Other intellectuals, such as Diodotus, emphasized governance over severe punishment, while Isocrates argued that education creates good citizens, not strict laws. He believed that virtue came from daily habits rather than written laws (184). Plato reinforced these ideas in *Gorgias*, suggesting punishment benefits the wrongdoer, and in *Laws*, where he proposed early moral education as essential for guiding citizens toward wisdom. For those deemed beyond reform, Plato saw the death penalty as a necessary measure (189).

II. Juvenile Delinquency

Although detailed records on juvenile delinquency in ancient Greece are limited, evidence shows that Greek culture influenced its examination and punishment. Today, scholars classify delinquency into three categories: socialized, neurotic, and psychopathic. Social delinquents often come from families marked by conflict, neglect, or rejection, and commit crimes to fit in with gangs. Neurotic delinquents, typically from middle-class backgrounds, suffer from anxiety and guilt, using crime as a temporary escape from stress. Psychopathic delinquents, usually from abusive and unstable households, are unable to form lasting relationships and feel little guilt for their actions (Garland).

Using the aforementioned modern framework as a lens through which to view Greek crimes, it is likely that psychopathic delinquency was most common among characters of Greek mythology. Greek gods such as Kronos (castrates his father Ouranos) and Hephaestus (chains up his mother Hera) as well as mortal heroes such as Oedipus (kills his father and marries his mother) and Orestes (kills his mother in cold blood) represent crimes rooted in problematic familial relationships. Though such gory, ancient myths may seem far-fetched as of today, the reality of ancient Greece, though lesser in degree, echoed myths. In Athens, for instance, there were laws in place against parental abuse due to sons who would beat their parents and abandon them once they had reached old age. In fact, parent-beating was such a recurrent phenomenon especially in fifth-century Athens that it would even casually be brought up in the plays of the ancient Greek playwright Aristophanes. Though it is unclear as to whether there is a direct correlation between the prevalence of youths committing acts of violence toward their parents in myths, many ancient Greek thinkers, including Plato, suspected so (Garland). Because Plato thought children took away moral values from the stories they were told, he believed in strict censorship, which he makes clear in Book II of the *Republic*.

The ancient Greeks recognized that adolescents were naturally prone to violent and impulsive behavior, making them susceptible to delinquency. Philosophers like Aristotle and the

Spartans shared this understanding, though the Spartans took a unique approach. Believing that all juveniles had the potential for wrongdoing, the Spartans sought to prevent misconduct early by imposing strict discipline and structure. From the age of seven, boys were enrolled in the *agoge*, an educational system that trained them to become skilled warriors and moral citizens. By ages seventeen to nineteen, they joined the *Krypteia*, a secretive military group tasked with terrorizing the *helots*, the enslaved population, to prevent rebellion. Through such practices, Spartans aimed to channel youthful aggression into socially beneficial roles while instilling discipline and order.

Law in Ancient Rome

Roman law was divided into three main categories: *ius civile*, *ius gentium*, and *ius commune*. *Ius civile*, or civil law, applied specifically to Roman citizens and covered statutory laws enforced by authorities like the senate, emperor, or jurists. For example, laws governing the writing of wills were restricted to Roman citizens. *Ius gentium*, or the law of nations, initially addressed interactions between Romans and non-citizens but later expanded to manage relations between Rome and other states. These laws were often used in court cases involving a Roman citizen and a foreigner, as the *ius civile* did not apply. Lastly, *ius commune*, or common law, encompassed more general and universal principles that applied to anyone in Rome, regardless of citizenship (Hall). Unlike *ius civile*, which dealt with specific issues, *ius commune* was broader and more fundamental.

I. Trial Procedure

Ancient Roman trials were more aligned with modern court systems than those of Ancient Greece. In the Roman Republic (509–27 BCE), legal procedures advanced considerably, especially after 80 BCE when jury courts became standard. Before this shift, judges determined compensation for victims, often requiring offenders to repay multiple times the stolen amount. For example, a thief might be required to repay four times the value of what they stole. However, after 80 BCE, juries of up to 75 randomly selected members decided both the verdict and compensation.

Much like modern U.S. courts, Roman trials operated on the principle of "innocent until proven guilty." Both accusers and defendants could challenge jurors, and defendants could choose to represent themselves or hire an advocate, akin to today's lawyers. Cicero, one of the most renowned Roman advocates, was famous not for his legal expertise but for his powerful oratory skills.

Roman trials followed a structured process, starting with opening statements, moving to witness examinations and evidence presentation, and concluding with closing arguments. Witnesses were expected to tell the truth, as perjury was a serious crime. After hearing the case, jurors would vote on the verdict, with a tied vote resulting in acquittal. Sentences ranged from fines and public service to harsher punishments like crucifixion, with lower-class offenders often

subjected to more severe physical penalties, while wealthier individuals typically received lighter sentences (Constitutional Rights Foundation).

II. Guardianship in Ancient Roman Law (*Tutela* and *Curatela*)

Ancient Romans recognized adolescent tendencies toward disobedience despite the focus on obedience in education. They understood that youths lacked full mental maturity and often attributed their mistakes to immaturity and inexperience, leading to lighter punishments. Misbehaviors ranged from minor infractions, like disobedience or theft, to more serious offenses like debauchery or gambling, which were viewed as morally corrupt. While occasional misbehavior was expected, writers like Plutarch (46–120 AD) emphasized the need to moderate and restrain the impulsive tendencies of young men (Antošovská, 12).

Possibly as a result of literature advising more careful supervision on the behaviors of their youth, ancient Romans established a system of guardianship, or tutelage. This system was originally created to protect those citizens who lacked practical or legal experiences due to their age, sex, or mental conditions. A guardian was someone who had taken on a role of leadership and authority over the protected person (Chevreau, 190). In ancient Roman law, there were two kinds of guardianship: *tutela*, meaning tutorship, and *curatela*, meaning curatorship. The first kind, *tutela*, was a system that applied only to Roman adolescents under the age of puberty, which was fourteen for boys and twelve for girls. Male citizens above the age of twenty-five could qualify as a tutor for the boys and girls. Once an individual was appointed as a tutor or a guardian, they could not refuse this role unless they were granted an exception by law. As a result, Roman citizens treated the duty of tutorship as a public burden.

There were four major classes of tutors: testamentary guardians (*tutores testamentarii*), statutory guardians (*tutores legitimi*), guardians appointed by courts (*tutores dativi*), and fiduciary guardians (*tutores fiduciarii*). *Tutores testamentarii*, the first class of guardians, were ones appointed by a father's will, just like the system in English law where a father reserves the right to appoint his child a guardian. *Tutores legitimi*, the second class, were appointed by the government in cases where the father did not appoint his child one. *Tutores dativi*, the third class, were appointed in default of the first and second class of guardianship, and finally *tutores fiduciarii* referred to when fathers had to act as a guardian for their sons themselves. The duties of a tutor, as well as how much power one would hold, were all determined based on how capable the pupil was. The farther away the pupil was from puberty, the less developed their capacities to sufficiently reason and make decisions were, meaning more authority and discretion were given to the tutor.

Once adolescents reached puberty, they were placed in a second system of guardianship, *curatela*. After puberty, adolescents entered the *curatela* system, where curators, aged 25 or older, were appointed to supervise them. This system differed from *tutela* in that other than the clinically insane and individuals engaged in lawsuits or in need of legal help, most were not obligated to be a part of *curatela*. Curators could be both special or general: special curators were appointed for the purpose of a specific affair such as a legal affair or transaction, whereas

general curators had a more holistic role of being appointed over the entire estate of the person in curatorship (Sherman, 126).

In modern law, tutorship is often combined with curatorship. An example of a legal system that includes an amalgamation of the two is the English legal system. In the English code of law, one cannot have full rights over their property until they reach the age of 25, which is also the age in Roman law when an individual officially no longer needs or can seek the assistance of a tutor or guardian. Until one turns 25 years of age, guardians exist to help individuals manage their property per English code of law, and the guardians fulfill the duties of both a tutor and curator in ancient Roman law.

Law in the United States

To understand how the U.S. legal system has been shaped by classical influences, it is essential to examine the trial procedure in federal criminal courts today. This system, while modernized, carries forward some structural elements from ancient judicial systems, such as the use of juries and the emphasis on fairness and due process.

I. Trial Procedure

In U.S. federal criminal trials, twelve jurors are randomly selected from a larger pool, with both the prosecutor and defense excusing a limited number of jurors. The jury must be diverse, and discrimination in selection is prohibited. The trial begins with opening statements, followed by witness examinations and objections. Once the prosecution rests, no further evidence or witnesses are introduced. After closing arguments, the jury deliberates, requiring a unanimous decision for conviction. If acquitted, the defendant is released; if convicted, sentencing occurs months later. Sentences follow federal guidelines, and the death penalty is reserved for capital crimes, although it is banned in many states. The Supreme Court prohibits the death penalty for minors or those with mental incompetence (U.S. Department of Justice).

In U.S. federal criminal trials, twelve jurors are randomly selected from a larger pool, with both the prosecutor and defense excusing a limited number of jurors. The jury must be diverse, and discrimination in selection is prohibited. The trial begins with opening statements, followed by witness examinations and objections. Once the prosecution rests, no further evidence or witnesses are introduced. After closing arguments, the jury deliberates, requiring a unanimous decision for conviction. If acquitted, the defendant is released; if convicted, sentencing occurs months later. Sentences follow federal guidelines, and the death penalty is reserved for capital crimes, although it is banned in many states. Finally, the Supreme Court prohibits the death penalty for minors or those with mental incompetence (U.S. Department of Justice).

II. Juvenile Delinquency

An estimated 423,077 juvenile delinquency cases are adjudicated annually in U.S. courts, with 52% of defendants found guilty in 2018. Juveniles, defined as individuals under 18, can

commit either criminal offenses or status offenses, such as underage drinking. While juvenile arrest rates have steadily declined, dropping 74% from 1997 to 2019, significant challenges remain (Youth.gov). The Supreme Court has acknowledged age-based protections for minors, but until 2005, children could face capital punishment. As of 2023, minors as young as eight can still be tried as adults, with extreme sentences depending on the crime (Equal Justice Initiative). However, juvenile sentencing is generally more lenient than adult sentencing, often involving house arrest, probation, or placement in juvenile detention. Judges also use alternatives like fines, counseling, or community service for minor offenses, with only verbal reprimands in the least serious cases.

The Influence of Classical Law Codes on the U.S. Law Code

There is no doubt that certain legal U.S. principles, procedures, writing, and even teaching have been influenced by those of the classical period, especially ancient Rome. U.S. law professors have referred to the use of Roman law by U.S. courts as “an integral part of the larger jurisprudential process by which American jurists reached back to find a line of argument to be employed in understanding the case” (Buchanan). An example of a modern legal concept that originates from Roman law is tort law, which states that offenders must redress the harm they caused their victim in some way, usually via monetary compensation (McGinley, 82). References to Roman law have also appeared in U.S. court when discussing commercial and trade disputes, agency law, and even admiralty law.

The roots of Roman law are founded in the laws, customs, and beliefs of ancient Greece. Even the very founding philosophy of Roman law—that law is founded upon ethics and that the specific rights and duties of men were derived from principles of natural justice—has its origins rooted in stoicism, an ancient Greek philosophy that was especially popular until the third century AD (71). That there is an inextricable connection between ancient Greek and Roman codes of law, means there is also an automatic connection between ancient U.S. law and ancient Greek law, even though the connection may not be as direct as with Roman law.

There are many more examples of ways in which Roman law has had a degree of bearing on U.S. law. Laws on children and childcare enforced during the reign of emperor Justinian (527-565 AD) still hold true in the modern day United States. Specifically, laws on adoption the U.S. has today were first established by the Roman Civil Code of Law. Justinian also made law to regulate the dominion of a *pater familias*, or the father of the family. Prior to Justinian, the father had absolute power over his family and could determine even the life and death of his child. However, Justinian curbed the father’s power so that he would have limited supervision and corrective power over his child. Justinian also abolished the ancient custom of selling free-born children and introduced the idea that even a child could have ownership over property. Many of these legal customs and principles are still in existence today, demonstrating the longstanding influence of ancient law on modernity (72).

The system of tutelage and curatorship explained in Section III: B is another example of Roman influence on modern law. Similar to how in ancient Rome a tutor was in charge of a

child's affairs from birth to age seven, from which age onward the child was believed to have developed sufficient reasoning capacities, in modern U.S. law, a minor can make transactions beneficial to him, provided that he receives guidance to prevent him from acting to his youthful impulses. Today, those who are legally of age and have sufficient reasoning capacities are entitled more agency in seeking legal guidance compared to ancient Rome, but the act of legal experts offering mentorship and consultation is still a similarity between the two systems.

Roman law has historically played a crucial role in the larger American jurisprudential process, as it offers lines of fundamental argument that the jurists reach back to when understanding or debating the case. An online search of actual references to "Roman law" in the Supreme Court yields 140 case references between 1788 and 1999 (Astorino, 633). In *Dred Scott v. Sandford* (1857), Justice Peter V. Daniel (1784-1860) presided over the notorious case, where he analyzed the issue of slavery in Roman law to draw parallels with slavery in the United States (Buchanan). Other examples of Court cases that made use of Roman law include *Columbian Insurance Company of Alexandria v. Ashby and Stribling* (1839), where the court analyzed Roman law concerning the bearing of risks in shipping contracts, and *Hurtado v. California* (1884), where the Court analyzed the due process of law clause in the Constitution while making references to Roman law. In *Coffin v. United States* (1895), the Court used Roman law as a basis under which they studied the presumption of innocence, and in *Geer v. Connecticut* (1896), the Court analyzed a "Connecticut statute that regulated the killing of game and prohibited its transportation outside of the state boundaries as an exception to the dormant Commerce Clause" under the framework of Roman law. In *Hovey v. Elliott* (1897), the Court reviewed Roman law on the right to appear and be heard and for contempt of court, and in *United States v. Chavez* (1899), the Court reviewed the stance of Roman law on adverse possession.

Research has shown that the Supreme Court still did make use of and references to Roman law during the twentieth century too, albeit less often than during the nineteenth century. In *Knowlton v. Moore* (1900), the Court looked into Roman law on the subject of taxes, and in *Greene v. McElroy* (1959), the Court stressed the importance of the Sixth Amendment, with a reference to Roman law for providing the origins of the rights granted within the amendment that covers the right to confront witnesses (634). Both *Johnson v. Louisiana* (1972) and *Taylor v. Kentucky* (1978) referred frequently back to *Coffin v. United States* (1895), a case particularly striking for the prevalence of references the judges made to Roman law. In *Johnson v. Louisiana*, Justice Byron White's majority opinion upheld a Louisiana statute that allowed criminal cases to have split verdicts in the jury. In *Taylor v. Kentucky*, Justice Lewis F. Powell declared that a jury must be informed of the defendant's presumption of innocence in criminal courts under the Due Process clause. Both cases reestablished the verdict of *Coffin v. United States*, that a defendant in criminal cases is considered innocent until proven guilty. In explaining his decision, Justice Edward White wrote a fifteen paged document, of which three pages were devoted to citing Roman law, and nearly two were on making references to different law codes, including those of Sparta and Athens. White even directly quoted past Roman emperor Julian (331-363 AD) on his views to justify his own verdict: "If it suffices to accuse, what will become of the innocent?"

(638). The impact of Greek and Roman law codes on the contemporary American legal system are evidenced by the numerous references of justices to the laws of these ancient civilizations.

Interestingly enough, there is barely any record of judges in the U.S. legal system citing ancient Roman law for cases of juvenile delinquency. Studying ancient Roman and Greek ways of handling juvenile crime may benefit judges because both civilizations had very unique and intricately structured ways of addressing delinquents and preemptive ways of confronting adolescents prone to committing crime, as showcased in Sections 3D and 4B.

A thorough investigation of various approaches to addressing juvenile crime could provide valuable insights for improving the U.S. juvenile justice system. Notable challenges include high juvenile incarceration rates, reports of abuse in detention facilities, and insufficient attention to the mental health needs of incarcerated youth. Although juvenile arrest and incarceration rates have steadily declined, the United States still held the highest incarceration rate globally as of 2021 (Gramlich). In 2020, an average of 25,014 youths were in juvenile justice facilities each day (Rovner). While the U.S. system emphasizes rehabilitation, the persistently high rates of incarceration suggest that public safety and risk management often take precedence. It remains uncertain whether a deterrence-focused approach is truly effective in reducing recidivism. Research shows that minors placed in adult prisons are 36 times more likely to commit suicide than those in juvenile facilities, and those who do not face a higher likelihood of re-arrest later in life, undermining the idea that harsher environments foster discipline (Equal Justice Initiative). For modern judges, revisiting historical approaches to juvenile crime might offer fresh perspectives, helping them draw on foundational principles and precedents to guide more effective and rehabilitative reform efforts.

Conclusion

The influence of classical law codes on the United States legal system is undeniable, with ancient Roman law, shaped by Greek principles, still relevant in U.S. courts today. Since the eighteenth century, U.S. judges have referenced Roman law for foundational reasoning in various modern cases. However, this practice is much less common in cases involving juvenile delinquency. Given current challenges like high recidivism and ineffective rehabilitation, revisiting ancient methods for addressing juvenile offenders could offer valuable insights and innovative approaches. A deeper understanding of classical models could inspire modern U.S. judges to adopt more effective, rehabilitative strategies for juvenile justice reform.

Works Cited

- Antošovská, Tereza. "Children as Culprits and Criminals: Children in Mischief, Delict, and Crime in the Roman Empire." *ResearchGate*, January 2019. Accessed July 20, 2024.
- Astorino, Samuel J. "Roman Law in American Law: Twentieth Century Cases of the Supreme Court." *Duquesne Law Review*, vol. 40, no. 4, 2002, article 3. Accessed July 27, 2024.
- Buchanan, Kelly. "References to Roman Law in US Courts." *Library of Congress Blogs*, 20 Mar. 2015. Accessed Aug. 18, 2024, <https://blogs.loc.gov/law/2015/03/references-to-roman-law-in-us-courts/>.
- Chevreau, Emmanuelle. "The Evolution of Roman Guardianship through the Mechanism of *Excusatio Tutelae*." *OpenstarTs*. Accessed June 30, 2024.
- Cohen, David. "Theories of Punishment." *The Cambridge Companion to Ancient Greek Law*, edited by Michael Gagarin and David Cohen, Cambridge University Press, 2005. Accessed July 1, 2024.
- Constitutional Rights Foundation*. "BRIA 17 4 b: When Roman Law Ruled the Western World." Accessed July 30, 2024, <https://www.crf-usa.org/bill-of-rights-in-action/bria-17-4-b-when-roman-law-ruled-the-western-world#:~:text=The%20trial%20procedure%20would%20be,evidence%20about%20the%20%20defendant's%20character>.
- Equal Justice Initiative*. "Children in Adult Prison." Accessed August 5, 2024, <https://eji.org/issues/children-in-prison/>.
- Gagarin, Michael. "Early Greek Law." *The Cambridge Companion to Ancient Greek Law*, edited by Michael Gagarin and David Cohen, Cambridge University Press, 2005. Accessed July 10, 2024.
- Garland, Robert. "Juvenile Delinquency in the Graeco-Roman World." *History Today*, vol. 41, no. 10, Oct. 1991. Accessed July 12, 2024.
- Gramlich, John. "America's Incarceration Rate Falls to Lowest Level Since 1995." *Pew Research Center*, 16 Aug. 2021. Accessed Aug. 1, 2024, <https://www.pewresearch.org/short-reads/2021/08/16/americas-incarceration-rate-lowest-since-1995/#:~:text=As%20crime%20has%20declined%2C%20so,number%20of%20people%20behind%20bars>.
- Hall, Eamonn G. "The Contribution of Roman Law to Modern Legal Systems." *VoegelinView*. September 19, 2019. Accessed July 20, 2024, <https://voegelinview.com/the-contribution-of-roman-law-to-modern-legal-systems/>.
- Justia*. "Criminal Sentencing in Juvenile Court & Possible Legal Penalties." Accessed July 7, 2024, <https://www.justia.com/criminal/offenses/other-crimes/juvenile-crimes/sentencing-in-juvenile-court/#:~:text=They%20may%20order%20the%20juvenile,go%20with%20a%20verbal%20reprimand>.
- McGinley, Gerald J. "Roman Law and Its Influence in America." *Notre Dame Law Review*, vol. 3, no. 2, 1927. Accessed July 10, 2024.

- Rovner, Joshua. "Youth Justice by the Numbers." *The Sentencing Project*, 16 May 2023. Accessed July 27, 2024, <https://www.sentencingproject.org/policy-brief/youth-justice-by-the-numbers/>.
- Seigle, Elizabeth, Natassia Walsh, and Josh Weber. "Core Principles for Reducing Recidivism and Improving Other Outcomes for Youth in the Juvenile Justice System." *The Council of State Governments (CSG) Justice Center*. Accessed July 15, 2024, <https://csgjusticecenter.org/publications/reducing-juvenile-recidivism/>.
- Sherman, Charles P. "The Debt of the Modern Law of Guardianship to Roman Law." *Michigan Law Review*, vol. 12, no. 2, 1913. Accessed July 2, 2024.
- Todd, S. C. "Law and Oratory at Athens." *The Cambridge Companion to Ancient Greek Law*, edited by Michael Gagarin and David Cohen, Cambridge University Press, 2005. Accessed July 7, 2024.
- Thür, Gerhard. "The Role of the Witness in Athenian Law." *The Cambridge Companion to Ancient Greek Law*, edited by Michael Gagarin and David Cohen, Cambridge University Press, 2005. Accessed July 4, 2024.
- United States Department of Justice. "Sentencing." *Executive Office for United States Attorneys*. Accessed August 2, 2024. <https://www.justice.gov/usao/justice-101/sentencing>.
- United States Department of Justice. "Steps in the Federal Criminal Process." *Executive Office for United States Attorneys*. Accessed Aug. 4, 2024, [https://www.justice.gov/usao/justice-101/trial#:~:text=The%20trial%20is%20a%20structured,committed%20the%20crime\(s\)](https://www.justice.gov/usao/justice-101/trial#:~:text=The%20trial%20is%20a%20structured,committed%20the%20crime(s)).
- Youth.gov. "Youth Involved with the Juvenile Justice System." Accessed Aug. 6, 2024. https://youth.gov/youth-topics/juvenile-justice/youth-involved-juvenile-justice-system#_ftn.

“Friendship Fallout: Teen Mental Health” By Syeda Rubab Bokhari

Introduction

Every teenager experiences a wide range of emotions and grows through their teenage years at their own pace, facing several new challenges, but along the way they also make new friendships. The significance of these teen friendships are in their ability to improve emotional growth and social skills. These friendships provide teens with a sense of belonging as well as understanding, as sometimes teens share thoughts with their friends that they wouldn't share with their parents. As a result, among teenagers, friendship breakups can cause a range of unwanted emotions, including confusion, hurt, and anger. Around the world, teenagers may struggle to cope with the fallout from friendship breakups since so many strong emotions might cause mental instability such as anxiety, depression or strong feelings of loneliness.

Causes and Consequences

Friendships eventually wilt and fade, just like leaves do on trees in the fall. Our friendships evolve with us as we mature and become unique individuals throughout time. In a study by Apostolou and Keramari (2021), 577 youths who were 18 years of age and older, identified the four major causes of friendship breakups: perceptions, selfishness, infrequent interaction, and romantic involvement. A friendship is always about give and take so that it can be fair for all parties. However, if one person consistently gives and gets little in return, friendships can become one-sided. Friendships could end as a result of this. Friends that don't meet often tend to drift away as they form new friendships. So, infrequent communication might lead to the end of friendships. Furthermore, romantic involvement ruins friendships as many new complex emotions are introduced. If your friend is interested in you but you aren't interested in them or if your friend is interested in someone you like or are with, this can cause rifts in the friendship, make it weak and eventually friendships tend to break as they become toxic. Lastly, perceptions can also result in the end of a friendship if one's family or other friends don't approve of the new friend.

Teenage friendships often end with feelings of betrayal, sadness, rage, and sometimes even numbness. Two teenagers can be such close friends one minute, and then just two people acting like they don't exist for the other. Teens in this situation may experience anxiety and depression, develop trust issues, feel insecure, isolate themselves from others, and, in the event that the friendship becomes extremely toxic and the teen is emotionally manipulated, they may even consider suicide or start self-harming because they feel unworthy and alone.

Perspectives

Losses among teenagers are very common and 5% western teenagers experience the loss of a friendship before they turn fifteen. In a study carried out by Kaitlin M. Flannery, an assistant professor in the Department of Psychology at SUNY Cortland and Rhiannon L. Smith, an associate professor in the Department of Psychological Sciences at the University of

Connecticut, the experiences of friendship breakups in teenagers were examined. The results of the study were that teenagers felt a mixture of emotions when they lost their friends and their emotional reactions could differ based on who ended the friendship, why it ended and how it ended.

According to 'Tweak India', one of their readers, Smitri Iyer wanted to "be sad for days, and months, on end" when she went through a friendship breakup.

Furthermore, although the mental health issues faced by teenagers are given world-wide importance and recognition, in Pakistan, this issue still remains a neglected area in both research and health policy of the country. According to a cross sectional sample of 1124 teenagers recruited from schools in Rawalpindi, Pakistan, 17.2% and 21.4% participants were identified to be probable cases of depression and anxiety, respectively. A recent survey revealed that 34% of Pakistan's youth are suffering from depression or anxiety disorder throughout the country. These high levels of anxiety and depression could be caused by exposure to toxic households, negative events in life or any kind of abuse. However friendships breakups may also be the cause of mental health issues in some teenagers.

To prove this claim, I carried out a survey which included questions with a yes/no answer related to my topic: 'Friendship Breakup'. Teenagers from the age of 13-18 were sent this survey. I received a total of 105 responses and the most common response given by most of these teenagers was that friendship loss played a huge role in affecting their mental health negatively. Many of them fell into depression for a while as they had lost a person they thought they'd be close to their entire life. Apart from depression, they felt lonely, unworthy and insecure as they thought they wouldn't be able to find close friends again for a long time.

Course of action

The following are a few ways to mitigate the negative impact of broken friendships on a teenager's mental health:

- 1) Teens should share their negative thoughts caused by friendships breakups with people online or in person, whom they are comfortable with. This could lessen the burden on their shoulders as their feelings are being shared and even understood by many, resulting in them feeling less heavy.
- 2) The one thing teenagers need to be able to move on from a friendship breakup is closure. They should give themselves the time to grieve their loss. To feel the sadness and pain. But then they should understand that what has happened won't change, and they need to move on in order to be able to give themselves the closure they need.
- 3) While couple therapy and individual therapy have been around for a long time, 'friendship therapy' is a new concept in which two or more friends, who are on the brink of losing their friendship, go for therapy sessions together in order to solve their differences and conflicts that led to their friendship's downfall to be able to strengthen

their friendship again. Teenagers can opt for this type of therapy to be able to avoid the mental health issues friendship breakups would cause them.

- 4) Exploring new hobbies and discovering new interests can help teens widen their social circle by meeting new people, and they can also make new friends. This will not only aid them in moving past the pain caused by a friendship breakup, but will also contribute towards personal development. Teens can start by making a list of all things they've wanted to try out, and can take the first step towards carrying out those activities.

Conclusion

To conclude, the effects broken friendships have on a teenager's mental health are very unpleasant and must be taken seriously. I was aware of the impact broken friendships had on one's mental health before choosing this topic for my report and starting my research, but I soon realized that these impacts weren't talked about enough. Thus I intended to shed light on this topic through my report.

In times like these, when teenagers are going through complex emotions and mental turmoils, we must remember the message given by UNICEF, "The feedback from youth is clear in that they are aware of their need for mental health support. Encouragingly, an increasing number of children and young people are engaging with their friends on mental wellbeing. We must ensure that national policy and community-based mental health support services put youth at centre."

Personal Reflection

I chose this topic for my report because, as a teenager, I myself have gone through friendship breakups, that left me feeling alone and insecure. Through my report, I want all the teenagers out there to know that they are not alone and that they are understood. While researching, I came across a lot of facts and statistics that I hadn't known before. Thus, I strongly believe that through conducting the research and preparing this report, I have strengthened my emotional intelligence, and have fully understood the seriousness of mental health issues. These issues can be extremely dangerous, and therefore, should be treated immediately if they are being experienced by a person.

Works Cited

- 7 Things That Kill Friendships. (n.d.). Retrieved from AnydayGuide: <https://anydayguide.com/post/650>
- Adolescents' mental health and well-being in developing countries. (2019). Retrieved from PubMed.Gov: <https://pubmed.ncbi.nlm.nih.gov/30451053/>
- Bockarova, M. (2022). 55 reasons why friendships end. Retrieved from Psychology today: <https://www.psychologytoday.com/us/blog/romantically-attached/202109/55-reasons-why-people-end-friendships#:~:text=The%20most%20significant%20factors%20in,giving%2C%20among%20the%20prime%20reasons.>
- Greenberg, B. (2021). 4 Things That Kill Friendships. Retrieved from Psychology today: <https://www.psychologytoday.com/us/blog/the-teen-doctor/202204/4-things-kill-friendships>
- Helping your teen through a friendship breakup. (2022). Retrieved from Pacific teen treatment: <https://pacificteentreatment.com/mental-health/helping-teen-through-friendship-breakup/>
- Kumar, P. (2022). Mental health. Retrieved from Tribune: <https://tribune.com.pk/letter/2253091/mental-health-1>
- Singh, R. (2022). Why the end of a friendship can hurt even more than your favourite boy band breaking up. Retrieved from Tweak India: <https://tweakindia.com/culture/discover/why-does-a-friendship-breakup-hurt-so-much/>
- Smith, K. M. (2021). Breaking Up (With a Friend) Is Hard to Do: An Examination of Friendship Dissolution Among Early Adolescents. Retrieved from SAGE Journals : <https://journals.sagepub.com/doi/10.1177/02724316211002266>
- Teen depression; a candid interview with teenager Kevin Breel. (2013). Retrieved from Mental Health Talk: <https://mentalhealthtalk.info/teen-depression>
- Tømmeraaas, I. J. (2022). Attachment and grief in young adults after the loss of a close friend. Retrieved from BMC Psychology: <https://bmcpublishing.biomedcentral.com/articles/10.1186/s40359-022-00717-8>
- Why Ending a Friendship Can Be Worse Than a Breakup. (n.d.). Retrieved from Time: <https://time.com/5402304/friendship-breakups-worse-romantic/>

Artificial Intelligence and the the Chrome Dino Game By Kevin Hsieh

Acknowledgements

Thank you for the guidance of Simon Alford, mentor from Cornell University in the development of this research paper.



I. INTRODUCTION

This research paper we experimented with different reinforcement learning algorithms and found different ways to train an artificial intelligence to play the Chrome Dino game. Our goal is to find different ways that we can allow the model to achieve the best performance. We tried training the model in different ways and managed to allow the dinosaur to be able to successfully jump over cacti, duck under birds, and jump over birds. We discovered that an evolutionary modification to Proximal Policy Optimization (PPO) worked best. We aimed to answer the question of what should be used to allow artificial intelligence to learn and play the game. We were trying to find a general algorithm that can be used across different games or projects besides the Chrome Dino game that uses reinforcement learning.

Related Work. Artificial Intelligence has been rapidly growing in the past few decades. Researchers discovered different algorithms that can be used for different topics. However, it is hard to create a general algorithm that can work in diverse environments because the AI has to be trained on specific data for the specific goal. Training artificial intelligence also takes a large amount of time. Because of this, researchers try to find ways to improve the generalization and efficiency of models. John Schulman proposed to do multiple gradient updates to every sample. Using new methods which they called proximal policy optimization—also known as PPO, they could outperform other methods by being simpler to implement, more general, and have better sample complexity (Schulman et. al., 2017). However there are many other algorithms. For example, Mastering Diverse Domains through World Models presents DreamerV3 which is a general algorithm which can be used over 150 different tasks. It was used to mine diamonds in minecraft using reinforcement learning. It used PPO as it has become a standard in reinforcement learning (Hafner et. al., 2023). Another paper, Playing Atari With Deep Reinforcement Learning, used reinforcement learning algorithms to play the Atari game. It combined deep learning with reinforcement learning and fed the computer frames of the game using convolutional neural networks. Using their algorithm, their AI could play six of the seven games it was tested on with no adjustment of the algorithms and parameters (Mnih et. al., 2013) . Additionally, an AI using reinforcement learning was trained to be able to play Go. The AI was able to achieve a 99.8% winning rate against other programs and professional human players (Silver et. al., 2016) .

II. METHODOLOGY

Plan. Our first step was to set up the environment for artificial intelligence. The experiment applied deep reinforcement learning algorithms in order for the model to learn. We used a combination of neural networks and reinforcement learning. The neural network used proximal policy optimization and stable baselines3 for learning. Stable baselines3 is a package of deep reinforcement learning algorithms. The game which was played was the Chrome T-Rex Rush which is a game that Chrome allows you to play when you lose internet connection. The Chrome T-Rex Rush was from an open source repository made by Shivam Shekhar (Shekhar. 2023) . In order for the AI to be able to input and gather data from the game. There is a channel used to channel data from the game to the AI. Gymnasium, which is a library for deep reinforcement learning environments, was used to create the environment which served as a gateway between the game and the AI.

Setup. After creating a custom environment for the Chrome T-rex game. The game was converted to be able to play a single frame of the game upon a function call. A channel was created to hold variables that were updated with game data which was given to the neural network upon each step. Data that was given was the x location of the dinosaur, the distance of the nearest bird, and the distance of the nearest cactus. The dinosaur's reward system was set up so that for every successful jump over the bird, it would gain a 1 reward upon every successful jump over a cactus or a bird. Upon death, the model would receive a -10 punishment.

Data Collection. Data collection was added in order to gather the data of the average lengths and rewards for every rollout of the game. The dinosaur was trained overnight and data was collected. Afterwards, the dinosaur was trained again over 3 nights (approx. 12 hours per night) and data was collected afterwards. This training process was repeated again. Following this, the seed of the game and the AI model was changed to not be randomly generated in hopes to allow more accurate data collection. We can see the data that was collected in the data graphs shown in the results below.

Experimentation. The speed of the game was also changed to be constant. The model was trained for a while and the ent-coef was increased to 0.01 when it began underperforming and getting stuck at a local minimum. Data was then collected. Afterwards, optimized parameters for CartPole from RL-baselines3-zoo (RL Baselines3 Zoo: A Training Framework for Stable Baselines3 Reinforcement Learning Agents, 2022) were then used for the dinosaur. The dinosaur was then trained over one night and data was collected. The dinosaur was then trained again over night and data was collected. After this, the dinosaur's observation space was increased from its current observations to include the bird's Y location and the distance between the next two obstacles. The distance between the bird and the dinosaur was also changed to be calculated through just the X value rather than the pythagorean theorem as the bird's Y value changes. The model was trained overnight and its performance data was collected. Next, the dinosaur's environment was changed to be able to support multiple environments running at the same time (Figure 7).

Evolutionary Algorithm. A new algorithm was created which hoped to imitate evolution. The best environment's model would be chosen every ~5 minutes and replaced with the best performing model. It would find the model with the most survival time and replace all other models with it and then

continue training. The models would also be backed up to its previous version if its performance dropped below 50% of its survival time length. Additionally, loading and saving models multiple times for the evolutions resulted in the model being no longer able to save and load. In order to combat this issue, it was made so that just the model's policy was saved and loaded during the evolutions. The model was then trained overnight on Macbook with 16 cores training 6 environments simultaneously and data was collected the following morning. The bird spawning was then increased to have a 50% spawn rate every frame with a maximum of one bird per frame. In order to compare the performance of this evolutionary algorithm. It was run multiple times with different amounts of environments. Data was collected for the different amounts of environments.

Evolutionary Algorithm. The dino was never able to figure out how to duck under birds. Even when the bird was too high for the dino, it still tried to jump over it and die. In order to get the dino to duck, an incentive was added. The dinosaur was given 1000% its successful jump reward for every frame it was under a bird that was equal to the Y level of the highest bird. After changing the death penalty to -1000 and the dino a reward of 100 for every second it was ducking, it finally learned to duck under birds. Its performance can be seen in Figure 13.

III. RESULTS

The first training session used the observations of only the dino X location, nearest bird distance, nearest cactus distance. The length of the training was around ~36 hours. The result was that the dinosaur learned to jump over cactuses but eventually lost its ability to do so after continual training.

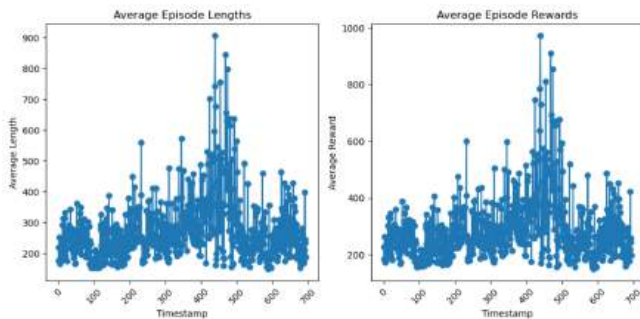


Figure 1.

In the second training session. The same observations and time was given as the first training session. The result was a slow training speed and then an eventual learned ability to be able to jump over cactuses. It was similar to the first session but slower.

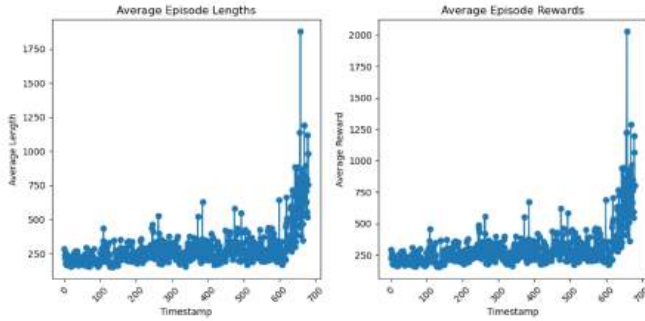


Figure 2.

In the third training session, the same parameters were given. It was training for a longer period of time. However, it suddenly lost all its skill as shown in timestamp ~1500 and eventually just tried to die as fast as possible by killing itself with the first cactus. This is shown by the second drop to its lowest point at timestep ~3000. In an attempt to salvage the model, the entropy coefficient was increased to 0.01 to encourage it to explore at timestamp ~3,400. This resulted in the model jumping over cactuses but performing poorly continuously.

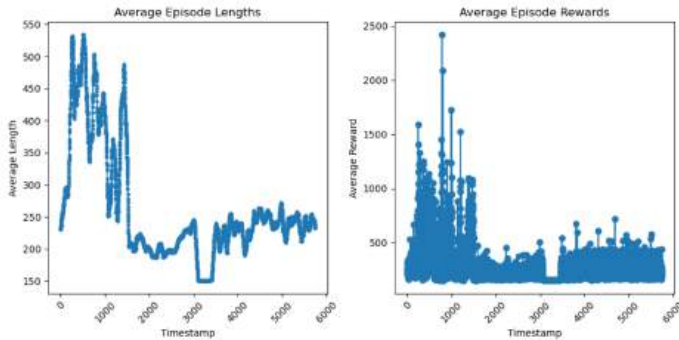


Figure 3.

In the fourth training session using the same parameters, the model was trained over 24 hours using the same observations but beginning with a entropy coefficient of 0.01. This resulted in an unstable model that never reached a good performance.

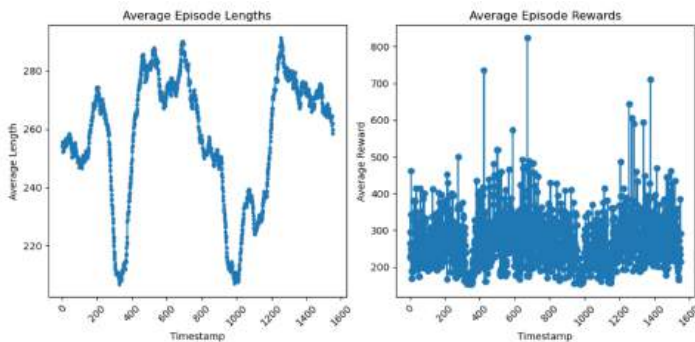


Figure 4.

In the fifth training session, the entropy coefficient was changed back to zero. The parameters were set to the optimized cartpole's parameters and then trained for 12 hours. Again, this resulted in an unstable model that performed poorly.

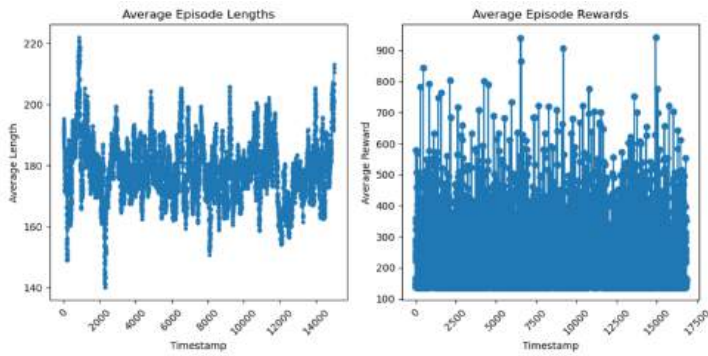


Figure 5.

In the sixth training session, the experiment was repeated with the optimized cartpole parameters but with an entropy coefficient of 0.01. The performance increased a bit but still performed fairly poor.

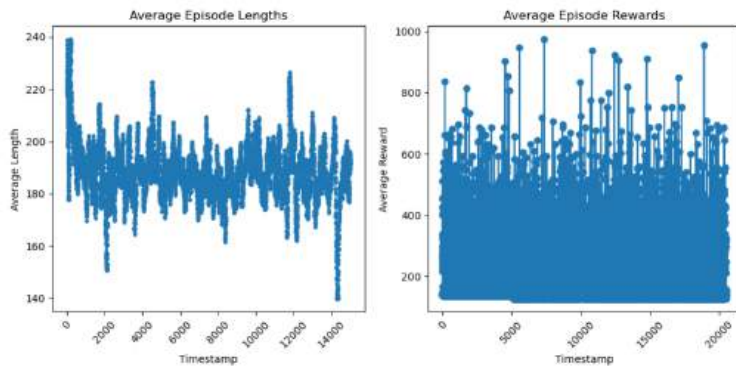


Figure 6.

In the seventh training session, the bird's Y location and obstacle distances were added to the model's observation space. This was also when multiprocessing and evolution was introduced. It was trained over 3 hours with 6 environments running simultaneously. The result was a continuous increase in performance.

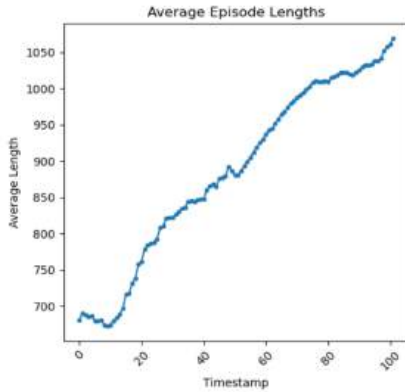


Figure 7.

It was discovered that using evolution, multiprocessing, and more observations, we were able to achieve a better performance in a shorter time span. In the following training sessions, we trained the models using different amounts of environments. Birds were also made to spawn with a 50% chance per frame.

The first run was with a single environment with no multiprocessing or evolution which ran for ~13 hours. This allows us to see the effectiveness of just increasing the observation space. The result was that it reached a similar performance to before the observation space was changed. However, its performance still improved as the model had to learn how to jump over birds along with the cacti, while the model did not have to before the observation space was increased.

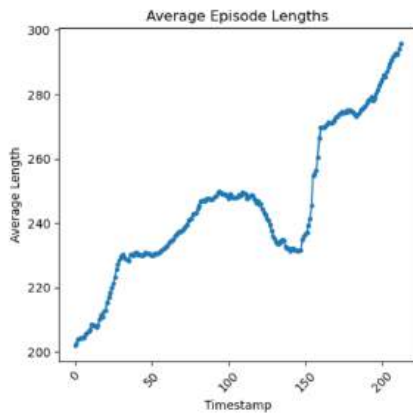


Figure 8.

The model was then run with two environments using multiprocessing and evolution for ~11 hours. However, if we just compare the models at time step 150. The model with two environments performed better as the first model reached 230, while the second model reached 350.

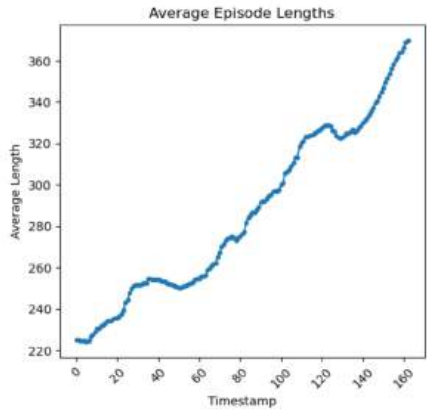


Figure 9.

Following this, another model was then run with four environments for ~14 hours. If we compare the models at time step 150. The model with two environments reached 350, while the model with four environments reached a little less than 600.

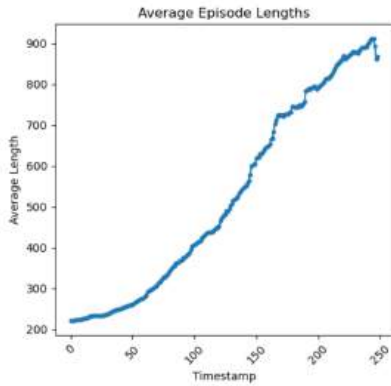


Figure 10.

The model was then trained with 6 environments. However, this is when the model's performance begins decreasing. It performs similarly to the model with 2 environments running.

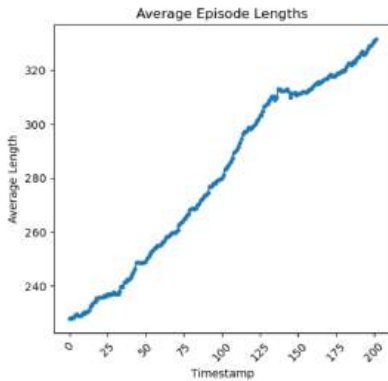


Figure 11.

Following the previous model, another model was run with 8 environments. This model reached 600 at time step 150 similarly to the environment with 4 models.

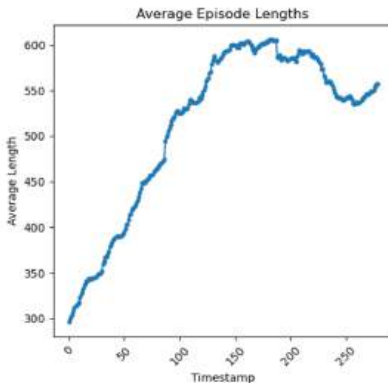


Figure 12.

A model was then trained with 16 environments. Its performance speed was much quicker than the previous models, however its performance began dropping after it reached time step 100 to lower than the previous three models.

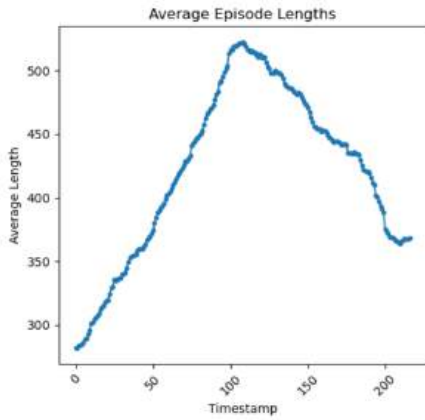


Figure 13.

After giving the dino an incentive to duck. It was finally able to duck which caused it to achieve a much higher performance than it had ever before.

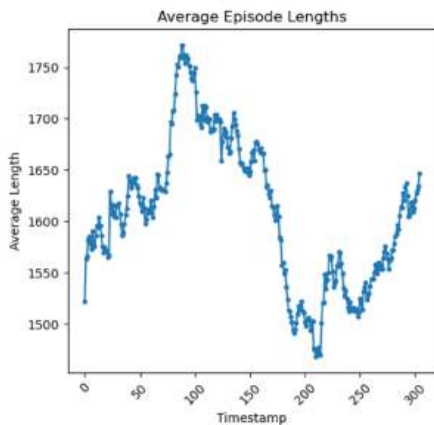


Figure 14.

After observing the data collected from this experiment. We can conclude that the model performs better and more efficiently with the evolutionary algorithm compared to without the algorithm. However, the data does not have a clear enough trend to allow us to understand how many environments result in a better performance.

IV. DISCUSSION

General Analysis. We can observe from Figure 1. And Figure 3. that the model would train for a certain amount of time and lose its capability. I hypothesize that the reason behind this is that the model learns to jump over a cactus and then reaches the section where the bird spawns. This would then result

in the model getting confused as it is introduced to a new obstacle that it has not seen before. It then attempts to experiment which results in the model getting confused with jumping over cacti. In order to combat this problem, the bird's spawning rate was increased and the model was made to take in more observations of the bird. As we can observe from Figure 8, this improves the ability for the model to learn to overcome this obstacle and seems to remove the sudden dip in ability. However, the model would need to be trained for a longer period of time to determine whether this is truly the factor which is causing its performance to drop. We can see that the increase in the entropy coefficient can be also used to prevent the model from finding a local optimum and staying there resulting in a low performance.

Evolutionary Algorithm Analysis. Besides this, we can see from the results that the model's performance increases with the evolutionary algorithm. The reason it performs better is probably because it allows the model to experiment and the model that finds does something correct would then replace all other models, while the model which performs poorly would then be removed from the evolution. This would then result in a continual growth in ability unless all models suddenly perform worse. Since the computer it is being trained on only has 16 cores, it should not be able to truly multi-process more than 16 environments. Figures 8-11 attempts to see if the amount of environments is truly affecting the model's performance. We can see a linear growth with environments 1, 2, and 4 as shown in Figures 8-10. However, the performance then decreases with 6 environments and onwards as shown in Figures 11-13. This shows that 4 environments seems to be the optimal, however there does not seem to be any reason that it is so. This would require further research in order to understand the amount of environments' equivalent performance. Despite this, it does show that this algorithm seems to be helpful when trying to increase the performance and speed of the reinforcement learning model.

V. CONCLUSION

In conclusion, this project shows that evolutionary algorithms should be further researched as it could provide a more efficient way for reinforcement models to learn. Further studies should experiment and discover what is the optimal amount of environments and ways to shape the evolutionary algorithm in order to improve learning. This experiment was limited in time which resulted in more unstable conclusions. I hope that future research could have more time to experiment in more detail and discover ways in which model training can be improved.

Bayesian Optimization. I came up with an idea to have a reinforcement learning algorithm, the "teacher," train the training algorithm, the "student," through setting its parameters. The teacher would try to optimize the training parameters such as reward amount, amount of environments, ent_coef, etc and getting the dino's score as a reward. As a first step towards training a teacher, I looked into hyperparameter tuning of the RL algorithm, aka discovering the best setting of the RL game to train the dino with. To conduct hyperparameter tuning, I used Bayesian Optimization (Bayesian Optimization, 2023). Bayesian Optimization allows us to optimize the parameters for the RL algorithm by modeling a distribution over best parameters and searching parameters most likely to improve the system. However, I unfortunately ran out of time to complete this experiment. I hope to develop this more in the future.

References:

1. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," arXiv.org, 2017. <https://arxiv.org/abs/1707.06347>
2. D. Hafner, J. Pasukonis, J. Ba, and T. Lillicrap, "Mastering Diverse Domains through World Models," arXiv.org, Jan. 10, 2023. <https://arxiv.org/abs/2301.04104>
3. V. Mnih et al., "Playing Atari with Deep Reinforcement Learning," arXiv.org, 2013. <https://arxiv.org/abs/1312.5602>
4. D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," Nature, vol. 529, no. 7587, pp. 484–489, Jan. 2016, doi: <https://doi.org/10.1038/nature16961>.
5. S. Shekhar, "shivamshekhar/Chrome-T-Rex-Rush," GitHub, Sep. 06, 2023. <https://github.com/shivamshekhar/Chrome-T-Rex-Rush>
6. "RL Baselines3 Zoo: A Training Framework for Stable Baselines3 Reinforcement Learning Agents," *GitHub*, May 23, 2022. <https://github.com/DLR-RM/rl-baselines3-zoo>
7. "Bayesian Optimization," GitHub, May 21, 2023. <https://github.com/bayesian-optimization/BayesianOptimization>

Title : Efficient Denoising Method to Improve the Resolution of Satellite Images

Submitted by: Jhanavi Hegde

Abstract

Satellites are widely used to estimate and monitor ground cover, providing critical information to address the challenges posed by climate change. High-resolution satellite images help to identify smaller features on the ground and classification of ground cover types. Small satellites have become very popular recently due to their cost-effectiveness. However, smaller satellites have weaker spatial resolution, and preprocessing using recent generative models made it possible to enhance the resolution of these satellite images. The objective of this paper is to propose computationally efficient guided or image-conditioned denoising diffusion models (DDMs) to perform super-resolution on low-quality images. Denoising based on stochastic ordinary differential equations (ODEs) typically takes hundreds of iterations and it can be reduced using deterministic ODEs. I propose Consistency Models (CM) that utilize deterministic ODEs for efficient denoising and perform super resolution on satellite images. The DOTA v2.0 image dataset that is used to develop object detectors needed for urban planning and ground cover estimation, is used in this project. The Stable Diffusion model is used as the base model, and the DDM in Stable Diffusion is converted into a Consistency Model (CM) using Teacher-Student Distillation to apply deterministic denoising. Stable diffusion with modified CM has successfully improved the resolution of satellite images by a factor of 16, and the computational time was reduced by a factor of 20 compared to stochastic denoising methods. The FID score of low-resolution images improved from 10.0 to 1.9 after increasing the image resolution using my algorithm for consistency models.

Introduction

Satellites play a crucial role in tracking and understanding climate change by providing images that can be used to monitor ground cover, offering critical information for environmental management, agriculture, forestry, and urban planning. High-resolution satellites like WorldView, GeoEye, and SPOT provide detailed images that can identify smaller features on the ground. These images are essential for precise mapping of ground cover types, such as distinguishing between different types of crops, forest species, or urban features. Figure 1 shows key Climate Change Indicators that can be monitored using satellite data (S. Zhao, 2023):

Vegetation Monitoring: Detect changes in forest cover, deforestation, and desertification using NDVI (Normalized Difference Vegetation Index) and other vegetation indices.

Ice and Snow Cover: Monitor changes in glaciers, polar ice caps, and snow cover to assess the impact of global warming.

Urban Expansion: Detect urban sprawl and changes in land use, which contribute to heat islands and increased carbon emissions.

Water Bodies: Monitor changes in the size and health of lakes, rivers, and coastal regions,

which can indicate rising sea levels and droughts.

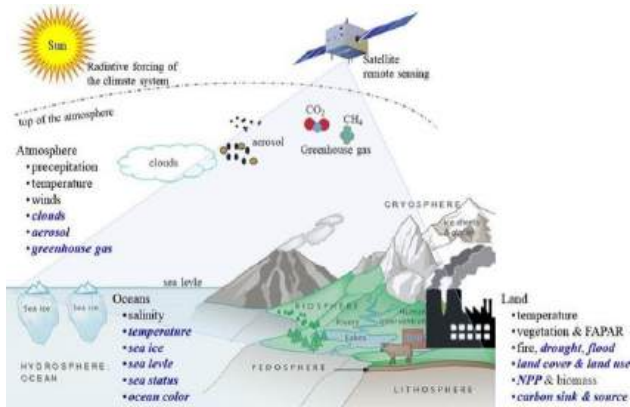


Figure 1 Climate Change Indicators

Computer vision (CV) is increasingly employed to monitor climate change by analyzing satellite data. This approach uses advanced image processing and machine learning techniques to extract critical insights from vast datasets generated by Earth-observing satellites. Machine learning models provide significant advantages over traditional handcrafted CV methods by automatically learning from data, adapting to various tasks, and managing complex patterns and large datasets. Deep learning architectures, in particular, excel in object detection, classification, segmentation, and change detection in satellite images. Given the immense volume of satellite data, machine learning models—especially those designed to handle large datasets—efficiently process and analyze this information at scale, offering substantial time savings with improved performance compared to traditional methods, making them essential for climate monitoring.

There are several practical challenges in using computer vision techniques on satellite images to achieve desired results. The performance of computer vision methods depends on the

resolution of images, and deep learning methods, such as Convolutional Neural Networks (CNNs) provide higher accuracy with high-resolution images. Cost-effective smaller satellites are becoming increasingly popular, but they often have lower spatial resolution, limiting the detection of small objects. Additionally, clouds and rain can obstruct the view, making it difficult to capture clear images and challenging to ensure high-quality images across diverse environments and conditions.

Pre-processing is crucial for enhancing the resolution of images to achieve optimal performance from machine learning models after initial image acquisition. Several methods are available to improve the resolution of satellite images, and generative models are increasingly being used for their ability to generate realistic images from degraded ones. DDMs have revolutionized the image Super-Resolution (SR) field, significantly narrowing the gap between image quality and human perceptual preferences (J.Ho, 2020). DDMs use conditional image generation and perform super-resolution through a stochastic denoising process. They are relatively easier to train and can produce high-quality samples that surpass the realism of those generated by previous methods such as GANs.

DDMs learn to transform a standard normal distribution into an empirical data distribution through a sequence of refinement steps. DDMs typically use a U-Net model (O. Ronneberger, 2015) that is trained with a denoising objective, iteratively removing varying levels of noise from the output. The SR3 model (C. Saharia, 2021) uses Denoising Diffusion Probabilistic Models (DDPMs) for conditional image denoising by introducing a simple yet effective modification to the U-Net architecture. The SR3 model outperformed traditional regression

models and GANs in producing super-resolution images. However, SR3 requires significant computational resources for denoising in the pixel space and takes several iterations to generate images. Recently, Stable Diffusion (R.Rombach, 2022), a Latent Diffusion Model (LDM) developed by researchers at Ludwig Maximilian University in Munich, implemented the denoising process in a compact latent space, significantly reducing computation. This LDM can be used for various image generation tasks, including super-resolution conditioned on low-resolution images.

Stable Diffusion enables high-quality image synthesis while avoiding excessive computational requirements by training the diffusion model in a compressed, lower-dimensional latent space. Figure 2 shows Stable Diffusion consisting of three components: the Variational Autoencoder (VAE), U-Net, and an optional text or image conditioning module. The VAE encoder compresses the image from pixel space to a smaller-dimensional latent space by removing perceptually irrelevant details and retaining only the fundamental semantic content of the image. This reduction in dimensionality aids in the super-resolution of high-dimensional images. For example, a color image with a 512x512 resolution has 786,432 possible values, while Stable Diffusion compresses it into a latent space that is 16 times smaller, with only 49,152 values. The U-Net model is used for noise removal, and denoising occurs in this compact latent space.

The main steps involved in an LDM (Latent Diffusion Model) are as follows:

Encoder: An input image is passed through an encoder model, which compresses it into a smaller latent representation. This latent code

captures the most important features and semantics of the image in a compact form.

Latent Diffusion: The denoising diffusion process is applied to the latent code rather than directly on the pixels. This allows the DDM to manipulate the image in a more controlled way by only modifying the latent code.

Decoder: Once the diffusion process modifies the latent code to generate the desired output image, a decoder model transforms the latent code back into pixel space, reconstructing the final high-resolution image.

LDMs can increase image resolution by 2x, 4x, or even higher by taking a low-resolution image as input, encoding it into the latent space, modifying the latent code, and then decoding it back into pixels. Stable Diffusion uses a Variational Autoencoder (VAE) to encode the image x_0 into a smaller-dimensional latent space z_T . The encoder preserves key details from the original image, which the decoder then expands into a highly realistic super-resolved version. The U-Net model is responsible for denoising, predicting a denoised image representation from noisy latents z_T . The actual latents z_0 are obtained by subtracting the predicted noise from the noisy latent z_T . This denoising step can be guided flexibly by conditioning it on a text or a blurred variant of the same image. To perform super resolution, the denoising step was conditioned on these low-resolution images. This approach—using a universal autoencoder once and then reusing it for different image generation tasks—is highly effective for improving the quality of satellite images with different resolutions, contrast and brightness. The pretrained Stable Diffusion model is used

as a base and fine-tuned for enhancing satellite image quality.

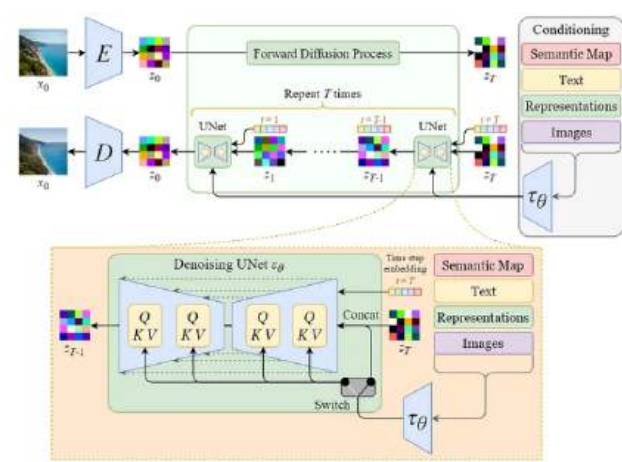


Figure 2. Stable Diffusion

The scope of this project is limited to modifying the DDM in Stable Diffusion to reduce computational time during denoising and implementing an inference pipeline by using the existing decoder to convert latents into RGB images. It is important to explain how the DDM works, followed by a description of the improvements made to the denoising process. The three main aspects of the DDM are outlined below:

Forward Diffusion Process: The model first corrupts the data by adding noise over several time steps. This is a Markovian process, where at each time step, Gaussian noise is added to the data, progressively degrading it into pure noise. This process can be described mathematically as:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{\alpha_t}x_{t-1}, (1 - \alpha_t)\mathbf{I})$$

Where x_t is the noisy version of the data at step t and α_t controls the noise level at each step. After many steps, the data becomes indistinguishable from random noise.

Reverse Diffusion Process: The goal of the model is to reverse this noising process: given a noisy sample, predict a less noisy version of it, eventually recovering the original data. The reverse process is parameterized by a neural network that learns to denoise the noisy data step by step. The probability distribution in the reverse process is also Markovian and given by:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

Where $\mu_\theta(x_t, t)$ and $\Sigma_\theta(x_t, t)$ are the mean and variance predicted by the neural network. The forward and reverse diffusion processes are shown in Figure 4 (J.Ho, 2020).

Training Objective: DDPM is trained by loss function that looks like the mean squared error (MSE) between the added noise and the noise predicted by the model at each diffusion step. This loss at each timestep is:

$$\mathcal{L}_t = \mathbb{E}_{x_0, \epsilon, t} [\|\epsilon - \epsilon_\theta(x_t, t)\|^2]$$

where x_0 is the original data ϵ is the Gaussian noise added in the forward process, and $\epsilon_\theta(x_t, t)$ is the noise predicted by the neural network.

The training and sampling (i.e., image generation) algorithm of DDPM are given below in Figure 3. Stable Diffusion uses the

same method as in SR3 for Image conditioned denoising as shown Figure 5 in (C. Saharia, 2021).

The low-resolution image x is concatenated with noisy image y_t and combined images are used as input to U-Net model. Predicted noise is subtracted from y_t to generate y_{t-1} . Successive denoising continue to use low-resolution image x to guide denoising by concatenating with noisy images. In Stable Diffusion, the noisy images y_t are replaced by latents z_t . To improve the speed of denoising, we need to simplify the sampling method used for the reverse diffusion process.

Algorithm 1 Training

- 1: **repeat**
 - 2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
 - 3: $t \sim \text{Uniform}(\{1, \dots, T\})$
 - 4: $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
 - 5: Take gradient descent step on
$$\nabla_{\theta} \|\epsilon - \epsilon_{\theta}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, t)\|^2$$
 - 6: **until** converged
-

Algorithm 2 Sampling

- 1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
 - 2: **for** $t = T, \dots, 1$ **do**
 - 3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
 - 4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}} \epsilon_{\theta}(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$
 - 5: **end for**
 - 6: **return** \mathbf{x}_0
-

Figure 3. DDIM Algorithm

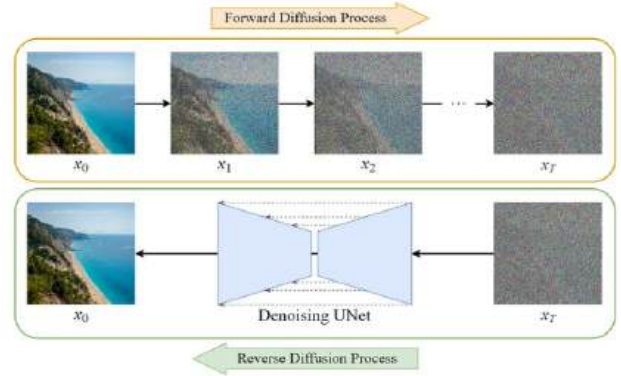


Figure 4. Denoising Diffusion

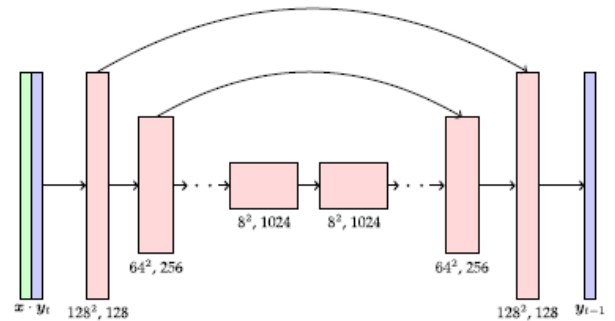
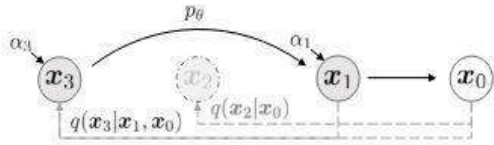


Figure 5. Image Conditioned U-Net

The DDPM used in Stable Diffusion employs the same number of steps in the reverse process as in the forward process. Denoising low-resolution images with the DDPM algorithm is very slow, as it takes hundreds of steps. For real-time applications, reducing time and computational resources is crucial. Denoising Diffusion Implicit Model (DDIM) was proposed to accelerate image generation (J. Song, 2022) and the denoising process in DDIM is shown in Figure 6, where the number of steps is reduced by sampling every n th step (with $n > 1$). In the example below, the value of n is 2.



Graphical model for accelerated generation, where $\tau = [1, 3]$.

Figure 6. Denoising diffusion implicit models

However, image quality deteriorates for large values of n , as the denoising trajectories remain stochastic. Recently proposed Consistency Models (CMs), directly map low-resolution images to high-resolution data and allow the use of deterministic denoising methods, as shown in Figure 7 (Y. Song, 2023).

The CM paper (Y. Song, 2023) proposed two methods for training Consistency Models, and in this project, Consistency Distillation (CD) method is used for training, where the CM is trained as a student model from a pre-trained teacher model. The training objective is to minimize the difference between the model outputs for pairs generated from the same trajectory.

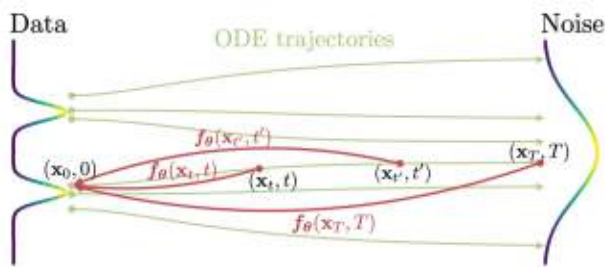


Figure 7. Deterministic Trajectory

The scope of this project is to train a DDM to use computationally efficient denoising based on deterministic trajectory and generate high-resolution images from low-resolution

satellite images. The method of training CM model and implementing the inference pipeline for satellite images will be discussed in the next section.

DOTA images are used as training and test datasets. Even though the size of the training and test datasets are quite small compared to typical size of datasets used for training generative models, the size is large enough to demonstrate improvement in denoising time using CMs.

Google Colab platform with V100 GPU was used for training the CMs and running inference pipeline to generate results. The source code in Huggingface diffusers repository is used as a baseline for this project. The CM training for text to image generation example in the repository is modified to support guided image generation conditioned on low resolution images. A new inference pipeline was implemented to perform super resolution using CM.

The PyTorch implementation of CM training and the inference pipeline are available on *dev_sr_branch* of my github repo - <https://github.com/jhanavi-h/diffusers>.

Methods

This method offers an advantage when a pre-trained model is already available from Hugging Face (HuggingFace, 2022), making the implementation more efficient. Algorithm of Teacher–Student Distillation process is described in Figure 8 (Y. Song, 2023).

Algorithm 2 Consistency Distillation (CD)

Input: dataset \mathcal{D} , initial model parameter θ , learning rate η , ODE solver $\Phi(\cdot, \cdot; \phi)$, $d(\cdot, \cdot)$, $\lambda(\cdot)$, and μ
 $\theta^- \leftarrow \theta$
repeat
 Sample $\mathbf{x} \sim \mathcal{D}$ and $n \sim \mathcal{U}[1, N - 1]$
 Sample $\mathbf{x}_{t_{n+1}} \sim \mathcal{N}(\mathbf{x}; t_{n+1}^2 \mathbf{I})$
 $\hat{\mathbf{x}}_{t_n}^\phi \leftarrow \mathbf{x}_{t_{n+1}} + (t_n - t_{n+1})\Phi(\mathbf{x}_{t_{n+1}}, t_{n+1}; \phi)$
 $\mathcal{L}(\theta, \theta^-; \phi) \leftarrow$
 $\lambda(t_n)d(\mathbf{f}_\theta(\mathbf{x}_{t_{n+1}}, t_{n+1}), \mathbf{f}_{\theta^-}(\hat{\mathbf{x}}_{t_n}^\phi, t_n))$
 $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}(\theta, \theta^-; \phi)$
 $\theta^- \leftarrow \text{stopgrad}(\mu \theta^- + (1 - \mu)\theta)$
until convergence

Figure 8. Consistency Distillation Training

The CM $f_\theta(\mathbf{x}, t)$ is characterized by the equation given below where it is a unity function at $t = \epsilon$.

$$\mathbf{f}_\theta(\mathbf{x}, t) = \begin{cases} \mathbf{x} & t = \epsilon \\ F_\theta(\mathbf{x}, t) & t \in (\epsilon, T] \end{cases}.$$

The pre-trained neural network $F_\theta(\mathbf{x}, t)$ provides the trajectory for training the CM $f_\theta(\mathbf{x}, t)$. The CM is trained over the time horizon (ϵ, T) with $N < T$ sub samples, at time instances $t_1 = \epsilon < t_2 \dots < t_{N-1} < t_N = T$. Given a datapoint x , the noisy sample $x_{t_{n+1}}$ is generated by adding noise with variance proportional to t_{n+1}^2 to x . Then a pair of adjacent data points $(x_{t_n}^\phi, x_{t_{n+1}})$ are generated where $x_{t_n}^\phi$ is generated by one step discretization step from the output $x_{t_{n+1}}$ from the teacher model. Then the student model $f_\theta(\mathbf{x}, t)$ is trained using back propagation to minimize the mean square distance - $d\left(f_\theta(x_{t_n}^\phi), f_\theta(x_{t_{n+1}})\right)$. The pretrained

DDIM from Huggingface is used as teacher model and CM training is done with mini-batch size of 12 with n values sampled from $U[1, N - 1]$. The CD training method in the CM paper (Y. Song, 2023) is modified to use guided image generation by conditioning the latents on low-resolution satellite images, as done in the SR3 implementation (C. Saharia, 2021). The CM paper (Y. Song, 2023) proposed sampling or image generation using a greedy algorithm, but this project proposed much simpler DDIM sampling that conditions the sampling process on the original image to remove Markovian dependency. This allows sampling with larger step sizes and enables super-resolution in just 3 to 4 steps.

Dataset

The DOTA v2.0 image dataset (Daa, 2018) is used for this project. DOTA consists of images collected from the Google Earth, GF-2, JL-1 satellites and aerial imagery provided by CycloMedia B.V. DOTA images are used to develop and evaluate object detectors needed for urban planning and ground cover estimation. The training data consists of 2,800 images, while the test data contains 200 images from the DOTA dataset. Even though the size of the training and test datasets are quite small compared to typical size of datasets used for training generative models, the size is large enough to demonstrate improvement in denoising time using CMs. The dataset has both RGB and grayscale images of varying sizes, all stored in PNG format. The images are labeled with various objects, but this information is ignored for the purposes of super-resolution. The ETL (extract, transform, load) pipeline selects only the RGB images, cropping them to a uniform size of 512 x 512 pixel. These

cropped images serve as the reference for training and performance evaluation. Additionally, the 512 x 512 images are down sampled by a factor of 4 to create low-resolution counterparts of size 128 x 128 pixels. The LDM is expected to enhance the resolution by a factor of 16, with performance assessed by comparing the model's output to the original 512 x 512 image.

Results

The consistency distillation example in the Huggingface Diffusers repository was modified to support guided image generation conditioned on low-resolution images. These changes were necessary to convert text-to-image generation into image-to-image generation. The training parameters for CD training are provided in Table 1.

Hyper-Parameter	Value
Learning Rate	0.000001
Loss Type	Huber
N	50
T	1000
Batch Size	12
Image Size	512x512
Time Step Scaling Factor	10

Table 1 Hyper-parameters

The hyperparameters of CD training were not changed, except for the batch size, which was adjusted to fit the training dataset into GPU memory. Stable Diffusion uses $T = 1000$ steps for forward diffusion process and the variance of the added noise is varied from $\beta_1 = 10^{-4}$ to from $\beta_T = 0.02$ and image data is normalized to $[-1, 1]$. The number of sub-samples, $N = 50$ is used for CD training. The batch size is set to 12 and limited by the GPU RAM size. Figure 9 shows plot of Huber loss of CD

training, and the learning curve does not show graceful reduction with number of samples. The training loss is small even at the beginning because the student model is initialized with same weights as pretrained teacher model.

It was difficult to judge the convergence of the model based on the training loss and the training loss was stopped after one epoch. There was no improvement seen by increasing the number of epochs. The stopping criteria for Teacher Student Distillation is not clear at this time and to be studied in the future.

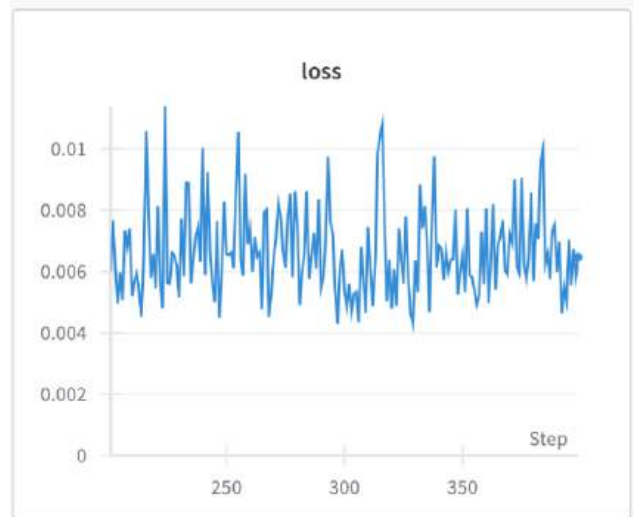


Figure 9 Training Loss

The Super Resolution process using Consistency Models (CM) is expected to be faster than DDPM, reducing the number of denoising steps from $T = 1000$ to 4, while maintaining comparable image quality. Traditional metrics like Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) are not ideal for evaluating generative models. Although high-resolution images produced by CM show only a modest improvement in PSNR

(up to around 4 dB), they exhibit substantial enhancements in Frechet Inception Distance (FID) scores. FID measures the similarity between the distribution of generated images and real images, with lower scores indicating more realistic outputs. Unlike MSE, which compares images pixel by pixel, FID evaluates the mean and standard deviation of activations in the deepest layers of the Inception v3 model. The FID score improved from 10 to 1.9 after improving the resolution using the proposed CM method. Lower FID scores imply improved feature extraction when enhanced satellite images are used in CNN-based object detection models for monitoring climate change. PSNR was improved by more than 3 dB for over 60% of the images.

Metric	Performance
PSNR	> 3 dB for 60% images
FID	1.9 (With Super Resolution) 10.0 (Before at Low Resolution)

Table 2 Improvement in Image Quality

The perceptual quality was significantly enhanced, as seen in the two examples given below.

Figure 10 (Right: Original, Left: Improved)



More examples of images with improved resolutions are available at my GitHub repo - <https://github.com/jhanavi-h/diffusers/images>.

Conclusion

The results indicate that modifying Stable Diffusion to use CM as a deterministic denoising model can significantly reduce computational resources and time. This model is a strong candidate for performing super-resolution on satellite images. Improved FID scores for super-resolution demonstrate that this implementation can be highly beneficial in the preprocessing module of satellites, enhancing the performance of downstream tasks such as object detection and classification, which are essential for tracking climate change.

The deterministic denoising method of CM demonstrates significant improvement in the perceptual quality of DOTA images, and it is worthwhile to apply this method and evaluate its performance on larger commercial datasets (e.g., Sentinel-2). However, further

improvements in the training algorithm are needed to generalize across larger datasets. The stopping criteria for improving the convergence of Teacher-Student Distillation training remains an active area of research, and adjustments will be necessary as better training methods become available. I am collaborating with the Huggingface AI community to release my modifications to the open-source community.

References

- J. Ho, A.Jain and P.Abbeel. *Denoising Diffusion Probabilistic Models*. Retrieved from arXiv: <https://arxiv.org/abs/2006.11239> (2020)
- O. Ronneberger, P. Fischer and T. Brox *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Retrieved from arxiv: <https://arxiv.org/abs/1505.04597> (2015).
- C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, M. Norouzi. *Image Super-Resolution via Iterative Refinement*. (arXiv) Retrieved from arXiv: <https://arxiv.org/abs/2104.07636> (2021)
- Daa. *Large Scale Benchmark and Challenges for Object Detection in Aerial Images*. Retrieved from DOTA: <https://captain-whu.github.io/DOTA/dataset.html> (2018).
- J. Song, C. Meng and S. Ermon. *Denoising Diffusion Implicit Models*. Retrieved from arxiv: <https://arxiv.org/abs/2010.02502> (2022).
- HuggingFace. *CompVis/stable-diffusion-v1-4*. Retrieved from HuggingFace: <https://huggingface.co/CompVis/stable-diffusion-v1-4> (2022)
- Huggingface. *huggingface / diffusers* . Retrieved from github: [https://github.com/huggingface/diffusers\(2024\)](https://github.com/huggingface/diffusers(2024))
- R. Rombach, A. Blattmann, D. Lorenz, P. Esser and B.Ommer. *High-Resolution Image Synthesis with Latent Diffusion Models*. Retrieved from arxiv: <https://arxiv.org/abs/2112.10752> (2022)
- S. Zhao, M. Liu, M. Tao, W. Zhou, X. Lu, Y. Xiong, F. Li, Q. Wang. The role of satellite remote sensing in mitigating and adapting to global climate change. *Science Direct* (2023).
- Y.Song, P.Dhariwal, M. Chen, I.Sutskever. *Consistency Models*. Retrieved from arxiv: <https://arxiv.org/abs/2303.01469> (2023).