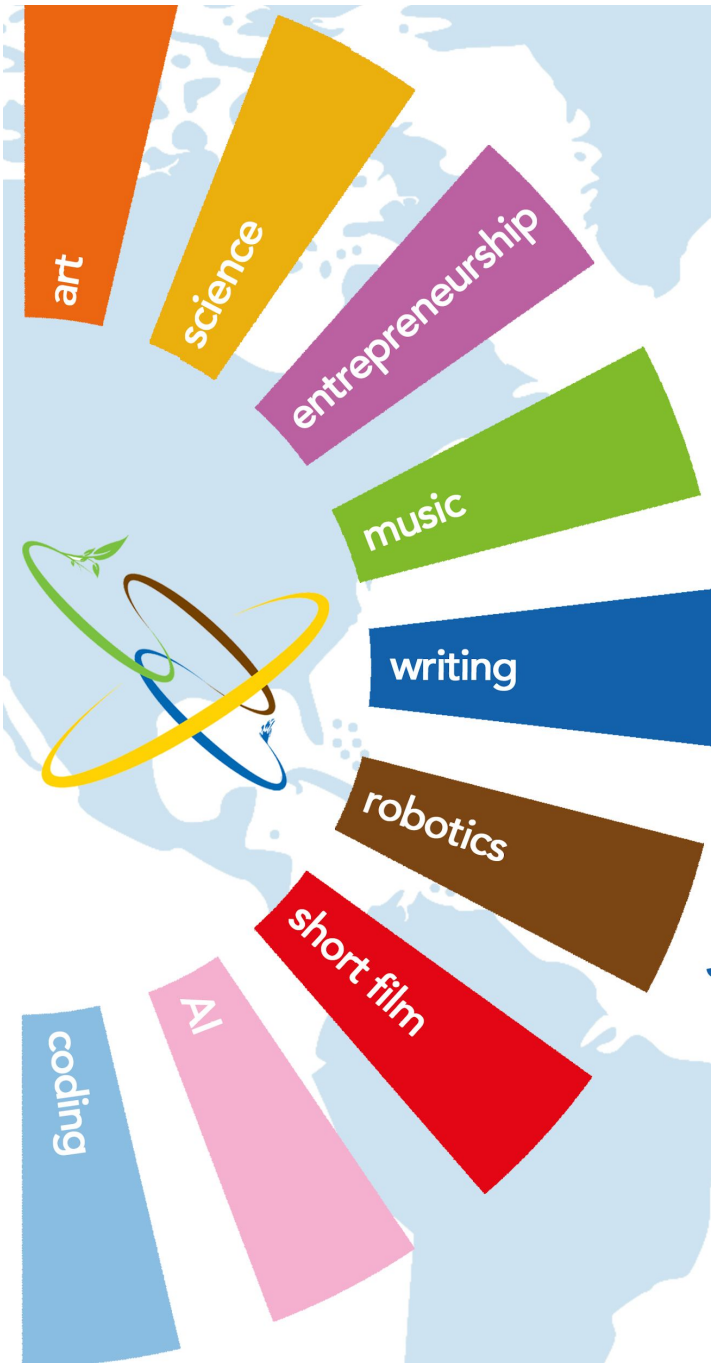
IJHSR International Journal of High School Research

June 2026 | Volume 8 | Issue 12
ijhsr.terrajournals.org
ISSN (Print) 2642-1046
ISSN (Online) 2642-1054

GENIUS OLYMPIAD

Let's build a better future together

International Project Fair focused on Sustainability and Environment for Grades 8-12



- Since 2011
- Hosted around 1400 participants in 2025 from 35+ states 70+ countries
- **Disciplines:** STEM, Coding, Robotics, AI, Speech, Entrepreneurship, Arts, Short Film, Music
- Applications start on December 1
- Application Deadline is March 1
- Finalists are announced by March 25
- Event is usually scheduled 2nd week of June
- Monday – Friday, includes a trip to Niagara Falls
- Hosted by large universities at Upstate New York
- Application Fee is $60/ project
- Participation Fee is $600/ person, w/ room and board
- Open buffet breakfast, lunch, and dinner
- Trip to Niagara Falls and boat tour is included
- **Instagram and Facebook @Geniusolympiad**
- For more information: **GENIUSOlympiad.org**
- Email: **info@geniusolympiad.org**

GENIUS Olympiad is organized by Terra Science and Education, a N.Y. based 501.c.3 non-profit organization dedicated for project-based learning

Marine Biology Research at Bahamas

Unique and exclusive partnership with the Gerace Research Center (GRC) in San Salvador, Bahamas to offer marine biology research opportunities for high school teachers and students.

- Terra has exclusive rights to offer the program to high school teachers and students around world.
- All trips entail extensive snorkeling in Bahamian reefs as well as other scientific and cultural activities.
- Terra will schedule the program with GRC and book the flights from US to the GRC site.
- Fees include travel within the US to Island, lodging, meals, and hotels for transfers, and courses.
- For more information, please visit terraed.org/bahamas.html

Terra is a N.Y. based 501.c.3 non-profit organization
dedicated for improving K-16 education

Table of Contents

June 2026| Volume 8| Issue 12

Editorial Board

International Journal of High School Research

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Perceptions of AI Among K–12 Educators and Administrators: A Mixed-Methods Study

Snowden Lange

Chatham High School, 255 Lafayette Avenue, Chatham, NJ 07928, USA; snowdenlange@icloud.com
Mentor: Amanda Rothlisberger

ABSTRACT: This project explores how teachers and school administrators in a New Jersey public school district perceive and use artificial intelligence (AI) in education. Using the Technology Acceptance Model (TAM) and the Zone of Proximal Development (ZPD) theory, the study examines educators' perceptions of usefulness, ease of use, preparedness, beliefs about student learning, and intention to adopt AI in K–12 settings. Seventy-one educators completed a survey, and responses were analyzed using descriptive statistics, correlations, k-means clustering, and one-way ANOVA with post hoc comparisons. Cluster-based analyses conducted with teachers identified three groups—AI Advocates, Curious but Cautious, and Skeptics—which differed significantly in their perceptions of AI's usefulness, ease of use, preparedness, and intention to adopt it. An interview and open-ended responses provided qualitative insights into educators' training needs, ethical concerns, and beliefs about student learning. Correlations between overall AI sentiment and TAM and ZPD measures highlighted the importance of mindset, confidence, and teacher beliefs in shaping AI adoption.

KEYWORDS: Behavioral and Social Sciences, Clinical and Developmental Psychology, Artificial Intelligence in Education, Technology Acceptance Model, Zone of Proximal Development.

Introduction

Artificial intelligence (AI) has become increasingly visible across many domains, including education, prompting discussion about its role in professional practice. With tools like ChatGPT, Gemini, and Microsoft Copilot becoming widely available, discussions about the benefits, challenges, and ethical concerns of AI in education have become increasingly prominent in educational environments. However, rapid technological advancements have outpaced research into educators' and administrators' real-world experiences and attitudes about AI.

Artificial intelligence (AI) is a term that encompasses a wide range of computational techniques and applications. In this study, AI is operationally defined as user-facing software tools that educators are likely to come across in contemporary K–12 settings. These include general-purpose AI assistants (e.g., ChatGPT, Gemini, Microsoft Copilot), AI-supported writing and grammar tools (e.g., Grammarly, Quillbot, Hemingway Editor), AI-based research assistance platforms (e.g., Elicit, Scite, Consensus), and AI-powered instructional or lesson-planning tools (e.g., Khanmigo, Curipod, MagicSchool AI).

Prior research suggests that AI-supported tools may support student learning in several ways, including assistance with homework and studying, personalized learning experiences, and skill development.[1] In 2021, Zhai *et al*. analyzed studies from 2010 to 2020 to identify trends and challenges of AI in education.[2] Much of this research was prior to the widespread availability of generative AI tools and focused on earlier forms of educational AI. From a theoretical perspective, these studies highlighted how AI systems were often designed to adapt content to learners' current abilities, which aligns with Vygotsky's Zone of Proximal Development (ZPD).

The Zone of Proximal Development (ZPD) is the range between what a student can do independently and what they can do with assistance from someone more experienced, such as a teacher, coach, or peer. It represents the "sweet spot" for learning, where students are challenged just beyond their current abilities but can still succeed with the right support. In this sense, AI-supported tools can be conceptualized as aligning with the ZPD through features such as:

- Customized learning activities that address the student's current gaps in knowledge and skills.
- Adaptive content that adjusts in real time to deliver the appropriate complexity based on the student's ongoing performance.
- Immediate feedback and corrections that help students understand their mistakes and learn from them in real time.
- Scaffolded support that provides temporary help for tasks just beyond the student's current ability, and is gradually removed as competence increases.

The use of AI in educational settings has raised substantial concerns. Educators and researchers have expressed apprehension about student overreliance on AI-generated content, erosion of critical thinking skills, accuracy of AI outputs, and challenges related to academic integrity and data privacy.[5] These concerns complicate educators' decisions about when and how to use AI in classrooms and emphasize the importance of examining perceived risks and ethical implications in addition to potential instructional benefits.

Some studies report that educators perceive AI tools as time-saving, as they support instructional planning and con-

DOI: 10.36838/IJHSR812.1

tent development, but educators also have concerns about reliability, accuracy, and ethical use.[1,3] In particular, Bergdahl and Sjöberg found that teachers have greater confidence in using AI chatbots for teaching when they feel clear about the ethics and safety of the tools.[3]

One factor associated with educators' use of AI in classrooms is their confidence in their ability to do so, which is often referred to as AI self-efficacy. Research shows that teachers with higher AI self-efficacy are more likely to have positive attitudes toward AI and integrate it into their teaching practices.[4] Several factors shape this confidence, including previous experience with technology, professional training, and whether teachers view AI as relevant to their work.[5]

The Technology Acceptance Model (TAM), developed by Fred Davis in 1989,[6] is one of the most widely used models to explain and predict the adoption of new technologies. While TAM originally focused on two core factors, this study incorporates two additional measures commonly used in extended versions of the model:

- Perceived Usefulness: how much a person believes a technology will enhance their job performance
- Perceived Ease of Use: how much effort they think it will take to use the technology effectively
- Preparedness: how ready they feel to integrate AI into their roles
- Intention to Use: how likely they are to adopt AI in the near future

In the original TAM framework, perceived usefulness and ease of use shape a person's attitude toward the technology, which then influences their intention to use it and, ultimately, their actual use. Although educational settings are inherently collaborative, the focus of this study is on educators' individual use of AI tools as professional supports (e.g., lesson planning, content creation, feedback, and administrative assistance), rather than on shared or mandated technologies. In this context, individual beliefs, confidence, and perceived readiness are central to adoption decisions, making the Technology Acceptance Model particularly appropriate. Intention to use serves as a meaningful outcome measure for understanding educators' likelihood of adopting AI in practice.

This study captures a comprehensive picture of educators' readiness and likelihood of adopting AI by incorporating preparedness and intention, extending Davis's model in ways supported by subsequent research, including Mun and Hwang's work on teacher adoption of digital tools.[7]

Research in language instruction settings suggests that some educators perceive AI tools, such as ChatGPT, as improving their efficiency, reducing some aspects of their workload, and increasing their confidence in technology use.[8] Similarly, studies report that educators may view AI as helpful for streamlining lesson planning and content creation, enabling them to produce customized materials while saving time.[9] Some research also indicates that school principals perceive AI as useful for automating administrative tasks and supporting data analysis.[10]

While AI's potential in education is widely discussed, prior research has disproportionately focused on technical applications rather than educators' perspectives, leaving gaps in understanding teachers' readiness, attitudes, and self-efficacy related to AI use.[5] There is also limited research examining training tailored to educators, as well as administrators' attitudes toward AI. Understanding these factors can help clarify how educators interpret, respond to, and make decisions about AI use in educational settings. Although AI is often framed as offering efficiency and instructional support, its educational value, particularly in the context of newer generative tools, remains contested, with educators weighing potential benefits against concerns related to ethics, learning, and student development.

This study uses a mixed-methods approach, based on the Technology Acceptance Model (TAM) and Vygotsky's Zone of Proximal Development (ZPD) theory, to investigate how K–12 educators and administrators perceive, use, and evaluate AI in their professional roles.

Methods

This study employed a mixed-methods approach to investigate how educators and administrators in K–12 schools perceive AI. A mixed-methods design was appropriate because quantitative survey data were used to identify broad patterns in attitudes and adoption, while qualitative data from open-ended responses and an interview were used to help contextualize and explain those quantitative findings, consistent with an explanatory mixed-methods approach. This explanatory approach enabled the interpretation of statistical results through educators' experiences and concerns. We collected data from three sources: survey responses, written comments, and an interview with a high school teacher. To analyze the results, we used a combination of statistical analysis, a review of common themes in written responses, and a machine learning method known as k-means clustering to group participants based on their responses to key questions related to the Technology Acceptance Model (TAM). This helped reveal patterns in attitudes toward AI that did not follow job roles or grade levels. By combining statistics with personal insights, the study captured both broad trends and deeper perspectives on educators' thinking about AI.

A total of 71 educators from a public school district in New Jersey participated in the survey. Respondents included 57 teachers and 14 administrators or staff members in student services. Because the sample included relatively few administrators (14 out of 71 respondents), most of the analysis focuses on patterns among teachers. Teachers represented a range of grade levels as shown in Table 1. Table 2 shows the number of teacher participants by subject area; participants could select more than one subject. Administrators held various leadership and support roles as shown in Table 3. Participants were recruited through district-wide email outreach and completed the survey voluntarily and anonymously during a two-week window in April 2025. The survey was distributed broadly via email, so a precise response rate could not be calculated, but with 71 completed responses in a district with approximately 600 total staff, the sample represents a sizable portion of the district's workforce.

Table 1: Number of teacher participants by grade level.

Elementary (K-5)	26
Middle School (6-8)	18
High School (9-12)	13
Total	**57**

Table 2: Number of teacher participants by subject area. Participants could select more than one subject.

English Language Arts	24
Social Studies	22
Math	20
Science	14
Special Education	8
Fine Arts	3
Physical Education	2
Electives	2
World Languages	2
Business	1

Table 3: Roles of administrator participants.

School or district leadership	6
Student services	6
Operations	1
Technology and IT	1
Total	**14**

The survey consisted of 52 items across three main areas: (1) Technology Acceptance Model (TAM) constructs such as perceived usefulness, perceived ease of use, behavioral intention to use, and preparedness; (2) Zone of Proximal Development (ZPD) aligned items assessing the ability of AI to support adaptive and independent learning; and (3) ethical concerns and perceived challenges related to AI use in education. Survey items were designed from established Technology Acceptance Model (TAM) instruments and supplemented with items aligned with current AI-in-education concerns. Items were informally reviewed for clarity and relevance to K–12 contexts. To support consistent interpretation, examples of widely used AI tools (e.g., ChatGPT, Grammarly, MagicSchool AI) were embedded in relevant survey items. This helped anchor responses to common applications, while acknowledging that individual definitions of AI might vary. A copy of the survey instrument is available online for review (https://tinyurl.com/5x9harxy). Most Likert-scale items used a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree).

To assess the internal consistency of the TAM instruments, we used Cronbach's alpha for each multi-item dimension. Internal consistency was good for perceived usefulness ($\alpha = .825$) and perceived ease of use ($\alpha = .826$), and acceptable for behavioral intention ($\alpha = .711$). Preparedness was assessed using a single item, so Cronbach's alpha was not applicable.

We used descriptive statistics and Pearson correlations to examine relationships between overall AI sentiment and the TAM and ZPD dimensions. One-way analyses of variance (ANOVA) with Tukey *post hoc* tests were conducted using SPSS to determine if mean differences across clusters were statistically significant.

To identify natural groupings in attitudes toward AI, exploratory k-means clustering was applied to the four TAM dimensions (perceived usefulness, ease of use, preparedness, and intention to use). This analysis was performed using the scikit-learn Python library (facilitated by ChatGPT). The researcher maintained full analytical oversight by specifying the input variables, determining the optimal number of clusters (k=3), and verifying all cluster memberships against the raw data to ensure accuracy.

We grouped open-ended survey responses by overall sentiment and looked for common themes. These included concerns about ethics, the benefits of AI, its application in education, and challenges to its broader adoption. We also interviewed a high school German teacher who volunteered to participate and had been using AI in her classroom. Her answers helped us see how the survey results connected to the ideas in our study, like the TAM and the ZPD. What she shared also helped us better understand the differences between the three groups we found through machine learning.

Results

To examine patterns in educators' attitudes toward AI, k-means clustering was conducted using 13 Likert-scale items related to perceived usefulness, perceived ease of use, preparedness, and behavioral intention to use. This analysis identified three distinct clusters (Cluster 1, n=21; Cluster 2, n=22; and Cluster 3, n=14). Mean scores for each Technology Acceptance Model (TAM) dimension by cluster are shown in Table 4.

Table 4: Mean Survey Scores by cluster across TAM dimensions.

Cluster	Perceived Usefulness	Behavioral Intention to Use	Preparedness	Perceived Ease of Use
Cluster 1	4.76	4.29	4.00	4.19
Cluster 2	4.29	3.86	2.79	2.07
Cluster 3	3.21	2.64	1.79	2.50

One-way analyses of variance (ANOVAs) were conducted to examine differences among the three clusters across the Technology Acceptance Model (TAM) dimensions using composite scores. Because the three clusters were generated from the TAM dimensions (Perceived Usefulness, Perceived Ease of Use, Behavioral Intention, and Preparedness), high F-statistics and significant p-values in the ANOVA results are expected. These analyses are presented not as independent hypothesis tests, but as a method of internal validation to profile the distinct characteristics of each cluster and to confirm the statistical separation achieved by the k-means algorithm. Significant differences were found across clusters for perceived usefulness, $F(2, 54) = 69.21$, $p < .001$; perceived ease of use, $F(2, 54) = 65.13$, $p < .001$; behavioral intention to use AI, $F(2, 54) = 40.99$, $p < .001$; and preparedness to integrate AI, $F(2, 54) = 31.88$, $p < .001$. These results indicate that cluster membership was associated with meaningful differences across all core TAM constructs and preparedness.

Post-hoc Tukey HSD tests were conducted to identify where specific cluster differences occurred. For perceived usefulness, all pairwise comparisons were statistically significant: Cluster

DOI: 10.36838/IJHSR812.1

1 differed from Cluster 2 (p = .0003) and Cluster 3 (p < .001), and Cluster 2 differed from Cluster 3 (p < .001). For behavioral intention to use AI, all cluster comparisons were also significant, including Cluster 1 versus Cluster 2 (p = .0326), Cluster 1 versus Cluster 3 (p < .001), and Cluster 2 versus Cluster 3 (p < .001). Similarly, preparedness differed significantly across all cluster pairs (Cluster 1 vs. 2, p < .001; Cluster 1 vs. 3, p < .001; Cluster 2 vs. 3, p = .0025). For perceived ease of use, Cluster 1 differed significantly from both Cluster 2 (p < .001) and Cluster 3 (p < .001), but the difference between Clusters 2 and 3 was not statistically significant (p = .202). While a slight trend suggests Skeptics found tools marginally easier than the Curious but Cautious group, the two clusters demonstrated comparable levels of perceived ease of use.

Based on these distinct statistical profiles and the qualitative themes described below, the clusters were labeled to reflect their overall perspectives on AI: Cluster 1 (AI Advocates), Cluster 2 (Curious but Cautious), and Cluster 3 (Skeptics). These differences are visualized in Figure 1, which illustrates the mean TAM scores across dimensions for each group.

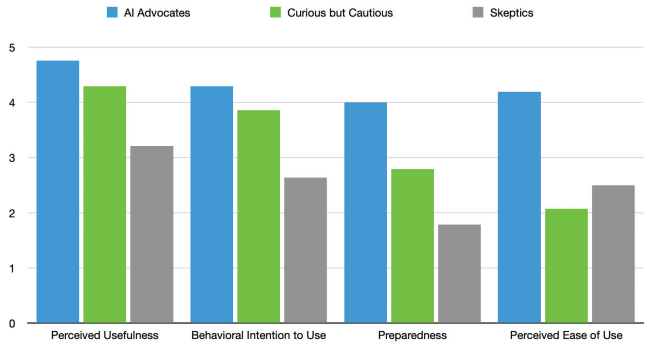


Figure 1: Average TAM scores by cluster. AI Advocates exhibited the highest mean scores across all four TAM dimensions: perceived usefulness, behavioral intention to use, preparedness, and perceived ease of use. Skeptics scored lowest on all dimensions except ease of use, where their scores were slightly above those in the Curious but Cautious cluster.

To provide context for these clusters, we examined general usage patterns across the district. A moderate correlation (r = 0.41, p < .001) was found between the frequency of teachers' use of AI assistants and their overall positive feelings about AI. This suggests that teachers who used AI more frequently were also more likely to hold favorable attitudes toward it. Educators and administrators reported moderate use of AI tools overall, though most had at least some experience with them. Writing tools like Grammarly appeared to be the most commonly adopted AI applications among teachers, with a large majority reporting at least occasional use. Elementary teachers were especially likely to use lesson planning platforms such as MagicSchool. Use of general AI assistants like ChatGPT varied. While some teachers used them often, most reported using them only occasionally or rarely. Data tools, such as AI dashboards or gradebook analytics, were the least used, with over half of teachers stating they never used them. In contrast, administrators reported using both data tools and general AI assistants more regularly, likely due to their operational responsibilities. These findings reflect self-reported usage trends rather than statistically tested differences.

AI Advocates (Cluster 1):

Qualitative data showed that AI Advocates viewed AI as a helpful, intuitive tool that could support differentiated instruction, content creation, and student learning. Many already use AI regularly and are confident in their ability to integrate it effectively. Their comments reflected comfort, enthusiasm, and practical alignment with teaching needs.

"I've used AI to reword things for different reading levels. It's been very helpful for students who need accommodations."

"AI can scaffold learning, but only if you know how to prompt it at the right level."

"Creating report card comments."

"Generating multiple choice questions. I'm terrible at writing those."

"Emails and writing report card comments!"

These educators believe in the ethical and purposeful use of AI and have already begun using it in their classrooms.

Curious but Cautious (Cluster 2):

Survey comments from this cluster showed moderate optimism about AI's potential but lower confidence and readiness. Curious but Cautious educators saw potential in AI but expressed hesitation about its limits, ethical implications, and impact on student learning. They were interested in learning more, but often cited a lack of training or clarity.

"It's helpful, but I don't feel confident enough to use it consistently yet."

"Teachers shouldn't lose that personal connection with their students' abilities and progress."

"I have limited knowledge of what is available and how to use it."

"Training would be helpful for understanding the best ways to use it effectively and ethically."

This group would benefit from more structured training, policy clarity, and real-world examples of how AI can align with their teaching goals.

Skeptics (Cluster 3):

This group scored lowest on perceived usefulness, behavioral intention, and preparedness, but reported slightly higher ease of use than the Curious but Cautious group. Skeptics were unsure of AI's educational value and were most concerned about its potential risks. Their comments reflected discomfort with AI as an instructional tool, particularly regarding student overreliance, loss of critical thinking, and academic integrity:

"Students… are not able to think critically anymore… they can barely write a complete sentence… AI just makes this more noticeable."

"It's like whack-a-mole… how am I going to figure out where the mole's coming from next?"

"AI could be used for the wrong reasons / the potential for students to misuse it is high."

"Plagiarism, not learning or creatively thinking, just copying."

Although some Skeptics acknowledged that AI could be useful for administrative tasks, they were unlikely to adopt it without strong guardrails.

Training Needs:

Training emerged as a common theme across all groups, but was especially important for the Curious but Cautious group. While this group rated AI as useful, they reported lower confidence in their ability to implement it effectively. Their Preparedness and Ease of Use scores were both substantially lower than their Usefulness and Intention scores, revealing a gap between interest and readiness.

The German teacher interviewed for this study echoed this sentiment:

"You can't just tell students to use AI and expect it to go well. You need to know how to prompt, guide, and set limits. That takes time and training."

Additional quotes reinforced this theme:

AI Advocate: "The potential is there, but we need time and training to do it right."

Curious but Cautious: "I want to use it more, but I don't know where to start or what tools are even allowed."

Curious but Cautious: "I feel like I'm being asked to use something I don't fully understand yet."

Technology Acceptance Model (TAM) Findings:

In addition to clustering, we also examined how teachers' overall perception of AI correlated with four TAM-related items. More positive sentiment about AI was associated with higher levels of:

- Perceived Usefulness ($r = 0.54$, $p < .001$)
- Behavioral Intention ($r = 0.57$, $p < .001$)
- Preparedness ($r = 0.37$, $p < .001$)
- Ease of Use ($r = 0.47$, $p < .001$)

All reported correlations were statistically significant ($p < .001$). These moderate correlations are consistent with the TAM framework, suggesting that as teachers view AI more favorably overall, they are more likely to find it useful, feel confident in integrating it, and intend to adopt it. The strongest correlation was between overall perception and Behavioral Intention.

Among administrators, the same sentiment question correlated even more strongly with TAM variables, including Perceived Usefulness ($r = 0.88$, $p < .001$), Behavioral Intention ($r = 0.82$, $p < .001$), and belief in AI's ability to personalize learning ($r = 0.78$, $p < .001$). These results suggest that administrators may be more attuned to AI's strategic potential in education, although the small sample size ($n = 14$) limits the strength of this conclusion.

Zone of Proximal Development (ZPD) Findings:

Teachers who felt more positively about AI were moderately likely to agree that it can support adaptive learning at the right level ($r = 0.54$, $p < .001$), but less so with its ability to support independent learning ($r = 0.31$, $p < .001$). Among all items, "Support adaptive learning" had one of the highest mean scores, showing broad agreement that AI can help tailor instruction. The related item about providing feedback for independent learning received lower but still favorable ratings. These findings suggest that while educators recognize the value of AI in scaffolding learning, they remain cautious about its potential to replace teacher-led guidance or foster autonomy.

Concerns and Challenges:

Educators expressed a range of concerns about using AI in schools. Figure 2 shows the frequency of concerns selected by respondents in a "select all that apply" survey item assessing perceived challenges related to AI use in education. Percentages reflect the proportion of respondents who selected each concern. The most commonly cited issues were the potential inaccuracy of AI-generated content, followed closely by the need for more training and the risk that AI could hinder student critical thinking. Ethical concerns, including bias, fairness, and privacy, were also frequently mentioned. Some respondents were skeptical of AI's guidance or reliability, and a smaller number felt that school or district policies limited their ability to use AI effectively.

These results highlight the importance of policies, professional development, and implementation strategies. While many educators are open to using AI in their work, their concerns suggest that effective integration depends not only on access to tools but also on support, trust, and alignment with educational goals.

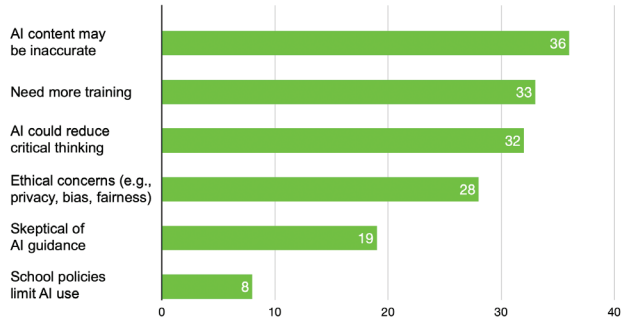


Figure 2: Frequency of top concerns about AI in education reported by all respondents. The most common concerns included the potential inaccuracy of AI-generated content, the need for more training, fears that AI could hinder students' critical thinking, and ethical concerns. A smaller number of respondents indicated distrust in AI's ability to provide reliable guidance to students, and a few noted that school or district policies limited their ability to use AI effectively.

Ethical Concerns:

Most teachers and administrators agreed that AI is acceptable for generating study materials, tutoring, and offering hints or suggestions as long as it helps students think for themselves rather than solving the problem for them. Unsurprisingly, nearly all strongly opposed using AI during tests or to complete assignments, reflecting shared concerns about academic integrity.

Teachers expressed greater concern about student over-reliance on AI (median 3.6) than administrators (median 2.36), as indicated by their agreement with the statement, "Students are becoming overly reliant on AI, which may negatively affect their ability to problem-solve without technological assistance."

Many teachers were especially concerned about how AI could affect student independence and long-term learning.

DOI: 10.36838/IJHSR812.1

One wrote, "I worry about students becoming over-reliant on AI. I also worry that they will not be able to discern whether or not what AI has provided is reliable or applicable to the situation." Another teacher noted, "Many students currently use it simply to complete assignments and are not able to or do not vet the accuracy of information that AI generates."

Some responses went beyond academic concerns and spoke to broader issues of motivation, curiosity, and skill development. One teacher explained, "It is essential to address several logistical, ethical, and critical thinking concerns that arise when students engage with this technology… With growing concerns about student independence and motivation, we must consider how the introduction of AI could either challenge or enhance these dynamics."

The most consistent message was that AI should support learning, but not replace the thinking and effort students need to grow.

Discussion

This study reveals that K–12 educators and administrators hold distinctly different perceptions of AI that cannot be explained by frequency of use or awareness alone. Through a mixed-methods approach combining clustering, survey data, and qualitative insights, three profiles emerged—AI Advocates, Curious but Cautious, and Skeptics—each characterized by differing views of AI's usefulness and ease of use, and their preparedness and intention to use it. These differences were also reflected in educators' beliefs about how AI should support student learning and in recurring concerns about training needs, ethical risks, student over-reliance, critical thinking, and academic integrity.

While survey-based research at the higher-education level has explored educators' attitudes toward AI, far fewer studies have examined the K–12 context, where instructional demands, developmental needs, and accountability pressures differ significantly. This study helps fill that gap by offering one of the first mixed-methods investigations of AI adoption in K-12 schools. These patterns echo findings from a 2025 multi-state study in which educators expressed both optimism and concern about AI's role in schools. While nearly 60% of educators in that study strongly agreed on the importance of having school or district policies guiding AI use, a high percentage disagreed that their district provided sufficient support or infrastructure for effective implementation.[11] That study, like the present one, revealed that many educators believed AI could enhance student learning, yet simultaneously worried about cheating, overreliance, and ethical boundaries. This tension was especially pronounced among the Curious but Cautious and Skeptics clusters in the current study, reinforcing the need for district-level policies and targeted professional development.

At the same time, a national survey of US K-12 educators found that many teachers were already using AI for instructional tasks, including personalizing learning (56%), providing real-time feedback (52%), lesson planning (44%), and proofreading (47%).[12] These use cases align closely with comments from AI Advocates in this study, who described using AI to reword content for different learners, co-create lesson materials, and scaffold student learning. The national survey and this local study suggest that when teachers see an instructional benefit, they are more likely to experiment with AI, particularly in ways that support differentiated instruction and reduced workload. Taken together, these findings can be interpreted through the lens of the Zone of Proximal Development (ZPD). Educators across groups in this study were more comfortable using AI to scaffold learning, such as adapting content or providing feedback at an appropriate level, than to support fully independent learning. This pattern reflects a view of AI as a supplementary tool that operates within students' proximal learning space rather than as a replacement for teacher guidance or student effort.

Limitations of the Research:

This study has several limitations that should be considered when interpreting the findings. First, the study was conducted with a small sample size within a single public school district in New Jersey, which may limit generalizability to districts with different demographics, governance structures, resources, or policies. The district examined is relatively well-resourced and higher-performing than many public school districts, which may influence educators' access to technology, professional development opportunities, and perceptions of AI. These variables were not directly measured in the present study and therefore cannot be shown through the reported data. Future research should examine how district-level context interacts with educators' attitudes toward AI to better assess the generalizability of findings across diverse educational settings. Additionally, the selection of a three-cluster solution represents one analytic choice; alternative cluster solutions may yield different groupings and should be explored in future research. Because k-means clustering requires the number of clusters to be specified a priori, its use in this study reflects an exploratory analytic judgment rather than a definitive determination of the true number of latent groups.

Participation was voluntary, introducing potential self-selection bias, as respondents may have had stronger opinions about AI or greater availability to complete the survey than non-respondents.

Although respondents were provided with examples of specific AI tools, participants may still have had differing interpretations of what the researcher meant by "AI," particularly regarding the differentiation between traditional educational software and newer generative AI systems. This construct ambiguity may have influenced self-reported usage levels, attitudes, and cluster membership. However, this limitation does not invalidate the findings. It suggests that reported perceptions reflect educators' understanding of AI as they experience it in their daily work. Future research could reduce ambiguity by focusing on specific categories of AI tools or by asking participants to define AI before completing the survey.

Several items in the survey were newly developed or adapted to address emerging AI-in-education contexts and have not undergone formal psychometric validation. As a result, some constructs, particularly those outside the core TAM measures,

should be interpreted as exploratory. Future research should further refine and validate these items using larger samples and psychometric techniques.

The qualitative component included a single interview and open-ended survey questions. The interview was conducted with an educator supportive of AI use, which may have biased some of the qualitative data toward more favorable insights. As a result, the qualitative findings may not fully represent the range of educator viewpoints. Future studies should include interviews with educators representing multiple sentiment groups and roles to capture a broader range of perspectives. Additionally, responses to the open-ended survey questions were optional, which may have introduced selection bias, as participants with stronger opinions or greater interest in AI could be more likely to respond. This could have influenced the themes that emerged from the qualitative data.

Given the exploratory nature of this study, findings that did not reach conventional levels of statistical significance should be interpreted as trends that point toward directions for future research.

While the Technology Acceptance Model (TAM) is useful for examining adoption attitudes, it does not explicitly model ethical concerns or perceived risks, which emerged prominently in qualitative responses. Although educational settings are inherently collaborative, the present study focused specifically on educators' individual perceptions, intentions, and preparedness related to AI use, rather than collective or policy-driven adoption. In this context, the TAM is well-suited to the research questions examined. Future research could extend this work by pairing TAM with complementary frameworks that more directly capture ethical considerations, perceived risk, and the collaborative dynamics of educational environments.

Because significant differences and correlations involving perceived usefulness were still observed despite lower internal consistency, these findings may represent conservative estimates of the true relationships, suggesting that effects could be stronger with a more refined instrument.

Conclusion

The three educator groups identified through k-means clustering—AI Advocates, Curious but Cautious, and Skeptics—differed significantly across all Technology Acceptance Model (TAM) measures. This suggests that attitudes toward AI reflect patterns in mindset, confidence, and instructional philosophy. Teachers who viewed AI more positively were more likely to feel confident using it, believe it was helpful, and report a stronger intention to adopt it in the future. Most educators saw value in AI's ability to adjust content to individual students' needs, but were less confident in its ability to support independent learning or replace teacher feedback. They were especially confident in AI's ability to scaffold learning at the right level, though less so in its potential to foster independent learning, which is consistent with Vygotsky's Zone of Proximal Development.

To support the responsible and effective use of AI in schools, leaders should prioritize teacher training, establish clear ethical guidelines, and provide real-world examples. Teachers need time and support to explore how AI can enhance planning, instruction, and differentiation while also ensuring that students stay engaged in thinking and learning for themselves. Schools that invest in this kind of support are more likely to see thoughtful, effective, and ethical adoption.

Future research should examine how AI use affects student outcomes over time, particularly how students at different developmental stages engage with these tools. This study suggests that teacher perception and preparation matter as much as the technology itself. To ensure AI enhances learning rather than undermines it, schools must invest in the people who use it by providing clear policies, strong training, and time to experiment.

Acknowledgments

The author would like to thank Dr. Michael LaSusa, Superintendent of the School District of the Chathams, for granting permission to conduct the study and for helping distribute the survey to teachers and administrators. Special thanks also go to Mrs. Amanda Rothlisberger, AP Statistics teacher at Chatham High School, for her thoughtful guidance and multiple reviews of the survey design and data analysis.

References

1. Labadze, L.; Grigolia, M.; Machaidze, L. Role of AI Chatbots in Education: Systematic Literature Review. *Int. J. Educ. Technol. High. Educ.* **2023**, 20, 56. DOI: https://doi.org/10.1186/s41239-023-00426-1
2. Zhai, X.; Chu, X.; Chai, C. S.; Jong, M. S. Y.; Spector, M.; Liu, J.-B.; Yuan, J.; Li, Y. A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, **2021**, Article ID 8812542. DOI: https://doi.org/10.1155/2021/8812542
3. Bergdahl, N.; Sjöberg, J. Attitudes, Perceptions and AI Self-Efficacy in K–12 Education. *Comput. Educ.: Artif. Intell.* **2025**, 6, 100358. DOI: https://doi.org/10.1016/j.caeai.2024.100358
4. Viberg, O.; Cukurova, M.; Feldman-Maggor, Y.; Alexandron, G.; Shirai, S.; Kanemune, S.; Wasson, B.; Spikol, D.; Milrad, M.; Coelho, R.; Kizilcec, R. F. Teachers' Trust and Perceptions of AI in Education: The Role of Culture and AI Self-Efficacy in Six Countries. arXiv **2023**, arXiv:2312.01627. DOI: https://doi.org/10.48550/arXiv.2312.01627
5. Zawacki-Richter, O.; Marín, V. I.; Bond, M.; Gouverneur, F. Systematic Review of Research on Artificial Intelligence Applications in Higher Education—Where Are the Educators? *Int. J. Educ. Technol. High. Educ.* **2019**, 16, 1. DOI: https://doi.org/10.1186/s41239-019-0171-0
6. Davis, F. D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **1989**, 13(3), 319–340. DOI: https://doi.org/10.2307/249008
7. Mun, Y. Y.; Hwang, Y. Predicting the Use of Web-Based Information Systems: Self-Efficacy, Enjoyment, Learning Goal Orientation, and the Technology Acceptance Model. *Int. J. Hum.-Comput. Stud.* **2003**, 59, 431–449. DOI: https://doi.org/10.1016/S1071-5819(03)00114-9
8. Ghafouri, M.; Hassaskhah, J.; Mahdavi-Zafarghandi, A. From Virtual Assistant to Writing Mentor: Exploring the Impact of a ChatGPT-Based Writing Instruction Protocol on EFL Teachers' Self-Efficacy and Learners' Writing Skill. *Lang. Teach. Res.* **2024**. DOI: https://doi.org/10.1177/13621688241239764

DOI: 10.36838/IJHSR812.1

9. Ma, S. How Generative AI Tools Assist with Lesson Planning. *Edutopia* **2024**. DOI: https://www.edutopia.org/article/how-generative-ai-tools-assist-lesson-planning
10. Banerji, O. How Principals Are Outsourcing Their Busywork to AI. *Educ. Week* **2024**. DOI: https://www.edweek.org/technology/how-principals-are-outsourcing-their-busywork-to-ai/2024/01
11. Riegel, C.; Brown, B.; Medley, A. Exploring AI in Education: A Multi-State Study on K–12 Teachers' and Administrators' Knowledge, Use, and Perceptions of Artificial Intelligence. *Issues and Trends in Learning Technologies* **2025**. DOI: https://journals.library-publishing.arizona.edu/itlt/article/id/7512/
12. Cambium Learning Group. K–12 Educator + AI Survey Report. **2024**. DOI: https://www.cambiumlearning.com/user_area/uploads/CambiumK-12EducatorAISurveyReport.pdf

Author

Snowden Lange is a high school senior interested in cognitive science, psychology, and human-computer interaction. His research explores how mindset, tools, and environments shape growth. His independent project examines educator perceptions of AI in schools, using behavioral frameworks and statistical methods to understand how technology intersects with learning and development.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Predictors of Subjective Cognitive Decline in Adults Aged 45 Years and Older: Findings from the 2022 BRFSS

Juwon Lee

Groton School, 282 Farmers Row, Groton, Massachusetts, USA; japple3.141592@gmail.com

ABSTRACT: Subjective cognitive decline (SCD) is a frequent early symptom of Alzheimer's disease and other dementias. This study used national survey data and machine learning to identify factors linked to SCD. Adults aged ≥45 years who completed the 2022 Behavioral Risk Factor Surveillance System cognitive decline module were included (n=62,743). The outcome was a yes/no report of worsening confusion or memory loss. Thirty-seven variables were selected from multiple domains and analyzed using four models: least absolute shrinkage and selection operator (LASSO), random forest (RF), extreme gradient boosting (XGB), and artificial neural network (ANN). A total of 11.0% of respondents reported SCD. Model accuracy was 0.818–0.834, and the area under the receiver operating characteristic curve (AUROC) ranged from 0.751 to 0.755. Aggregated feature importance ranked depression highest, followed by chronic obstructive pulmonary disease, lack of emotional support, unemployment, arthritis, stroke, low income, no physical activity, chronic heart disease, and short sleep duration. Logistic regression confirmed these associations, with depression showing an odds ratio of 2.46 (95% CI 2.23–2.72). SCD was associated with several modifiable factors. Interventions addressing mental health, chronic disease management, social support, and adequate sleep may reduce the risk and the burden of cognitive decline.

KEYWORDS: Medical and Health Sciences, Public Health, Subjective Cognitive Decline, Depression, Chronic Diseases and Social Support.

Introduction

Subjective cognitive decline (SCD), a frequent early symptom of Alzheimer's disease (AD) and other dementias, is reported by 9.6% of adults aged ≥45 years in the United States.[1] SCD is defined as a self-reported observation of worsening confusion or memory loss and is usually identified by the Subjective Cognitive Disorder Questionnaire (SCD-Q). While the incidence of SCD is on a steady rise among older adults, the number of individuals who experience age-associated cognitive decline is expected to grow. The estimated lifetime risk of AD for a 45-year-old is currently 19.5% for women and 10.3% for men.[2] In 2023, the national cost for Alzheimer's and other dementias was an estimated $345 billion, and this figure is projected to rise to nearly $1 trillion by 2050, as stated in the Alzheimer's Association's 2023 Facts and Figures report.[3] In addition to the financial burden, the disease places a significant amount of emotional and physical distress on caregivers.[4] Given such profound impact of cognitive diseases on individuals and their families, there is an urgent need to identify modifiable risk factors that could lead to better prevention and management strategies.

A series of studies have analyzed national survey data, such as the National Health and Nutrition Examination Survey (NHANES) and the Behavioral Risk Factor Surveillance System (BRFSS), to investigate the predictors of cognitive decline. A past NHANES study suggested that certain dietary factors, such as an anti-inflammatory diet like the Mediterranean diet, could have a positive impact on cognitive performance.[5] Research using BRFSS data has identified multiple modifiable factors associated with SCD; state-level analyses revealed that poverty, diabetes, and hypertension were strongly correlated with SCD and functional impairments.[6] Smoking also showed a graded association with SCD, with the highest prevalence in current smokers compared to never smokers.[7] Sleep problems, particularly comorbid insomnia and sleep apnea, were related to almost double the odds of SCD compared to healthy sleep patterns.[8] Additionally, insufficient social and emotional support was associated with a higher prevalence of SCD.[9] These findings demonstrate the value of BRFSS data for identifying modifiable risk factors at a population level; however, the previous studies were limited by outdated datasets and a narrow focus on specific factors rather than a comprehensive evaluation of a wide range of potential predictors.

In the past decade, machine learning techniques have provided new analytic ways for large-scale survey datasets, such as NHANES and BRFSS, which enable the identification of complex relationships and the discovery of unexpected predictors in the field of epidemiology. By leveraging methods such as clustering, predictive modeling, and feature selection, machine learning can process high-dimensional data and reveal nuanced patterns often missed by traditional statistical approaches.[10] This capability is particularly valuable in exploring the interactions between socioeconomic, behavioral, and medical factors and their collective impact on cognitive health, which can lead to more tailored and effective interventions.[11]

Our study sought to apply various machine learning algorithms to identify risk factors associated with SCD, using data from the 2022 BRFSS. The findings of this research were intended to guide interventions and help reduce the burden of cognitive decline on society as well as individuals and fam-

DOI: 10.36838/IJHSR812.9

ilies. To achieve this, we selected a broad range of variables from multiple domains and applied different machine learning algorithms to determine the factors linked to SCD. Feature importance was calculated for each algorithm, and the scores were combined to identify variables that were consistently significant across all models.

Methods

Study population:

The BRFSS is a nationwide, population-based survey that includes a large sample size and collects a wide range of demographic, socioeconomic, and health-related data. Among the most recent publicly available BRFSS datasets, the 2022 BRFSS provides the widest coverage of non-clinical risk factors, including socioeconomic indicators and health-related behavioral factors. Therefore, it was selected for this study to examine the associations between memory loss and modifiable, non-clinical risk factors among U.S. adults. In the 2022 BRFSS dataset (n = 445,133), only adults aged 45 years and older (n = 311,176) were eligible for the Cognitive Decline Module. Among these respondents, adults who answered either "yes" or "no" to the question, "*During the past 12 months, have you experienced confusion or memory loss that is happening more often or is getting worse?*" were included in this analysis. The final study population consisted of 62,743 adults.

Variable selection:

In the study, the outcome variable, SCD, was a binary dependent variable determined through a yes/no response to the CMEMLOS questionnaire item. During the variable selection process, we excluded variables that were directly related to sampling methods, inherently sex-specific (e.g., mammography, prostate-specific antigen testing), consequential to SCD (e.g., decision-making difficulties and self-reported health status), or derived or overlapping variables of other variables under consideration. As a result, a total of 37 variables were selected. These variables were classified into six categories (1) Demographic variables: sex (male or female), age groups (45–65 or 65+), race (non-Hispanic white or others), living arrangements (living alone or together), marital status (married, separated/divorced/deceased, and unmarried), veteran status, and living with children; (2) socioeconomic variables: income levels (<35,000, 35,000-70,000, or 70,000+), educational level (graduated college, some college, high school or less than high school), housing type (own or rent), and current employment status; (3) medical condition variables: the presence of chronic diseases such as asthma, chronic obstructive pulmonary disease (COPD), chronic kidney disease, cancer, stroke, coronary artery disease/myocardial infarction, diabetes, arthritis, depression, tooth loss, and obesity (body mass index > 25); (4) accessibility to medical care variables: having a personal medical care provider, health insurance, and doctor visit in the preceding year; (5) health-related behavioral variables: heavy drinking (more than 14 drinks per week for males and more than 7 drinks per week for females), smoking status (never smoker, former smoker or current smoker), flu vaccination, frequency of intended physical activity in the past 30 days, and sleep duration; and (6) social and emotional well-being variables: presence of adverse childhood experiences, frequency of feeling socially isolated (always/usually or sometimes/never), and the availability of social and emotional support (always/usually or sometimes/never).

Data preprocessing:

To address multicollinearity, predictors were evaluated using the variance inflation factor (VIF). Variables with VIF values greater than 10 were removed iteratively. Missing data was handled using multivariate imputation by chained equations (MICE), which imputes missing values while accounting for relationships between variables.[12] Variables with more than 20% missing responses were excluded from imputation to reduce bias.[13] In this process, three variables were excluded: living with children (high multicollinearity), frequency of feeling isolated, and COVID-19 vaccination status (high proportion of missing responses).

Machine learning models and feature importance aggregation:

Four advanced machine learning models were used to evaluate predictors and determine their importance. These models included least absolute shrinkage and selection operator (LASSO) regression, a random forest classifier (RF), an extreme gradient boosting algorithm (XGB), and an artificial neural network (ANN). LASSO is a linear model with L1 regularization that selects features by shrinking irrelevant coefficients to zero.[14] Both XGB and RF are tree-based models capable of exploring non-linear relationships. XGB uses boosting, a sequential process where trees are constructed to correct errors from previous ones,[15] whereas RF uses bagging. In this parallel process, multiple decision trees are independently built on bootstrapped subsets of the data.[16] ANN is a nonlinear model composed of interconnected layers of neurons that can learn complex patterns.[17] Each model was trained and validated using a 5-fold stratified cross-validation procedure to ensure robustness and reduce overfitting. Model performance was evaluated using accuracy, area under the receiver operating characteristic curve (AUROC), and F-1 score.

Feature importance was primarily calculated using model-specific methods if available intrinsically. For LASSO, the absolute values of the coefficients were used, as it performs variable selection by shrinking irrelevant coefficients to zero. In XGB and RF, split-based metrics were applied, with gain used for XGB and mean decrease in impurity for RF, to evaluate how features improved model splits. In contrast to the other three algorithms, ANN lacks internal measures of feature importance; therefore, permutation-based importance scores were calculated to measure the effect of each feature on model performance.[18] To ensure comparability, the importance scores were normalized by dividing each score by the total importance within each model. Then the normalized scores were summed across models to produce aggregated feature importances to identify variables consistently important across models. Additionally, to provide an interpretable measure of the relationships between predictors and the outcome (i.e.,

SCD), the ten most significant predictors identified through machine learning were further evaluated using logistic regression and presented as odds ratios (ORs) with 95% confidence intervals (CIs).

Statistical analysis:

All analyses were conducted using Python (version 3.10.11). Data preprocessing was performed using pandas (version 2.2.3) and NumPy (version 1.25.2). LASSO and RF were implemented using scikit-learn (version 1.7.2). XGB was implemented using XGBoost (version 3.1.3), and ANN was implemented using PyTorch (version 2.4.1). GLM analyses were performed using statsmodels (version 0.14.3). A P-value of less than 0.05 was considered statistically significant.

Results

Study population characteristics:

The study included 62,743 adults aged 45 years and older, representing 14.1% of survey respondents. Among them, 6,912 (11.0%) reported SCD. Compared with adults without SCD, those with SCD were more likely to be aged ≥65 years, non-White or Hispanic, unmarried, veterans, and living alone (all p<0.001; Table 1). They had lower income, lower educational attainment, lower home ownership, and higher unemployment (p<0.001). Depression (43.7% vs. 16.5%), asthma, chronic obstructive pulmonary disease, chronic kidney disease, cancer, coronary heart disease, stroke, diabetes, and arthritis were more prevalent in the SCD group. In contrast, obesity showed no significant difference (p=0.379). Adults with SCD were less likely to report having a personal healthcare provider or health insurance and more likely to have shorter sleep duration, lower physical activity, and current smoking (all p<0.001). Adverse childhood experiences, lack of emotional support, and frequent social isolation were also more common in the SCD group (all p<0.01).

Table 1: Characteristics of adults with and without subjective cognitive decline (SCD). Adults with SCD had a higher prevalence of depression and multiple chronic diseases, along with socioeconomic disadvantage and lower levels of protective health behaviors.

	Adults without SCD (n = 55,831, 89.0%)	Adults with SCD (n=6,912, 11.0%)	P-value
Demographic variables			
Sex, male	24,894 (44.6%)	3,113 (45.0%)	0.486
Age, 65+	30,427 (54.5%)	4,003 (57.9%)	<0.001
Race, Non-white or Hispanic	7,103 (13.1%)	1,022 (15.3%)	<0.001
Veteran, yes	8,453 (15.2%)	1,324 (19.2%)	<0.001
Marital status, unmarried	21,990 (40.4%)	3,511 (52.4%)	<0.001
Living alone, yes	17,499 (31.5%)	2,587 (37.7%)	<0.001
Socioeconomic variables			
Income level			<0.001
<$35,000	18,843 (41.7%)	1,386 (24.7%)	
$35,000–70,000	14,519 (32.1%)	1,699 (30.3%)	
>$70,000	11,846 (26.2%)	2,527 (45.0%)	
Level of education			<0.001
Graduated college	24,947 (44.8%)	2,335 (33.9%)	
Some college	15,312 (27.5%)	2,085 (30.3%)	
High school	12,956 (23.3%)	1,915 (27.8%)	
Less than high school	2,464 (4.4%)	551 (8.0%)	
Home ownership, no	8,242 (14.8%)	1,715 (25.0%)	<0.001
Employment status, unemployed	32,801 (59.1%)	5,239 (76.2%)	<0.001
Medical conditions			
Depression	9,191 (16.5%)	2,991 (43.7%)	<0.001
Asthma	7,380 (13.3%)	1,489 (21.9%)	<0.001
Chronic obstructive pulmonary disease	5,333 (11.1%)	1,455 (30.7%)	<0.001
Chronic renal disease	3,199 (5.8%)	732 (10.7%)	<0.001
Cancer	8,940 (16.1%)	1,456 (21.3%)	<0.001
Chronic heart disease	6,400 (11.6%)	1,511 (22.3%)	<0.001
Stroke	2,874 (5.2%)	884 (12.9%)	<0.001
Diabetes	9,715 (17.4%)	1,873 (27.2%)	<0.001
Arthritis	24,728 (44.5%)	4,360 (63.6%)	<0.001
Teeth removal	29298 (53.9%)	4535 (67.8%)	<0.001
Obesity	37,077 (70.7%)	4,693 (71.2%)	0.379
Access to medical care			
Personal health care provider, yes	3,612 (6.5%)	399 (5.8%)	0.029
Health insurance, no	1,501 (2.8%)	217 (3.3%)	0.024
Health-related behavioral variables			
Health check-up within 1 year, no	7,066 (12.8%)	731 (10.7%)	<0.001
Heavy drinking, yes	3,531 (6.4%)	444 (6.6%)	0.731
Smoking status			<0.001
Never	31,639 (57.1%)	3,094 (45.2%)	
Former	17,954 (32.4%)	2,658 (38.8%)	
Current	5,802 (10.5%)	1,099 (16.0%)	
COVID-19 vaccination, no	2,412 (13.9%)	277 (12.9%)	0.228
Flu vaccination, no	21,661 (39.0%)	2,632 (38.4%)	0.340
Physical activity, no	13,279 (23.9%)	2,701 (39.2%)	
Sleep duration			<0.001
6 hours or less	15,739 (28.5%)	2,716 (40.1%)	
7 hours	17,402 (31.5%)	1,377 (20.3%)	
8 hours or more	22,090 (40.0%)	2,678 (39.6%)	
Social and emotional well-being variables			
Adverse childhood experience, yes	16,245 (29.1%)	2,132 (30.8%)	0.003

Performance of the prediction model:

The performance of the four machine learning models was evaluated for predicting SCD using accuracy, AUROC, and F1-score (Table 2). Overall, model performance was comparable across algorithms. Accuracy ranged from 0.818 to 0.834, and AUROC values ranged from 0.751 to 0.757. F1-scores were similar across models, ranging from 0.359 to 0.363.

Identification of significant predictors across all models:

Figure 1 shows both consistent and model-specific patterns in variable importance measures. Depression was the most important predictor, ranked first across all models. Lack of emotional support and unemployment were also consistently ranked among the top five across models. Arthritis, stroke, low income, and no physical activity were generally ranked within the top 10, indicating their relevance across multiple approaches. Some variables displayed model-specific differences. For example, COPD was ranked second in XGB and RF but had lower importance in LASSO (rank 17) and ANN (rank 14). Similarly, short sleep duration was ranked 9–10 in XGB and RF but lower in LASSO and ANN (ranks 19 and 10). Variables such as no flu vaccination, no health insurance, and heavy alcohol consumption were consistently among the least important predictors across all models.

The aggregated feature importance analysis identified depression as the most important variable, with a higher importance score (0.7625) compared to other variables (Figure 2). COPD, lack of emotional support, unemployment, and arthritis were among the top five variables, with scores ranging from 0.2200 to 0.3127. Other predictors in the top 10 included stroke, low income, no physical activity, heart disease, and short sleep duration, with scores ranging from 0.1065 to 0.1854.

DOI: 10.36838/IJHSR812.9

Table 2: Comparison of performance to predict subjective cognitive decline among the four machine-learning models. All four models demonstrated comparable predictive performance, indicating stable and reliable identification of factors associated with subjective cognitive decline.

	Accuracy	AUROC	F1-score
LASSO	0.827 (0.811–0.842)	0.752 (0.748–0.757)	0.360 (0.354–0.367)
XGB	0.834 (0.832–0.836)	0.757 (0.752–0.761)	0.363 (0.355–0.370)
RF	0.832 (0.829–0.834)	0.751 (0.748–0.754)	0.359 (0.353–0.366)
ANN	0.818 (0.798–0.837)	0.755 (0.751–0.760)	0.361 (0.354–0.367)

Data are presented as means with 95% confidence intervals in parentheses.

Abbreviations: AUROC, area under the receiver operating curve; LASSO, least absolute shrinkage and selection operator; XGB, extreme gradient boosting algorithm; RF, random forest; ANN, artificial neural network.

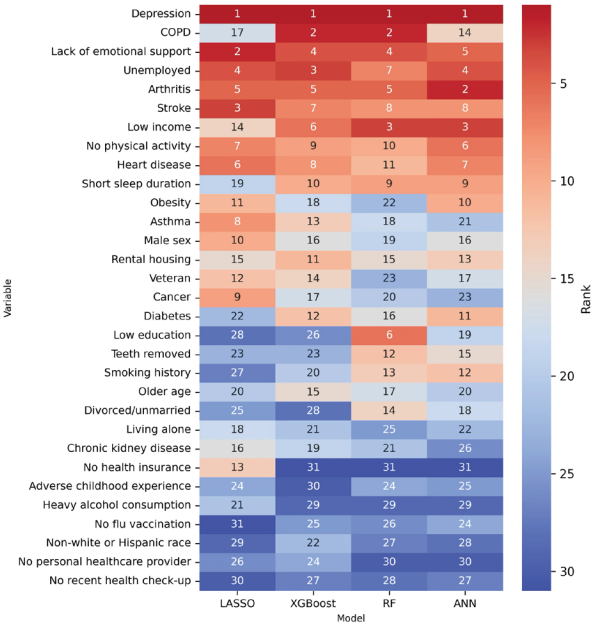


Figure 1: Variable importance ranks across models. Several factors, such as depression and lack of emotional support, maintained high importance across all models, whereas others varied substantially among models. This pattern indicates both shared and model-specific determinants of subjective cognitive decline. Darker red indicates higher importance (lower rank), and darker blue indicates lower importance (higher rank). COPD = chronic obstructive pulmonary disease; LASSO = least absolute shrinkage and selection operator; XGB = extreme gradient boosting algorithm; RF = random forest; ANN = artificial neural network.

Confirmation of significant predictors using logistic regression:

Logistic regression was conducted to evaluate the association between the top 10 variables identified through aggregated feature importance and SCD (Table 3). Depression was the strongest predictor (OR: 2.461; 95% CI: 2.229–2.717). Lack of emotional support (1.936; 1.773–2.114) and stroke (1.698; 1.556–1.854) also showed strong associations. COPD, arthritis, unemployment, and chronic heart disease showed moderate associations with ORs ranging from 1.289 to 1.516, all statistically significant ($p<0.001$). Low income, no physical activity, and short sleep duration were also significant, with smaller effect sizes (OR, 1.119–1.361).

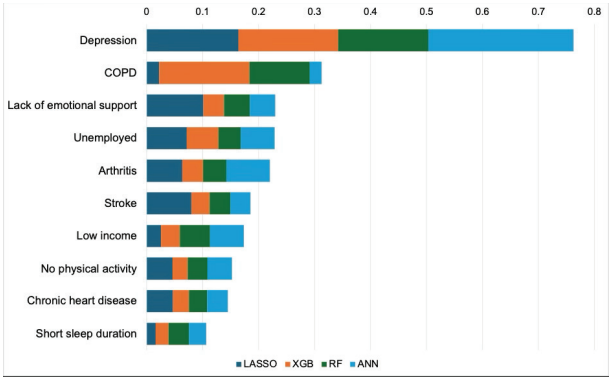


Figure 2: Top 10 variables by aggregated feature importance across the four advanced machine-learning models. Depression showed the highest overall importance, followed by COPD, lack of emotional support, unemployment, and arthritis. Stroke, low income, no physical activity, chronic heart disease, and short sleep duration also ranked among the top ten. COPD = chronic obstructive pulmonary disease; LASSO = least absolute shrinkage and selection operator; XGB = extreme gradient boosting algorithm; RF = random forest; ANN = artificial neural network.

Table 3: Odds ratios estimated by logistic regression for the top 10 predictors identified by aggregated feature importance. These results indicate that the highest-importance factors identified by the machine-learning models were statistically significant predictors of subjective cognitive decline.

Variable	Odds ratio (95% Confidence intervals)	P-value
Depression	2.46 (2.23–2.72)	<0.001
Chronic obstructive pulmonary disease	1.29 (1.17–1.42)	<0.001
Lack of emotional support	1.94 (1.77–2.11)	<0.001
Arthritis	1.46 (1.38–1.54)	<0.001
Unemployment	1.52 (1.42–1.62)	<0.001
Stroke	1.70 (1.56–1.85)	<0.001
Low-income level	1.23 (1.18–1.27)	<0.001
No physical activity	1.36 (1.29–1.44)	<0.001
Chronic heart disease	1.45 (1.35–1.55)	<0.001
Short sleep duration	1.12 (1.08–1.15)	<0.001

Discussion

Main Findings:

This study analyzed predictors of subjective cognitive decline (SCD) among adults aged 45 years and older using the 2022 BRFSS. The prevalence of SCD was 11.0%, similar to prior BRFSS reports.[19] Depression was the most significant predictor. Other significant predictors included lack of emotional support, COPD, arthritis, unemployment, stroke, low income, no physical activity, chronic heart disease, and short sleep duration.

Depression and Dementia Risk:

A series of large population-based studies has demonstrated that depression significantly increases the risk of cognitive decline and dementia.[20-22] Additionally, each episode of depression is associated with an increased risk of all types of dementia, with a 14% increase per episode.[23] Multiple biological mechanisms are known to link depression to dementia, including steroid and neurotransmitter dysregulation, hippocampal atrophy, inflammation, increased deposition of beta-amyloid, and cerebrovascular changes.[24,25] Animal and human studies have suggested that antidepressant medication may reduce dementia risk by attenuating hippocampal structural damage, enhancing neuroplasticity, and reducing amyloid and tau pathology.[26] Depression should therefore be considered an important modifiable risk factor in strategies to prevent or delay dementia.

Chronic Medical Conditions:

Managing chronic medical conditions such as arthritis, COPD, and stroke may reduce the risk of SCD.[27-29] These conditions are associated with a higher prevalence of SCD and may contribute to cognitive impairment through mechanisms including inflammation and vascular changes. Addressing chronic diseases may also alleviate depression, an established risk factor for SCD. Stroke, identified as an important factor in our analysis, is a chronic condition that can be both prevented and managed. Physical activity has also been reported to influence the relationship between depression and cognition.[30] Recent studies have examined this association further. A randomized clinical trial showed that cognitive remediation combined with transcranial direct current stimulation slowed cognitive decline in older adults with remitted major depressive disorder or mild cognitive impairment.[31] These findings suggest that targeted interventions may mitigate the cognitive effects of depression, although the link between depression treatment and dementia prevention remains complex and requires further study.

Emotional and Social Support:

Emotional support is a crucial factor for mental and physical well-being. Studies have demonstrated that adequate emotional support can improve cognitive function and reduce memory decline in different populations.[32,33] Social connections may also protect against depression, which is closely related to SCD.[34] In addition, individuals with stronger social engagement often maintain positive relationships, which may lower the risk of dementia.[35] In this context, strengthening emotional and social support may help reduce the risk of SCD and dementia.

Sleep and Cognitive Function:

Inadequate sleep duration was identified as a predictor of SCD in our analysis. Sleep is a modifiable behavior and may represent a potential target for intervention, unlike predictors such as economic hardship and education level, which are less amenable to change. An inverted U-shaped association has been reported between sleep duration and cognitive function, with both short and long sleep linked to poorer cognitive outcomes and subsequent decline.[31,36] Extreme sleep durations (≤4 or ≥10 hours per night) have been associated with lower baseline cognitive function and with faster decline during follow-up. These findings underscore the need to monitor and adjust sleep duration. At the same time, further studies are required to clarify underlying mechanisms and to assess the role of sleep quality as well as duration.

Data Processing Considerations:

Regarding the preprocessing of our data, the cohort was highly imbalanced, with only 11% of respondents classified as having SCD. In addition, the overall F1-scores were relatively low across models, which probably reflected the difficulties in accurately identifying the minority SCD group, i.e., participants with SCD, despite high overall accuracy and AUROC. Oversampling techniques such as the Synthetic Minority Over-sampling Technique combined with Tomek Links (SMOTE-Tomek) are often used in previous studies to address imbalance by generating synthetic minority samples.[37] While these methods can enhance predictive performance, they may also distort the original associations between variables and outcomes, which was a central concern in this study.[38] Our primary aim was not to optimize model performance but to identify and evaluate predictors of SCD. To this end, we aggregated feature importance scores across models, which helped reduce bias inherent to any single method and provided more robust rankings of predictors. This approach also has limitations, as normalization and aggregation may dilute model-specific signals and obscure predictors that are important in one model but not in others. In some cases, consensus across models may overemphasize shared predictors while underrepresenting unique insights from individual models.

Limitations:

This study has several limitations. The cross-sectional design of the BRFSS limits causal inference. Reliance on self-reported measures of both SCD and predictor variables may also introduce recall bias. Excluding variables with a high proportion of missing responses, such as adverse childhood experiences, may have reduced the scope of the analysis. Future research should use longitudinal designs to clarify causal pathways and validate these findings.

■ Conclusion

In this study, we identified key predictors of SCD using machine learning models applied to 2022 BRFSS data. Depression was the most consistent risk factor, and chronic medical conditions, inadequate emotional support, and insufficient sleep were also significant contributors, indicating the multifactorial nature of SCD. These results clarify the associations between multiple predictors and SCD. Combining statistical approaches with machine learning may improve the interpretability of large survey data and inform public health strategies.

■ References

1. Wooten, K. G.; McGuire, L. C.; Olivari, B. S.; Jackson, E. M. J.; Croft, J. B., Racial and Ethnic Differences in Subjective Cognitive Decline - United States, 2015-2020. *MMWR Morb Mortal Wkly Rep* **2023**, *72* (10), 249-255.
2. Murman, D. L., The Impact of Age on Cognition. *Semin Hear* **2015**, *36* (3), 111-21.
3. 2023 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* **2023**, *19* (4), 1598-1695.
4. Zhang, J.; Wang, J.; Liu, H.; Wu, C., Association of dementia comorbidities with caregivers' physical, psychological, social, and financial burden. *BMC Geriatrics* **2023**, *23* (1), 60.
5. Xie, S.; Li, Z.; Yao, Q.; Zhang, Y.; Ou, Y., Adherence to Mediterranean diet and female urinary incontinence: Evidence from the NHANES database. *PLoS One* **2024**, *19* (10), e0311771.
6. de Havenon, A.; Stulberg, E. L.; Littig, L.; Wong, K. H.; Sarpong, D.; Li, V.; Sharma, R.; Falcone, G. J.; Williamson, J. D.; Pajewski, N. M.; Gottesman, R. F.; Brickman, A. M.; Sheth, K. N., Socioeconomic and medical determinants of state-level subjective cognitive

DOI: 10.36838/IJHSR812.9

decline in the United States. *Alzheimer's Dement* **2024**, *20* (11), 7567-7579.
7. Rajczyk, J. I.; Ferketich, A.; Wing, J. J., Relation Between Smoking Status and Subjective Cognitive Decline in Middle Age and Older Adults: A Cross-Sectional Analysis of 2019 Behavioral Risk Factor Surveillance System Data. *J Alzheimers Dis* **2023**, *91* (1), 215-223.
8. Huang, J.; Spira, A. P.; Perrin, N. A.; Ellis, A.; Hsu, E. C.; Kaufmann, C. N.; Li, J., Latent classes of sleep problems and subjective cognitive decline among middle-aged and older adults in the United States. *Arch Gerontol Geriatr* **2025**, *129*, 105657.
9. Weng, X.; George, D. R.; Jiang, B.; Wang, L., Association between subjective cognitive decline and social and emotional support in US adults. *American Journal of Alzheimer's Disease & Other Dementias®* **2020**, *35*, 1533317520922392.
10. Ngiam, K. Y.; Khor, W., Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology* **2019**, *20* (5), e262-e273.
11. Topol, E. J., High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* **2019**, *25* (1), 44-56.
12. Van Buuren, S.; Groothuis-Oudshoorn, K., mice: Multivariate imputation by chained equations in R. *Journal of statistical software* **2011**, *45*, 1-67.
13. Jakobsen, J. C.; Gluud, C.; Wetterslev, J.; Winkel, P., When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. *BMC Medical Research Methodology* **2017**, *17* (1), 162.
14. Tibshirani, R., Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **1996**, *58* (1), 267-288.
15. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, August 13-17, 2016; pp 785–794. DOI: 10.1145/2939672.2939785.
16. Breiman, L., Random forests. *Machine learning* **2001**, *45* (1), 5-32.
17. LeCun, Y.; Bengio, Y.; Hinton, G., Deep learning. *Nature* **2015**, *521* (7553), 436-444.
18. Khan, A.; Ali, A.; Khan, J.; Ullah, F.; Faheem, M., Using Permutation-Based Feature Importance for Improved Machine Learning Model Performance at Reduced Costs. *IEEE Access* **2025**.
19. Olivari, B. S.; Baumgart, M.; Taylor, C. A.; McGuire, L. C., Population measures of subjective cognitive decline: A means of advancing public health policy to address cognitive health. *Alzheimer's & Dementia: Translational Research & Clinical Interventions* **2021**, *7* (1), e12142.
20. Mirza, S. S.; Wolters, F. J.; Swanson, S. A.; Koudstaal, P. J.; Hofman, A.; Tiemeier, H.; Ikram, M. A., 10-year trajectories of depressive symptoms and risk of dementia: a population-based study. *Lancet Psychiatry* **2016**, *3* (7), 628-35.
21. Elser, H.; Horváth-Puhó, E.; Gradus, J. L.; Smith, M. L.; Lash, T. L.; Glymour, M. M.; Sørensen, H. T.; Henderson, V. W., Association of Early-, Middle-, and Late-Life Depression With Incident Dementia in a Danish Cohort. *JAMA Neurol* **2023**, *80* (9), 949-958.
22. Li, C.; Wang, W.; Wei, Y. X.; Lu, K.; Wang, J. Y.; Yao, M. H.; Du, Q. Q.; Li, X. L.; Li, S.; Tian, X. Y.; Yin, F.; Zhang, T.; Ma, Y., Association between cognitive decline and depression in middle-aged and older adults: Findings from six large cohorts in different countries. *J Affect Disorders* **2025**, *371*, 215-223.
23. Dotson, V. M.; Beydoun, M. A.; Zonderman, A. B., Recurrent depressive symptoms and the incidence of dementia and mild cognitive impairment. *Neurology* **2010**, *75* (1), 27-34.
24. Dafsari, F. S.; Jessen, F., Depression-an underrecognized target for prevention of dementia in Alzheimer's disease. *Transl Psychiat* **2020**, *10* (1).
25. Byers, A. L.; Yaffe, K., Depression and risk of developing dementia. *Nat Rev Neurol* **2011**, *7* (6), 323-331.
26. Stephan, B. C.; Brayne, C.; Savva, G. M.; Matthews, F. E., Occurrence of medical co-morbidity in mild cognitive impairment: implications for generalisation of MCI research. *Age and ageing* **2011**, *40* (4), 501-507.
27. Jinshil, K.; Eunok, P.; Minjeong, A., The cognitive impact of chronic diseases on functional capacity in community-dwelling adults. *Journal of Nursing Research* **2019**, *27* (1), e3.
28. Bakouni, H.; Guerra, S. G.; Chudzinski, V.; Berbiche, D.; Vasiliadis, H. M., One-year prospective study on the presence of chronic diseases and subsequent cognitive decline in older adults. *J Public Health-Uk* **2017**, *39* (4), E170-E178.
29. Singh, B.; Mielke, M. M.; Parsaik, A. K.; Cha, R. H.; Roberts, R. O.; Scanlon, P. D.; Geda, Y. E.; Christianson, T. J.; Pankratz, V. S.; Petersen, R. C., A Prospective Study of Chronic Obstructive Pulmonary Disease and the Risk for Mild Cognitive Impairment. *Jama Neurology* **2014**, *71* (5), 581-588.
30. Hu, L.; Smith, L.; Imm, K. R.; Jackson, S. E.; Yang, L., Physical activity modifies the association between depression and cognitive function in older adults. *J Affect Disorders* **2019**, *246*, 800-805.
31. Ma, Y. J.; Liang, L. R.; Zheng, F. F.; Shi, L.; Zhong, B. L.; Xie, W. X., Association Between Sleep Duration and Cognitive Decline. *Jama Netw Open* **2020**, *3* (9).
32. Fratiglioni, L.; Paillard-Borg, S.; Winblad, B., An active and socially integrated lifestyle in late life might protect against dementia. *Lancet Neurol* **2004**, *3* (6), 343-353.
33. Weng, X. R.; George, D. R.; Jiang, B. B.; Wang, L., Association Between Subjective Cognitive Decline and Social and Emotional Support in US Adults. *Am J Alzheimers Dis* **2020**, *35*.
34. Schwarzbach, M.; Luppa, M.; Forstmeier, S.; König, H. H.; Riedel-Heller, S. G., Social relations and depression in late life: A systematic review. *Int J Geriatr Psych* **2014**, *29* (1), 1-21.
35. Wang, H. X.; Karp, A.; Winblad, B.; Fratiglioni, L., Late-life engagement in social and leisure activities is associated with a decreased risk of dementia: A longitudinal study from the Kungsholmen project. *Am J Epidemiol* **2002**, *155* (12), 1081-1087.
36. Leng, Y.; Yaffe, K., Sleep Duration and Cognitive Aging-Beyond a U-Shaped Association. *Jama Netw Open* **2020**, *3* (9).
37. Wang, Z.; Wu, C. H.; Zheng, K. F.; Niu, X. X.; Wang, X. J., SMOTETomek-Based Resampling for Personality Recognition. *IEEE Access* **2019**, *7*, 129678-129689.
38. Buda, M.; Maki, A.; Mazurowski, M. A., A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks* **2018**, *106*, 249-259.

■ Author

Juwon Lee is a student at Groton School, passionate about neurological disorders and brain science. She is eager to explore behavioral and biological factors associated with these conditions. She hopes to contribute to research that advances prevention, early detection, effective intervention, and equitable access to health care.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Brick-by-Brick: Enhancing Creativity and Problem-Solving in Under-Resourced Schools Through LEGO Interventions

Vihaan Tannan

Gilbert Building, Babulnath, 2nd Cross Road, Mumbai, Maharashtra 400007, India; vihaantannan10@gmail.com
Mentor: Aashna Saraf

ABSTRACT: This study examines the effects of traditional LEGO brick interventions on creativity and problem-solving among primary school students from low socio-economic backgrounds in under-resourced Indian schools. While extensive research has investigated LEGO robotics in Western contexts, limited empirical evidence exists regarding the utilization of conventional LEGO bricks in developing countries. Creativity, conceptualized by Torrance as the ability to generate novel ideas and relationships, and problem-solving, defined as the application of higher-order cognitive functions to achieve goals, are critical skills often underdeveloped in India's rigid, instruction-focused education system. This research employs a structured "Brick-by-Brick" intervention with 3rd and 4th-grade students, utilizing the Alternative Uses Test (AUT) as a pre- and post-intervention measure of creativity. It is hypothesized that participants will demonstrate significant gains in creativity and problem-solving scores following the intervention. By addressing gaps in context, population, and longitudinal research, this study contributes to the evidence base for low-cost, scalable pedagogical tools that foster 21st-century skills in socioeconomically disadvantaged educational settings.

KEYWORDS: LEGO Bricks, Creativity, Problem-Solving, Socio-Economic Status, Alternative Uses Test (AUT).

Introduction

This research paper aims to understand the improvement traditional LEGO bricks can make to creativity and problem-solving in under-resourced schools in India, with students and parents belonging to a low socio-economic status. The term creativity was defined by Torrance March as "the capacity to detect gaps, propose various solutions to solve problems, produce novel ideas, re-combine them, and intuit a novel relationship between ideas."[1] For this research, the term "problem solving" is established as "the process by which individuals attempt to overcome difficulties, achieve plans that move them from a starting situation to a desired goal, or reach conclusions through the use of higher mental functions, such as reasoning and creative thinking."[2] While the present study focuses on creativity as measured by the AUT, references to problem-solving are framed as related cognitive processes supported by task design rather than as a directly measured outcome using a validated problem-solving instrument. Finally, according to the American Psychology Association, socioeconomic status "encompasses not only income but also educational attainment, occupational prestige, and subjective perceptions of social status and social class. SES encompasses quality-of-life attributes and opportunities afforded to people within society and is a consistent predictor of a vast array of psychological outcomes."[3]

Traditional schools in India follow a rigid curriculum, with little to no freedom given to the teachers to improve the quality of education. Although there has also been some priority toward a holistic education by the Indian government, which highlights the importance of the environment, there has been little action taken towards the enhancement of creativity and problem-solving skills. Additionally, traditional schooling focuses on well-structured instructions, which directly hinders problem-solving and creativity. A majority of LEGO-based research is carried out using LEGO robotics, while very little research uses conventional LEGO bricks [such as][4]. A few attempts have been made to define the relationship between creativity and LEGO bricks. One study showed that using pre-established, instruction-based LEGO kits has an inverse effect on creativity. This 90-minute study was conducted with 136 undergraduate college students and aimed at differentiating the effects of ill-defined vs. well-defined tasks on creative thinking.[5]

Another study at Canterbury Christ Church University College explored the positive correlation between limiting the number of LEGO bricks used and creativity.[6] The correspondence between LEGO and problem-solving has not been explored to the same depth as the one between creativity and LEGO. Previous research on this topic ranged from using LEGO with 3rd-grade students to higher education learners. For example, a study conducted with 40 elementary school children saw an increase in mean scores from 9.22 (pre-intervention) to 12.2 (post-intervention). Whereas, in a study conducted in China, the use of LEGO has been used by engineering students to heighten problem-solving.[7]

While some studies provide successful primary research, multiple pieces are missing in the research. Firstly, many studies have taken place in Western contexts, China,[8] and Hong Kong, but no study related to traditional LEGO bricks, creativity, and problem-solving has been conducted with the Indian population (adults or children). Secondly, this research also aims to work with students from a lower socioeconomic background. A majority of studies are based in private schools, such as the

DOI: 10.36838/IJHSR812.15

D1 district in Iran.[9] Additionally, a longitudinal study has not been conducted to explore the effects of creativity and problem-solving using LEGO bricks. The research, which was conducted for younger students in Tabriz, Iran, aimed to discover the advantages of using Lego in science.[10]

With the rise of artificial intelligence, creativity and problem-solving are more important than ever. Artificial intelligence can help conduct everyday tasks, but we need these skills to build originality. Students in private schools, studying in international programs, still might have the opportunity to learn with a holistic education, but children belonging to a lower socioeconomic status have no way to learn these important life skills. Moreover, these skills will help them excel in different fields that require a certain way of thinking. If the Lego toolkits do improve creativity in lower socioeconomic backgrounds, the goal of the study is to give access to the toolkits to multiple government schools and education-focused NGOs.

As a result, it is hypothesized that students in under-resourced Indian schools will demonstrate significantly higher scores on the Alternative Uses Test (AUT) after participation in the traditional LEGO-based intervention (Brick-by-Brick) compared to their baseline scores.

Methodology

Participants:

This research aimed to study whether exposure to the 8 Lego problem-solving and creativity toolkits over 8 days leads to significant improvement in participants' (3rd and 4th graders, aged 11-12 years) problem-solving abilities and creative thinking, as measured by the Alternate Usage Test.

Students aged between 11 and 12 years of age (3rd & 4th grade) were recruited for this study. Ensuring their residence in Mumbai (India), selected students were from lower SES (the children participating were either paying less than ₹15,000($174) in yearly fees, or were sponsored by the Teach For India Foundation). Recruitment was undertaken through a government school in Mumbai, and a convenience sampling strategy was used to include an equal number of boys and girls. An important criterion for consideration was a lack of previous exposure to LEGO bricks. Participants were given the option to be a part of the study by the principal as a free after-school activity.

The data was gathered from the school and parents of the children through a demographic survey and past exposure to Lego bricks of the children. The Lego bricks were assembled into 8 unique kits. Each kit had its own difficulty and required the students to work together in groups of 4-5 to build creativity and solve problems.

Materials and Measures:

The first toolkit was built to promote creativity among learners. This was done to try to educate students using a holistic curriculum. This type of education would not only improve creativity but also allow the students to enjoy the learning process.

- ***Storytelling with Bricks:***
  - The first kit, "Storytelling with Bricks", Figure 1, helped students get a feel of the pieces through building objects and creating stories as a group.

Figure 1: Storytelling with Bricks.

- ***Stability Test:***
  - Next, the students were made to perform a stability test as shown in Figure 2. The students built a bridge-like structure with the stands 6 inches apart. After this, a one-kilogram weight was placed on the structure to check its integrity.

Figure 2: Stability test.

- ***Drawing to Lego:***
  - Thirdly, the "Drawing to Lego", Figure 3, activity was introduced. This consisted of students drawing anything they could imagine and converting that image into a tangible and 3-dimensional structure made of Lego.

Figure 3: Drawing to Brick Models.

- ***Non-functional to functional:***
  ○ For toolkit 4, the students received cars and planes, which were incomplete and 'non-functional' as seen in Figure 4. The activity aimed to fix these objects to the best of the students' ability.

Figure 4: Fixing Non-Functional Objects.

- ***Instructions to Object:***
  ○ Next, learning was incorporated into an activity by teaching the students the intricacies of the concept of 'inventions' (Figure 5). After learning what inventions are, the candidates start building a unique invention that hasn't been built before.

Figure 5: Teach and Build.

- ***2-D to 3-D:***
  ○ For activity 6, the toolkits contained cards with pictures of different shapes on them, which can be seen in Figure 6. The goal of the activity was to make the shapes into objects made of Lego.

Figure 6: 2-D to 3-D.

- ***Lego Relay:***
  ○ The next day, the students performed a Lego relay as shown in Figure 7. In this exercise, each group member built half of an object and passed it to the next member. The new member has to add to that object without communicating with other members. This princess continued until every member worked on every object.

Figure 7: Brick Relay.

- ***Rube Goldberg Machine:***
  ○ Lastly, the candidates took part in a culminating activity, in which they built a Rube Goldberg Machine (Figure 8). The objective was to build a ramp to take a marble from the top to the bottom. The students worked creatively on balancing the ramp perfectly to align with the wooden block bases.

Figure 8: Rube Goldberg Machine.

To quantify the data received pre- and post-testing, the Alternative Use Test (J.P. Guilford) was used. The purpose of this test is to generate useful and original uses for everyday objects. The Alternative Usage Test is used to measure an individual's creativity. The goal of the test is to name as many possible unique uses for 4 regular objects (a brick, a newspaper, a hanger, and an envelope). The test is scored based on 4 aspects: originality, fluency, categories, and elaboration.

Multiple software applications were used during this study. Canva played an important role in creating the design for the toolkits. Microsoft software was also used to digitize the data from the Alternative Use Test. Moreover, studio.io was used to list the Lego bricks needed and to create the instruction manuals for some of the activities.

DOI: 10.36838/IJHSR812.15

Ethical Considerations:

Permission to conduct the study was obtained from the school administration. Written parental consent and verbal student assent were secured prior to participation. Participation was voluntary, and students were informed that they could withdraw at any time without penalty. All data were anonymized prior to analysis.

Procedure

The process began by brainstorming different ideas for the activities in the toolkit. 12 possible ideas for toolkits were listed. Next, each toolkit was broken down into the following categories: materials required, aim, and procedure. After analyzing each one, the 8 most effective toolkits were chosen based on the skills the kit develops. Then, the design for the boxes was created, within which the Lego bricks, a guide, and other miscellaneous items were included. These kits were then assembled one at a time with bricks, sand timers, and miscellaneous items into the boxes; the kits were also checked multiple times to ensure no box was incorrectly assembled. Next, the pretest was conducted using the Alternative Use Test. The students were given 2 minutes per category (newspaper, brick, envelope, hanger). To accommodate language diversity, the AUT instructions were read aloud with clarifications provided in Hindi when needed. Students were permitted to respond in either Hindi or English, and responses were later translated where necessary for scoring. For the next 8 days, each of the different kits was deployed in order.

Each activity was given an average time of 1 hour in the same classroom. There were 2 teachers present on behalf of the school. These teachers were given the role of making sure the class was quiet when the instructions were being given. The teacher-to-student ratio was 1:20. All instructions were given in the toolkit and were read out as given in the toolkit.

Design:

An experimental research design was used. The independent variable was the use of the Brick-by-Brick toolkits. The dependent variables were creativity and problem-solving measured through the Alternative Use Test. There were multiple control variables that were kept constant during the pre- and post-test and the study period. The first control variables were the age, grade, and socio-economic status of the students; these were major constants during our study. Next, the variables in the school were kept constant through the location and time of day of the study. Moreover, the duration of the study on each day was averaged at 1 hour.

Data Analysis:

A Paired Sample T-Test was used to compare the results between the pre-test and the post-test. First, the data was gathered on an Excel sheet to tabulate all of the results. To analyze the data, JASP software was used. Additionally, data integrity was used throughout the entire process. During the pre- and post-test, no talking between participants, teachers, or the instructors was permitted unless a relevant question was asked.

There were a few results that could not be used due to reasons that were not controlled by the researchers. A few questionnaires were nameless, which prevented them from being used as a result. Another group of students was present for only the pre-test or post-test. This meant no change in creativity could be observed.

Standard AUT scoring conventions were used to score the responses of the participants in four dimensions, namely originality, fluency, categories (flexibility), and elaboration. Two independent raters were trained on the same scoring rubric and sample answers before the scoring. The differences would be discussed and agreed upon. The score on originality was determined using the infrequency of responses statistically in the sample as the measure, and category score as the number of conceptual groupings (distinct) represented.

Results

Students in under-resourced Indian schools demonstrated significantly higher scores on the Alternative Uses Test (AUT) following participation in the traditional LEGO-based intervention, Brick-by-Brick, compared to their baseline scores. The AUT assesses creativity by requiring participants to generate multiple solutions to a single problem across four dimensions: originality (uniqueness of responses), fluency (number of ideas), categories (range of ideas), and elaboration (level of detail). The study included n = 30 participants, with a mean age of 10.20 years (SD = 0.961). Measures of skewness and kurtosis fell within the acceptable range, indicating normality of the sample distribution. A paired-sample t-test was conducted, confirming that the differences observed were statistically significant. Given the sample size and normal distribution, the results can be considered representative of the target population. Of the students initially recruited, 30 participants provided complete matched pre- and post-test data and were included in the final analyses; incomplete or unmatched responses were excluded.

Table 1: Descriptive Statistics for Age. The table shows the statistics for the age range of 9-12.

Statistic	Value
Mean	10.20
Standard Deviation	0.96
Skewness	0.07
Standard Error of Skewness	0.43
Kurtosis	-1.14
Standard Error of Kurtosis	0.83
Minimum	9.00
Maximum	12.00

Note: N = 30.

The final sample consisted of 30 students enrolled in 3rd and 4th grade, ranging in age from 9 to 12 years (M = 10.20, SD = 0.96).

In this experiment, each participant completed the AUT (Alternative Uses Test) twice, before and after using the Brick-by-Brick Toolkits. There were a total of 8 kits, which consisted of the following:

- Storytelling with Bricks
- Stability Test
- Drawing to Lego
- Non-functional to functional

- Instructions to Object
- 2-D to 3-D
- Lego Relay
- Rube Goldberg Machine

Measure 1	Measure 2	t	df	p
Pre-test Sum_o	Post-test Sum_o	-4.891	29	< .001
Pre-test Sum_f	Post-test Sum_f	-7.360	29	< .001
Pre-test Sum_c	Post-test Sum_c	-5.152	29	< .001
Pre-test Sum_e	Post-test Sum_e	-1.216	29	.117

Although no formal correction for multiple comparisons was applied due to the exploratory nature of the study, three of the four AUT subscales demonstrated p-values below .001, which would remain significant under conservative correction procedures.

Given the exploratory and classroom-based nature of this study, statistical inference focused on within-subject comparisons using paired-sample t-tests. While effect sizes and confidence intervals are recommended for interpreting magnitude, the present analysis emphasizes statistical significance and directionality of change. Future work will incorporate full descriptive statistics and standardized effect size estimates to support meta-analytic comparison.

For all tests, the alternative hypothesis specifies that Measure 1 is less than Measure 2. For example, Pre-test Sum_o is less than Post-test Sum_o.

Sum_o: Sum of all originality scores

Sum_f: Sum of all fluency scores

Sum_c: Sum of all categories' scores

Sum_e: Sum of all elaboration scores

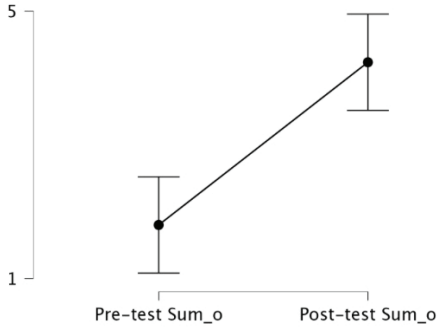


Figure 9: Comparison of the originality pretest and post-test. Shows the increase in originality from pre-test to post-test.

Comparison of the average test score in the originality pretest and post-test (before and after using the Brick-by-Brick toolkits).

A Paired Sample T-test was used to compare the paired data for an improvement in originality. A p-value of <0.001 was yielded, which was below that of the set alpha level of 0.05. This suggests that the hypothesis was supported; participants performed better on the Paired Sample T-test in originality after using the Brick-by-Brick toolkits than before using the Brick-by-Brick toolkits.

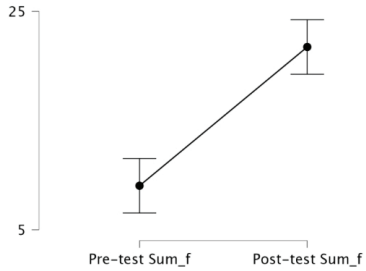


Figure 10: Comparison of fluency pretest and post-test. Shows the increase in fluency from pre-test to post-test.

Comparison of the average test score in fluency pretest and post-test (before and after using the Brick-by-Brick toolkits).

A Paired Sample T-test was used to compare the paired data for an improvement in fluency. A p-value of <0.001 was yielded, which was below that of the set alpha level of 0.05. This suggests that the hypothesis was supported; participants performed better on the Paired Sample T-test in fluency after using the Brick-by-Brick toolkits than before using the Brick-by-Brick toolkits.

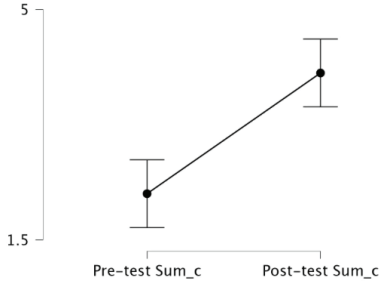


Figure 11: Comparison of categories' scores pretest and posttest. Shows the increase in categories from pre-test to post-test.

Comparison of the average test score in categories pretest and posttest (before and after using the Brick-by-Brick toolkits).

A Paired Sample T-test was used to compare the paired data for an improvement in categories. A p-value of <0.001 was yielded, which was below that of the set alpha level of 0.05. This suggests that the hypothesis was supported; participants performed better on the Paired Sample T-test in categories after using the Brick-by-Brick toolkits than before using the Brick-by-Brick toolkits.

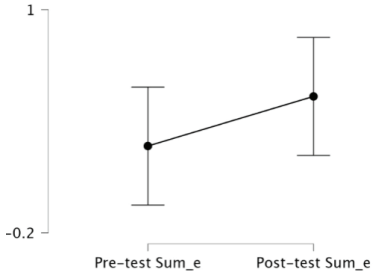


Figure 12: Comparison of elaboration scores pretest and post-test. Shows a smaller increase in elaboration from pre-test to post-test.

Comparison of the average test score in the elaboration pretest and post-test (before and after using the Brick-by-Brick toolkits).

DOI: 10.36838/IJHSR812.15

A Paired Sample T-test was used to compare the paired data for an improvement in elaboration. A p-value of 0.117 was yielded, which was above that of the set alpha level of 0.05. This suggests that the hypothesis was not supported. This could be because the students in this study were not fluent in English, which restricted them from forming complete sentences. If the study were re-conducted and the participants had English as their first language rather than their second language, the p-value of elaboration could be expected to decrease. The lack of a statistically significant improvement in elaboration may be partly attributable to language constraints, as many participants were not fluent in written English, potentially limiting their ability to produce detailed responses despite conceptual understanding.

Discussion

This study was designed to explore the relationship between creativity, problem-solving, and Lego Brick-by-Brick toolkits. The aim was to test whether exposure to 8 Lego Brick-by-Brick toolkits over 8 days improves creativity and problem-solving in 9-12-year-old students. The objective of this study was to examine whether structured Lego-based activities would significantly enhance scores on the Alternative Uses Test (AUT). The Alternative Uses Test (AUT) directly corresponds to creativity and has been used in multiple studies. For this study, the AUT test was conducted before and after the Lego toolkits were deployed.

The setting of this study was in lower socio-economic setting schools in Mumbai, India. Students aged between 9 and 12 (3rd and 4th graders) and having no prior experience with LEGO were recruited for this study. Most young students in India do not get the opportunity to learn and grow using tools such as LEGO; therefore, this study tried to address the gaps in creative pedagogy resources for underprivileged learners.

Upon completion of the LEGO toolkit after 8 days, the findings indicated significantly improved results when compared to the pre-test. This improvement shows the effectiveness of the Brick-by-Brick Lego Toolkits and the impact they can have on students in India. The improvements varied for each student, but most students improved in Fluency, Categories, and Originality. There were major advancements in the fluidity of their thought process; the students came up with more ideas than before, and listed their thoughts down with precision. Due to this, there was more flexibility in the uses. Additionally, there were a variety of different genres of ideas in the post-test. Out of all of the kits, the "Drawing to Lego" and the "Rube Goldberg Machine" stood out as some of the most creative toolkits. The "Drawing to Lego" activity allowed the students to make their drawings come to life with Lego Bricks. This allowed them to think and imagine for themselves with no preset template. The "Rube Goldberg Machine" was really exciting for the students. They competed with each other to build the fastest and most creative way to transport a marble

from one place to another. The exposure to the toolkits had a measurable, positive effect on creativity and problem-solving.

This form of learning is holistic and hands-on, which propels a higher order of thinking. In this digital age, having a tangible experience with the Lego toolkits helped students integrate learning and play, while still conforming to the constructivist learning theory. This form of learning and testing follows Guilford's model of divergent thinking, which is measured by the Alternative Uses Test (AUT). His model requires participants to find different and creative uses for everyday objects that are unique. Moreover, the activity was done in groups of 4 or more (5 to 6)-this enhanced peer interaction, which fostered peer learning and problem solving beyond individual creativity. Working in groups brought a completely new dynamic to the way the activity was conducted; students listened to ideas from their teammates and learnt to collaborate with people they might not have worked with before. There were multiple places where some students adapted faster than others. For example, some students preferred doing functional tasks such as the "Non-Functional to Functional Objects". In contrast, others excelled in more open-ended and creative tasks such as the "Storytelling with Bricks" activity. Additionally, every student started at a different entry point in the pretest, which varied their scope to improve.

The results of this study show the major impact the Lego toolkits can have on students' creativity and problem-solving. These kits are scalable throughout India and can be a low-cost option to build creativity-based interventions in under-resourced classrooms. Furthermore, these toolkits can be integrated with government initiatives such as India's National Education Program 2020, which aims to revolutionize the Indian education system through lower and higher education.

Strengths and Limitations

This is a study of its kind, which uses students from lower socioeconomic status in India, and there has been no prior research done on the impact of Lego on this group. This experiment also has definitive pre- and post-test results from the students. Moreover, these toolkits and activities are carefully curated and designed over a long period of time. They are made to balance functional and abstract tasks while continuously improving creative skills.

Nonetheless, the sampling strategy (convenience sampling) from one school in Mumbai makes generalizability difficult. The interventions were short-term as they only lasted 8 days in total, and longer and sustained efficacy cannot be measured. Additionally, all results were based on the Alternative Uses Test (AUT); this indicates that there was only a single creativity measure. Other dimensions, like collaboration or emotional engagement, need further exploration.

One of the weaknesses of this study is that there was no control or comparison group. Consequently, post-test improvements in AUT scores might be in part due to practice effects of repeated exposure to divergent thinking tasks and not to the intervention itself. Such divergent thinking measures as the Alternative Uses Test have been reported to be susceptible to familiarity effects. Thus, the results can be regarded as the initial sign of the relationship and not a causal effect. Research

conducted in the future needs a randomized or waitlist control group and parallel-type AUT prompts to better isolate intervention effects.

The study was also based on one standardized measure of creativity, which, again, constrained convergent validity. The research was limited to the use of the AUT because it was free of license and had creative commons, and future research would consider using various instruments that have high validity to measure creativity, executive function, or problem-solving to enhance construct interpretation.

Future Research Directions

There are multiple ways of building and enhancing this research. Expansion to larger and more diverse samples, such as other socioeconomic statuses, different cities, and more rural settings, is needed. This will help in gaining more accurate results with a diverse group of communities. There is also the opportunity to conduct a long-term study over multiple months. This will help track the longevity and persistence of the gains in creativity and problem-solving. These toolkits can also be deployed at NGOs and Foundations aiming to shrink the education disparity in India. This disparity exists in most local schooling systems in the country. These organizations can also help bridge the socioeconomic status-based skill gap using playful and powerful tools such as the Lego toolkits. These kits include instructions, but are best used with a mentor or teacher for guidance. Due to this, the teachers need to go through training on how to use these kits to the best of their ability. However, these kits can also be used to highlight the importance of structured facilitation with minimal resources.

From a scalability perspective, the Brick by Brick intervention is designed to be low-cost and reusable, relying on conventional LEGO bricks and printed task cards rather than electronic components. With minimal teacher training focused on facilitation rather than technical instruction, the model is suitable for deployment in government schools and NGO-run learning environments.

Conclusion

The Lego toolkit program significantly improved the creativity and problem-solving abilities of lower socioeconomic status 9–12-year-olds. The findings and results highlight the power of structured and creative play in fostering divergent thinking and collaboration. These results offer evidence to collaborate with governments and NGOs to magnify the effect of the toolkits and integrate them within education policies and curricula in India. This study demonstrates how small, well-designed interventions can unlock the creative potential of children who might otherwise be excluded from such opportunities. A holistic education is important to train the youth to think differently and contribute to society in ways that help shape the nation.

Acknowledgments

I would like to thank my mentor, Ms. Aashna Saraf, for all of her guidance and support throughout this entire process. I would also like to thank New Era School for allowing me to pilot Brick-by-Brick with their students. I would also like to thank my parents and school, without whom none of this would have been possible.

References

1. Torrance, E. P. *Creativity as Manifest in Testing. The Journal of Creative Behavior* 1969, *3* (3), 137–154. ScienceDirect
2. American Psychological Association. *Problem Solving*. In the APA Dictionary of Psychology. https://dictionary.apa.org/problem-solving
3. American Psychological Association. *Socioeconomic Status (SES)*. In the *APA Dictionary of Psychology*. https://dictionary.apa.org/socioeconomic-status
4. Chaudhary, V.; Agrawal, A.; Sureka, A.; Sureka, P. *An Experimental Study on the Learning Outcome of Teaching Elementary Level Children using Lego Mindstorms EV3 Robotics Education Kit. arXiv preprint* 2016, arXiv:1610.09610. https://arxiv.org/abs/1610.09610
5. Swan, G. M.; The LEGO Group. *LEGO® Creativity: Building Skills for Life*. University of Kentucky. https://www.uky.edu/~gmswan3/544/LEGO_Creativity.pdf
6. LeGoff, D. B. *Use of LEGO® as a Therapeutic Medium for Improving Social Competence. Pastoral Care in Education* 2002, *20* (2), 27–35. https://journals.sagepub.com/doi/pdf/10.2304/plat.2002.2.2.87
7. Li, Y.; Huang, R.; Shi, Y. *The Effect on Pupils' Science Performance and Problem-Solving Ability through LEGO: An Engineering Design-Based Modeling Approach. Journal of Educational Technology & Society* 2016, *19* (3), 143–156.
8. Li, Y.; Huang, R. *The Effect on Pupils' Science Performance and Problem-Solving Ability through LEGO: An Engineering Design-Based Modeling Approach. Semantics Scholar* 2016. https://www.semanticscholar.org/paper/The-Effect-on-Pupils%27-Science-Performance-and-Lego%3A-Li-Huang/3efbfe62b30e9c3a12ec3e2c19a9e71d56f3b40b
9. Abbasi, M.; Hosseinzadeh, M.; Shafiee, M. *The Effect of LEGO Educational Robots on Students' Creativity and Learning Outcomes. Journal of Technology of Education (JTE)* 2020, *14* (1), 67–82. https://jte.sru.ac.ir/article_437_0279b00216f5dad805969f0ac39e2e52.pdf

Author

Vihaan Tannan is an enthusiastic 11th-grade student from Bombay International School, passionate about creativity and design, who has built over 300 LEGO sets. With a keen interest in problem-solving and innovation, they enjoy teaching and mentoring younger students, inspiring them to explore hands-on learning through building, teamwork, and imaginative thinking.

DOI: 10.36838/IJHSR812.15

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Adolescent Sleep Deprivation and Susceptibility to Later Development of Alzheimer's Disease

Mingwei Ma

Phillips Exeter Academy, 20 Main St, Exeter, NH, 03833, USA; melodyma6@icloud.com

ABSTRACT: Adolescence is a critical period for brain development, and sleep deprivation during this time may lay the groundwork for Alzheimer's disease (AD). AD is a neurodegenerative disorder characterized by memory loss and cognitive decline. Research has shown that sleep loss in older adults potentiates the risk of AD. Interestingly, approximately 80% of teenagers in the U.S. suffer from sleep deprivation. This narrative review paper aims to evaluate how the biological and structural changes induced during sleep deprivation in adolescence affect the susceptibility to AD in later adulthood. In recent years, literature reviews have focused on either the short-term effects of sleep loss during adolescence or the bidirectionality of sleep and AD in older adults. However, due to the difficulty of longitudinal studies across an individual's lifespan, there is a lack of research connecting the two fields. By bridging evidence from both fields, this paper proposes that persistent sleep deprivation during adolescence induces biological and structural changes that may increase the susceptibility to AD development later in life. This review advocates for this crucial gap in research as it may identify an early-life risk factor for AD and point to key public health and policy implications, such as later school start times.

KEYWORDS: Medical and Health Sciences, Neuroscience, Alzheimer's Disease, Sleep Deprivation, Adolescence.

Introduction

Adolescence is a critical period for brain development and growth.[1] The accumulation of sleep debt during this time may increase an individual's susceptibility to Alzheimer's disease (AD) in the future. Sleep deprivation occurs when an individual consistently gets fewer than the recommended hours of sleep for their age. For teenagers, the recommended amount is 8-10 hours of sleep per night, but self-report surveys indicate that approximately 80% of teenagers in the U.S. sleep less than the recommended amount and therefore suffer from sleep deprivation.[2,3] Similar results are found across the globe, in countries like Korea, Brazil, and Taiwan.[4] Contributing to these statistics are both sociological and biological factors. Social factors include academic stress, early school start times, and addictive technology, while biological factors include the circadian rhythm shift due to hormonal fluctuations during puberty.[5] Chronic sleep loss induces short and long-term effects on a teenager's mental health, physical well-being, and brain structure.[6] However, the long-term effects remain understudied. This is due to the difficulty of tracking brain changes over an extended period of time. This limitation is especially challenging during adolescence, when the brain is rapidly undergoing major structural changes and synaptic pruning of neurons.[7]

Emerging evidence from animal and neuroimaging studies shows that sleep loss potentiates the risk of neurodegenerative diseases in older adults. AD is a neurodegenerative disorder characterized by deteriorating memory and cognitive functioning. It is ranked seventh in the United States for leading causes of death among older adults. In addition, the worldwide cost of AD is estimated to reach $2 trillion USD in 2030.[8,9] Decreased sleep quality in older adults is a common aspect of aging. It results in decreased function of the glymphatic system, which is responsible for draining waste proteins from the brain.[10] This causes a build-up of amyloid-beta and tau proteins, increased neuroinflammation, and structural changes in brain regions vulnerable to AD. These consequences of sleep loss are also indicators or biomarkers for AD, reinforcing the idea that sleep loss is a direct risk factor for the pathogenesis of AD.[11]

By bridging two traditionally separate areas of research, this narrative review proposes that persistent sleep deprivation during the adolescent years induces biological and structural changes that may increase the susceptibility to developing Alzheimer's disease in later adulthood. In recent years, extensive studies have been done on either the short-term effects of sleep loss during teenage years or the bidirectionality of sleep and AD in older adults. However, due to the difficulty of longitudinal studies across an individual's lifespan, there is a lack of research bridging the two fields. This review works to generate hypothesized AD mechanistic pathways from adolescence to adulthood for future investigation.

As the number of sleep-deprived teenagers across the globe skyrockets, it becomes increasingly crucial to investigate the long-term neurological effects of chronic sleep loss in adolescence. Addressing this crucial gap in research could potentially identify an early-life risk factor for AD, which has significant public health implications for scientists, doctors, and policymakers. If the hypothesis is proven to be true, then improving sleep duration and quality in adolescents will not only provide immediate benefits like academic performance but may also reduce the risk of future neurodegeneration. In turn, this could lead to decreased death rates due to AD and reduced medical costs. Early implementation of preventive measures, such as later school start times and increased control over technology

DOI: 10.36838/IJHSR812.22

usage, can reduce the number of sleep-deprived adolescents before it is too late.

This paper is organized as follows. First, it discusses the methodology used for this narrative review. Next, it reviews previous research on the causes and effects of sleep deprivation in adolescents, emphasizing its impacts during this critical window of brain development. Then, it examines findings on sleep loss and its specific neurological effects on the brain. The third section provides evidence for the link between the neurological effects of sleep loss and the risk factors of AD in older individuals. The following section analyzes all of the previous evidence to generate hypothesized pathways for the progression from biomarkers induced during adolescence to AD in late adulthood. Finally, the discussion outlines possible early preventive measures for AD, solutions to improve adolescent sleep habits, and ideas for future areas of research.

Methodology

This paper is a narrative review of the research question: How might the biological and structural changes induced by chronic sleep loss during adolescence affect the susceptibility to AD in later adulthood? This paper solely focuses on the issue of sleep deprivation, rather than oversleeping or sleep disorders, due to its higher prevalence among adolescents worldwide. The specific biological and structural changes discussed in this paper refer to the glymphatic system, neuroinflammation, and changes in brain region volume and density. This paper investigates AD in particular because of the urgency to find a cure or prevention, as well as the recent discovery that sleep loss is a direct risk factor for AD.

Databases such as PubMed, Google Scholar, and Science Direct were used to find credible and recent sources. Other sources were obtained by searching through the reference lists of other review articles. This paper considers studies published between 2012 and 2025, with the exceptions of three studies from 1999, 2002, and 2008, as they represent foundational discoveries. Key terms included "adolescent sleep deprivation" and "Alzheimer's disease biomarkers". Key terms also included combinations of phrases from different fields, such as "glymphatic system", "neuroinflammation", "brain region changes", "Alzheimer's pathology", AND "sleep loss". The inclusion criteria were peer-reviewed studies of human or animal models examining the biological consequences of sleep loss or the biomarkers of AD, as well as review articles for a comprehensive background on the topic. Due to the limitations of testing directly on human participants, this paper includes animal studies and evaluates the validity of these studies based on their sample size, methodology, and relevance. The exclusion criteria were studies that focused on sleep disorders or acute sleep loss.

A narrative review approach was taken due to the interdisciplinary nature of this topic, as it bridges two traditionally separate areas of research together. This paper critically analyzes various past empirical research and synthesizes new hypotheses for future research.

Results

Causes and Consequences of Sleep Deprivation in Adolescents:

As the teenage brain undergoes its second major developmental phase, chronic sleep loss disrupts its growth and induces both short- and long-term changes in the individual.[12] A full understanding of the pathways of brain maturation and the impact of sleep debt during this critical period is needed for hypothesizing how early-life sleep loss can potentiate AD in later adulthood.

In recent years, the number of teenagers sleeping less than the recommended 8-10 hours per night is increasing at an alarming rate. This trend is driven by a combination of biological and sociological factors. Adolescence is typically defined as the stage in life between childhood and adulthood, ages 10 to 19. It is characterized by the hormonal fluctuations associated with puberty.[13] These hormonal fluctuations disrupt the circadian rhythm, delaying the natural sleep-wake cycle and making it harder for teenagers to sleep earlier. The circadian rhythm is regulated by a system of internal "clocks" that match up the body's sleep patterns to the 24-hour light-dark cycle.[14] Specifically, the suprachiasmatic nucleus in the hypothalamus receives light from the environment and coordinates bodily functions such as hormone production, temperature fluctuations, and metabolic processes to direct sleep activity.[15] This biological aspect of adolescence naturally reduces sleep duration in teenagers. On top of this innate delay in bedtime, adolescents are exposed to external pressures, such as increased academic stress, early school start times, strenuous after-school activities, and late-night athletic commitments. These pressures can directly cause shorter sleep duration for adolescents. In addition, habits before bedtime, such as high caffeine intake and exposure to technology (social media, video games, television), can disrupt sleep times and decrease sleep quality.[16] The sociological factors can amplify the circadian rhythm shift, increasing the amount of sleep debt per night.

Consequences of chronic sleep loss in adolescents can appear as short-term effects or as long-term effects, summarized in Table 1. Short-term effects of sleep loss are commonly studied through self-report surveys, hormonal assays, and neurocognitive tests. They include increased stress levels, headaches, stomach pain, decreased memory and focus, worsened mental health, decreased cognitive abilities, and mood swings.[17] These consequences can be shown through poor academic performance and burnout.[18] This may lead to even less sleep per night to catch up on school work, forming a bidirectionality between the causes and effects of sleep loss and creating a cycle that leads to chronic sleep deprivation.[18] The long-term consequences of sleep debt are poorly understood, due to the difficulty of longitudinal studies tracking the teenage brain into adulthood. However, growing evidence suggests that chronic sleep loss can lead to obesity, hypertension, mental health disorders, and cognitive decline.[19] Although these long-term effects are unpredictable due to the multi-factorial nature of brain development, it is imperative that we understand them fully.

Table 1: Short- and long-term effects of chronic sleep loss in adolescents.

Table 1. Effects of Chronic Sleep Loss in Adolescents	
Short-term	**Long-term**
Increased stress levels	Obesity
Headaches	Hypertension
Stomach pain	Mental health disorders
Decreased memory and focus	Cognitive decline
Worsened mental health	
Decreased cognitive abilities	
Mood swings	

Adolescence marks the period where the brain undergoes its second major developmental phase.[20] Human longitudinal studies reveal that the brain reorganizes itself during this time by performing key processes.[21] To increase the efficiency of neural processing and communication between brain regions, the brain undergoes synaptic pruning and myelination. Synaptic pruning is the process by which the brain filters and removes unused or weak synaptic connections, and myelination is the process of insulating axons with myelin sheaths to increase the speed of signal conduction.[7,22] The brain also undergoes structural changes. The prefrontal cortex–the part of the brain responsible for executive function, such as impulse control and risk assessment–rapidly develops throughout adolescence.[23] The limbic system, which controls emotions and reward processing, also becomes heightened in reactivity. This causes adolescents to become emotionally sensitive to social and reward stimuli.[24] Lastly, it is crucial to note that the teenage brain is highly plastic, meaning it is very sensitive to environmental factors such as sleep. Accumulation of sleep debt could take advantage of the brain's plasticity during this time and alter the brain's development, leaving long-lasting consequences that carry into adulthood.

This evidence indicates that sleep deprivation due to biological and sociological factors during adolescence can have immediate effects on physical states, mental health, and cognitive function, although the long-term consequences are less understood. The adolescent brain is particularly vulnerable to damage from sleep loss, as it is undergoing major structural and functional changes during this time period. Future research continuing to investigate the long-term effects is crucial for implementing public health policies. In addition, more insight into how exactly the brain changes during its development phase in response to sleep loss will allow us to understand the possible connections to neurodegenerative diseases in the future. The following section will begin to explore this potential link by analyzing the specific neurological effects of sleep deprivation on a molecular and structural level.

Neurological Effects Due to Sleep Loss:

In all individuals, sleep plays a role in critical processes such as memory consolidation, synaptic homeostasis, cognition, sensory processing, critical thinking, and neural restoration.[25] Sleep deprivation disrupts these processes, resulting in significant consequences for the brain and body. This section examines the molecular mechanisms underlying key functions of sleep and how they are impaired by sleep loss.

First, the brain drains waste proteins during sleep in order to maintain tissue and synaptic homeostasis. The glymphatic system is the waste clearance system for excess cerebrospinal fluid (CSF) and other solutes produced by neurons.[26] This system is most active during the third stage of non-rapid eye movement sleep (NREM3) and works to prevent a build-up of waste proteins in the brain by removing them through a network of channels formed by astrocytes.[27] As shown in Figure 2, CSF enters the perivascular space and gets transported by aquaporins to the astrocyte network. As the CSF flows through the astrocytes, it filters out the metabolic waste. The excess CSF and metabolic waste are transported through aquaporins into the perivascular space, where they are excreted from the brain.[28] Sleep loss or poor sleep quality impairs the glymphatic system's function, allowing waste proteins to accumulate and form aggregates such as Tau tangles or amyloid-beta plaques.[29] As waste build-up increases in the brain, the drainage channels become clogged, causing decreased CSF flow in the glymphatic system, creating a toxic cycle (Figure 3).[10] This glymphatic system dysfunction is most commonly seen in older adults, who experience reduced NREM3 sleep due to aging.

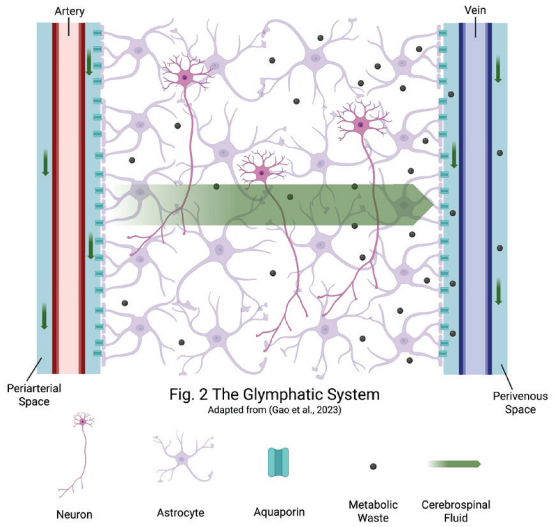


Figure 2: The glymphatic system removes excess CSF fluid and metabolic waste to maintain tissue and synaptic homeostasis.

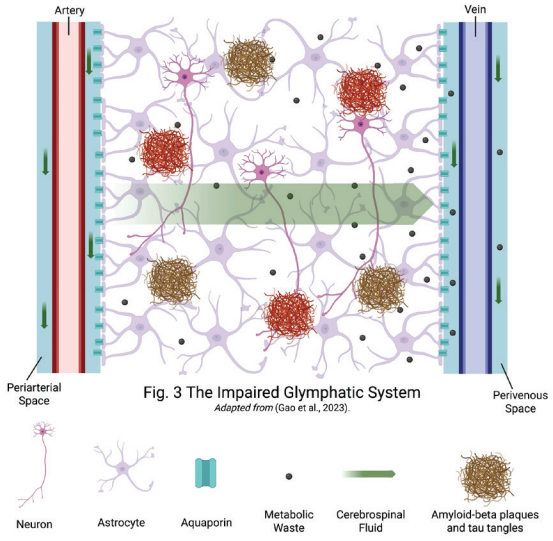


Figure 3: Accumulation of waste proteins leads to an impaired glymphatic system with reduced CSF flow and drainage capabilities.

DOI: 10.36838/IJHSR812.22

Second, the brain's immune system shifts from defensive to maintenance mode during sleep. The brain's immune system–composed mostly of microglia, leukocytes, T-cells, and cytokines–reduces its inflammatory activity and instead focuses on synaptic pruning and memory storage of previous antigens.[30] Although pro-inflammatory cytokines are still present to protect the body from antigens during sleep, the brain also promotes anti-inflammatory cytokines to maintain a balanced neural environment.[30] This mechanism helps to prevent chronic neuroinflammation. Persistent sleep deprivation triggers the production of pro-inflammatory cytokines in the brain, disrupting this balance and causing neuroinflammation.[31] Neuroinflammation–an inflammatory response in the brain–may lead to irreversible consequences such as demyelination of axons, neural tissue damage, disruption of synaptic connectivity, and cognitive decline.[32]

Third, sleep plays a critical role in shaping brain structure through processes like consolidating memory, dendritic spine remodeling, and synaptic pruning. During sleep, the brain has the optimal conditions for memory consolidation, the process of encoding short-term memory into long-term memory.[33] The core of this concept of memory formation is the strength of the synaptic connections relating to this memory, also known as long-term potentiation (LTP).[33] LTP, a form of synaptic plasticity, works in tandem with synaptic pruning, as the latter removes weak neural connections and strengthens strong ones. The loss of sleep impairs the strengthening of these neuronal connections, therefore leading to decreased cognition, memory, and critical thinking.

These findings demonstrate that sleep deprivation impairs numerous neurological processes that occur during sleep, such as waste clearance, immune system maintenance, synaptic pruning, and structural development. Disruption to these essential restorative functions in the brain leads to lasting consequences. Additionally, while current research mostly explores the consequences of insufficient sleep, it would be beneficial for future studies to investigate the consequences of excess sleep or "catch-up" sleep on weekends, especially in teenagers. To answer the research question, the next section will analyze evidence connecting these neurological effects of sleep deprivation to the biomarkers and hallmarks of AD.

Evidence Linking Sleep Deprivation to Alzheimer's Disease:

The overlap between the specific neurological changes discussed above and the pathology of AD suggests that sleep deprivation may be a direct risk factor. Both animal and human studies indicate that the sleep-deprived brain shares key characteristics with brains suffering from AD. These characteristics include accumulation of waste protein aggregates, increased neuroinflammation, and structural changes in vulnerable brain regions. This section analyzes AD pathology and emerging evidence linking it to sleep deprivation. It is worth noting that these studies are focused on the matured adult brain, not the adolescent one. Translating the mechanisms of this link–between sleep loss and adult brains–to adolescent brains is only a speculation and not directly observed.

AD is a neurodegenerative disease that targets the brain's neural connections and results in memory loss. The basis of this cognitive decline involves the decline in activity of the glymphatic system, leading to the accumulation of waste proteins, which induce inflammatory responses and ultimately result in disruption of synaptic activity.[34] AD can be diagnosed by the detection of amyloid-beta and tau protein clumps in the brain through magnetic resonance imaging (MRI) or positron emission tomography (PET).[35] Interestingly, these hallmarks of AD–amyloid-beta aggregates and tau tangles–also appear in sleep-deprived brains, such as older individuals who struggle with entering NREM3 sleep or those who suffer from sleep disorders like insomnia.[36] Researchers investigating the similarity of sleep-deprived brains and AD-affected brains showed that sleep loss acts as a direct risk factor for AD. A 2018 PET study of human participants reveals that after one night of sleep deprivation, the amyloid-beta levels in the right hippocampus and thalamus significantly exceeded the baseline levels of the control group.[37] Another study uses self-reported sleep lengths from human participants and MRI/PET brain scans to analyze amyloid-beta deposition, finding that shorter sleep duration results in greater amyloid-beta burden.[38] In addition, an animal study testing the effects of orexin levels on mice found that sleep deprivation leads to an increased amount of amyloid-beta.[39] This evidence reveals how sleep loss, even just one night, causes the build-up of waste proteins, harming neuron health and synaptic communication.

The accumulation of waste proteins, like amyloid-beta, induces neuroinflammation. Receptor protein LTR2 triggers neurotoxic pro-inflammatory mediators when amyloid-beta binds to it.[40] Neuroinflammation contributes to AD progression through a positive feedback cycle. It increases oxidative stress, degrades neurons, and loses the ability to destroy neurotoxic proteins like amyloid-beta and tau, which in turn triggers more inflammation.[41] Inflammation levels can be measured by miRNA, snippets of non-coding RNA that can act as biomarkers.[42] A human study tracking circulating miRNAs in the blood of participants the morning after a night of no sleep found that two out of the three miRNAs associated with neurodegeneration in AD increased compared to the baseline values.[43] Specifically, this group of researchers found that the levels of miR-127-3p and miR-142-3p, both of which activate inflammation, were greater after a night of no sleep versus a night of sleep. These results further corroborate the idea that sleep loss directly causes inflammation, which in turn causes neurodegeneration.

Taken together, emerging evidence from both human and animal studies indicates that sleep loss is a direct risk factor for AD, as its consequences overlap with the pathology and hallmarks of AD. However, while the neuroscience community is starting to understand the molecular mechanisms of waste protein accumulation and neuroinflammation, little is known about these processes in the sleep-deprived adolescent brain. This research gap may exist due to the recency of the consensus that sleep loss is a direct risk factor for AD, meaning longitudinal studies tracking brains from adolescence to late adulthood have not yet been completed. Due to the lack of direct evidence

connecting sleep deprivation in adolescence to the risk of AD in later adulthood, the following section proposes hypothetical pathways based on current understandings of AD pathology for possible future investigation.

Hypothesized AD Mechanistic Pathways from Adolescence to Adulthood:

Using known evidence of the pathology of AD in adults and the consequences of sleep deprivation, this section works to create extrapolated hypotheses for the possible mechanistic pathways of AD from adolescence to late adulthood.

One pathway involves the impaired glymphatic system. The effectiveness of the glymphatic system is a major concern when considering the long-term neurological consequences of sleep loss in adolescence. As the circadian rhythm shifts back and teenagers fall to the temptation of technology before bedtime, both sleep quantity and quality are compromised. Because the glymphatic system is most active during NREM3 sleep, the fragmented or decreased durations of NREM3 sleep could limit its ability to remove CSF and waste proteins. These proteins, such as amyloid-beta and tau, can cause a build-up in the neurons, which impairs synaptic connections. Persistent sleep deprivation of consecutive nights of sleep loss may cause substantial waste protein accumulation. The glymphatic system may attempt to compensate and work harder to remove these waste proteins during the next NREM3 sleep, but this may backfire and result in a damaged glymphatic system from overactivation. The teenage brain is undergoing rapid brain development, which leaves it very vulnerable to damage during this adolescent period. This impairment and overactivation of the glymphatic system may persist into adulthood. When aging naturally causes an increase in the build-up of neurotoxic proteins, the damaged glymphatic system may not be as effective in removing waste, leading to a faster accumulation of tau or amyloid-beta. This early-life build-up of amyloid-beta and tau aggregates may predispose the brain to faster accumulation of neurotoxic proteins in later adulthood by wearing down the glymphatic system.

A second mechanism is chronic neuroinflammation. The persistent activation of the immune system is a long-term consequence of sleep deprivation in adolescence that may carry into adulthood. A night of sleep loss increases neuroinflammation and results in increased oxidative stress, which impairs neuronal health and synaptic plasticity.[44] Persistent sleep deprivation–common in teenagers due to bad sleep habits and early school start times–causes recurring inflammation in the brain. This chronic neuroinflammation, where microglia cells activate strong immune responses, can damage healthy neurons nearby, including the blood-brain barrier (BBB).[45] The BBB is a selectively permeable membrane between the blood and the interstitium of the brain and acts as a line of defense against foreign antigens.[46,47] During adolescence, the BBB undergoes alterations in its permeability, suggesting that early damage to the barrier in adolescence may decrease the barrier's resilience against antigens in the future.[48] In addition, astrocytes and microglia cells undergo major development during adolescence.48 In late adulthood, the cells may remember the mechanisms for inflammation from previous episodes, resulting in a faster and more aggressive inflammation response when triggered by age-related sleep loss. Due to the brain and immune system's high plasticity during adolescence, chronic inflammation may result in decreased protection against antigens and increased responsiveness or overactivation of immune responses. These consequences may follow into late adulthood, leading to a faster rate of neurodegeneration.

The third pathway involves structural changes in the brain. During periods of sleep loss, brain regions associated with learning, memory, and executive functioning show reductions in volume and activity. Adolescence is when the brain undergoes major growth in key brain regions such as the hippocampus and prefrontal cortex. The hippocampus is responsible for memory consolidation, and the prefrontal cortex is responsible for decision-making and cognition.[49] Sleep deprivation during this stage impacts the growth of these brain regions by reducing grey matter volume, synaptic density, and dendritic spines. As the adolescent brain continues to develop with these limitations, the result may be an adult brain with lower baseline levels of cognition. Human neuroimaging studies have shown that individuals suffering from AD have decreased functioning and volume in the hippocampus and prefrontal cortex, both of which are regions associated with cognition and memory.[50] This early-life underdevelopment of the cognitive functioning regions in the brain may increase the vulnerability to neurodegeneration in later adulthood.

In sum, these three potential pathways, hypothesized from the evidence surrounding glymphatic system function, neuroinflammation, and brain structural changes, suggest that sleep deprivation in adolescence may predispose an individual to AD when older. Future research should acquire more longitudinal data, such as surveys or brain scans about sleep duration and neurodegeneration across lifespans, in order to confirm or deny these hypotheses. With all of this extensive literature review in mind, the paper will discuss the importance of this field and potential research directions.

■ Discussion

This review uses converging evidence from the consequences of sleep loss and pathological characteristics of AD to outline three potential pathways of AD progression: impaired glymphatic system, chronic neuroinflammation, and reduced development of brain regions vulnerable to neurodegeneration. Although this paper analyzes them separately, it's important to note that these three pathways are interconnected in many ways. For example, an impaired glymphatic system leads to the build-up of amyloid-beta and tau proteins, which activate neuroinflammation; inflammation, in turn, hinders cell communication and further reduces glymphatic efficiency.[51] In addition, both the accumulation of waste proteins and inflammation can disrupt brain region development and reduce synaptic density.[51] These three main mechanisms together create a self-perpetuating cycle. On the other hand, human and mouse studies have shown that the accumulation of amyloid-beta proteins results in shortened and disturbed sleep.[52]

DOI: 10.36838/IJHSR812.22

This bidirectional relationship between sleep and AD forms a continuous loop that leads to chronic sleep deprivation.

More attention should be directed to this critical gap in research. Existing studies have established the bidirectional relationship between sleep and AD, yet most of them focus on the aged adult brain, not the adolescent one. The lack of studies investigating the link between sleep loss in teenagers and AD in later life can be attributed to several limitations and reasons. First, there are logistical difficulties in performing longitudinal studies across decades. It is a challenge to find willing participants, ensure these participants are able to perform periodical check-ups, and finish the decades-long experiment until completion. Second, the period of adolescence is a time of big developmental changes in the brain. These neurological changes are unpredictable and different for everybody, making investigating the effects of sleep loss during adolescence hard and inaccurate. Third, the connection between sleep deprivation in adolescence and neurodegenerative diseases may seem weak. Big lifestyle changes can happen between adolescence and late adulthood, such as diet, physical activity, and mental health. These factors may affect the brain more than sleep and complicate the results. It is difficult to determine causality when there are so many factors at play. However, as technology continues to improve, new research methods or laboratory equipment may allow researchers to overcome these limitations. By demonstrating the possibility of chronic sleep loss in adolescence being an early-life risk factor for AD, this paper hopes to invite future research on this topic.

From a public health standpoint, finding an early-life risk factor for AD is critical, as it allows for early preventive measures. Sleep loss in adolescence is a modifiable risk factor, meaning it can be altered through individual and systemic interventions, unlike other risk factors for AD, like genetic predispositions. Public policies such as later school start times and increased control on technology usage can help reduce the number of sleep-deprived adolescents. If the hypothesis is proven to be true, improving sleep duration and quality in adolescents will not only provide immediate benefits like academic performance and mental health, but may also reduce the risk of future neurodegeneration. This leads to reduced medical costs and deaths due to AD, which will increase the quality of life for everyone.

Future research should also consider the role of other factors. For example, certain genes in individuals are found to increase the risk of AD.[53] The role of genetic predispositions to AD in adolescents during sleep deprivation is a promising area for future research. In addition, the long-term effects of "make-up sleep" in teenagers, such as naps or excess sleep on weekends, are also worth investigating. Though this paper focuses on sleep deprivation due to its prevalence among teenagers, the consequences of oversleeping are another concern and can be a future area of research. Lastly, research can investigate how chronic sleep loss during adolescence may increase the susceptibility to other brain disorders or diseases. Examples include Parkinson's disease, multiple sclerosis, stroke, and epilepsy. By bridging two traditionally separate fields of study, this paper opens the possibilities for a wide variety of future research.

■ Conclusion

This review proposes that persistent sleep deprivation during the adolescent years induces biological and structural changes that may increase the risk of developing Alzheimer's disease later in life. By analyzing evidence on the causes and consequences of sleep loss during adolescence, neurological effects on the brain as a result of sleep deprivation, and the role of sleep loss as a direct risk factor of AD, this paper outlines hypothesized mechanistic pathways linking early-life sleep debt to neurodegeneration in later adulthood. The hypothesized pathways are the overactivation of the glymphatic system, chronic neuroinflammation, and reductions in key brain regions associated with cognitive functions. Future research testing the amount of waste protein build-up, neuroinflammation, and structural changes in the adolescent brain after various periods of insufficient sleep, as well as longitudinal data are needed to test these hypotheses. Other variables, such as genetic predisposition to AD and oversleeping, need to be considered as well. By addressing this major gap in research, this paper highlights the potential for identifying an early-life risk factor for AD and its public health policy implications. If the hypothesis is proven to be true, later school start times and increased control over technology could increase the amount of sleep a teenager gets per night and help decrease the susceptibility to AD in later years. Reduced neurodegeneration in the elderly population can reduce medical costs and deaths, and lead to a better quality of life for all involved. Given the sharp increase in the number of sleep-deprived teenagers who will all grow into the elderly population in several decades, public health protocols must be implemented early to prevent neurodegeneration before it is too late.

■ Acknowledgments

This work was supported by Dr. Jorge Avila and Andrew Huckins-Noss from Indigo Research. I attest that the ideas, graphics, and writing in this paper are entirely my own.

■ References

1. Larsen, B.; Luna, B. Adolescence as a Neurobiological Critical Period for the Development of Higher-Order Cognition. *Neurosci. Biobehav. Rev.* **2018**, *94*, 179–195. https://doi.org/10.1016/j.neubiorev.2018.09.005.
2. American Academy of Sleep Medicine. *Teen Sleep Duration Health Advisory - AASM Recommendation*. American Academy of Sleep Medicine – Association for Sleep Clinicians and Researchers. https://aasm.org/advocacy/position-statements/teen-sleep-duration-health-advisory/ (accessed 2025-11-27).
3. National Sleep Foundation. *Teens' Sleep Health and Mental Health Strongly Linked*. National Sleep Foundation. https://www.thensf.org/teens-sleep-health-and-mental-health-strongly-linked/ (accessed 2025-07-26).
4. Leger, D.; Beck, F.; Richard, J.-B.; Godeau, E. Total Sleep Time Severely Drops during Adolescence. *PLOS ONE* **2012**, *7* (10), e45204. https://doi.org/10.1371/journal.pone.0045204.
5. Owens, J. Insufficient Sleep in Adolescents and Young Adults: An Update on Causes and Consequences. *Pediatrics* **2014**, *134* (3), e921–e932. https://doi.org/10.1542/peds.2014-1696.

6. Johri, K.; Pillai, R.; Kulkarni, A.; Balkrishnan, R. Effects of Sleep Deprivation on the Mental Health of Adolescents: A Systematic Review. *Sleep Sci. Pract*. **2025**, *9* (1), 9. https://doi.org/10.1186/s41606-025-00127-w.
7. Cornell, J.; Salinas, S.; Huang, H.-Y.; Zhou, M. Microglia Regulation of Synaptic Plasticity and Learning and Memory. *Neural Regen. Res*. **2021**, *17* (4), 705–716. https://doi.org/10.4103/1673-5374.322423.
8. *Alzheimer's Disease Fact Sheet*. National Institute on Aging. https://www.nia.nih.gov/health/alzheimers-and-dementia/alzheimers-disease-fact-sheet (accessed 2025-07-26).
9. Tay, L. X.; Ong, S. C.; Tay, L. J. *Economic Burden of Alzheimer's Disease: A Systematic Review - Value in Health Regional Issues*. https://www.valuehealthregionalissues.com/article/S2212-1099(23)00094-8/fulltext (accessed 2025-11-27).
10. Nedergaard, M.; Goldman, S. A. Glymphatic Failure as a Final Common Pathway to Dementia. *Science* **2020**. https://doi.org/10.1126/science.abb8739.
11. Sadeghmousavi, S.; Eskian, M.; Rahmani, F.; Rezaei, N. The Effect of Insomnia on the Development of Alzheimer's Disease. *J. Neuroinflammation* **2020**, *17* (1), 289. https://doi.org/10.1186/s12974-020-01960-9.
12. Anastasiades, P. G.; de Vivo, L.; Bellesi, M.; Jones, M. W. Adolescent Sleep and the Foundations of Prefrontal Cortical Development and Dysfunction. *Prog. Neurobiol*. **2022**, *218*, 102338. https://doi.org/10.1016/j.pneurobio.2022.102338.
13. National Academies of Sciences, E.; Division, H. and M.; Education, D. of B. and S. S. and; Board on Children, Y.; Applications, C. on the N. and S. S. of A. D. and I.; Backes, E. P.; Bonnie, R. J. Adolescent Development. In *The Promise of Adolescence: Realizing Opportunity for All Youth*; National Academies Press (US), **2019**.
14. Uccella, S.; Cordani, R.; Salfi, F.; Gorgoni, M.; Scarpelli, S.; Gemignani, A.; Geoffroy, P. A.; De Gennaro, L.; Palagini, L.; Ferrara, M.; Nobili, L. Sleep Deprivation and Insomnia in Adolescence: Implications for Mental Health. *Brain Sci*. **2023**, *13* (4), 569. https://doi.org/10.3390/brainsci13040569.
15. Coomans, C. P.; Ramkisoensing, A.; Meijer, J. H. The Suprachiasmatic Nuclei as a Seasonal Clock. *Front. Neuroendocrinol*. **2015**, *37*, 29–42. https://doi.org/10.1016/j.yfrne.2014.11.002.
16. Owens, J. Insufficient Sleep in Adolescents and Young Adults: An Update on Causes and Consequences. *Pediatrics* **2014**, *134* (3), e921–e932. https://doi.org/10.1542/peds.2014-1696.
17. Medic, G.; Wille, M.; Hemels, M. E. Short- and Long-Term Health Consequences of Sleep Disruption. *Nat. Sci. Sleep* **2017**, *9*, 151–161. https://doi.org/10.2147/NSS.S134864.
18. Evers, K.; Chen, S.; Rothmann, S.; Dhir, A.; Pallesen, S. Investigating the Relation among Disturbed Sleep Due to Social Media Use, School Burnout, and Academic Performance. *J. Adolesc*. **2020**, *84* (1), 156–164. https://doi.org/10.1016/j.adolescence.2020.08.011.
19. Johri, K.; Pillai, R.; Kulkarni, A.; Balkrishnan, R. Effects of Sleep Deprivation on the Mental Health of Adolescents: A Systematic Review. *Sleep Sci. Pract*. **2025**, *9* (1), 9. https://doi.org/10.1186/s41606-025-00127-w.
20. Sisk, C. L.; Foster, D. L. The Neural Basis of Puberty and Adolescence. *Nat. Neurosci*. **2004**, *7* (10), 1040–1047. https://doi.org/10.1038/nn1326.
21. Giedd, J. N.; Blumenthal, J.; Jeffries, N. O.; Castellanos, F. X.; Liu, H.; Zijdenbos, A.; Paus, T.; Evans, A. C.; Rapoport, J. L. Brain Development during Childhood and Adolescence: A Longitudinal MRI Study. *Nat. Neurosci*. **1999**, *2* (10), 861–863. https://doi.org/10.1038/13158.
22. Corrigan, N. M.; Yarnykh, V. L.; Hippe, D. S.; Owen, J. P.; Huber, E.; Zhao, T. C.; Kuhl, P. K. Myelin Development in Cerebral Gray and White Matter during Adolescence and Late Childhood. *NeuroImage* **2021**, *227*, 117678. https://doi.org/10.1016/j.neuroimage.2020.117678.
23. Arain, M.; Haque, M.; Johal, L.; Mathur, P.; Nel, W.; Rais, A.; Sandhu, R.; Sharma, S. Maturation of the Adolescent Brain. *Neuropsychiatr. Dis. Treat*. **2013**, *9*, 449–461. https://doi.org/10.2147/NDT.S39776.
24. Casey, B. J.; Jones, R. M.; Hare, T. A. The Adolescent Brain. *Ann. N. Y. Acad. Sci*. **2008**, *1124*, 111–126. https://doi.org/10.1196/annals.1440.010.
25. Zielinski, M. R.; McKenna, J. T.; McCarley, R. W.; Zielinski, M. R.; McKenna, J. T.; McCarley, R. W. Functions and Mechanisms of Sleep. *AIMS Neurosci*. **2016**, *3* (1), 67–104. https://doi.org/10.3934/Neuroscience.2016.1.67.
26. Iliff, J. J.; Wang, M.; Liao, Y.; Plogg, B. A.; Peng, W.; Gundersen, G. A.; Benveniste, H.; Vates, G. E.; Deane, R.; Goldman, S. A.; Nagelhus, E. A.; Nedergaard, M. A Paravascular Pathway Facilitates CSF Flow through the Brain Parenchyma and the Clearance of Interstitial Solutes, Including Amyloid β. Sci. *Transl. Med*. **2012**, *4* (147), 147ra111. https://doi.org/10.1126/scitranslmed.3003748.
27. Jessen, N. A.; Munk, A. S. F.; Lundgaard, I.; Nedergaard, M. The Glymphatic System – A Beginner's Guide. *Neurochem. Res*. **2015**, *40* (12), 2583–2599. https://doi.org/10.1007/s11064-015-1581-6.
28. Gao, Y.; Liu, K.; Zhu, J. Glymphatic System: An Emerging Therapeutic Approach for Neurological Disorders. *Front. Mol. Neurosci*. **2023**, *16*. https://doi.org/10.3389/fnmol.2023.1138769.
29. Voumvourakis, K. I.; Sideri, E.; Papadimitropoulos, G. N.; Tsantzali, I.; Hewlett, P.; Kitsos, D.; Stefanou, M.; Bonakis, A.; Giannopoulos, S.; Tsivgoulis, G.; Paraskevas, G. P. The Dynamic Relationship between the Glymphatic System, Aging, Memory, and Sleep. *Biomedicines* **2023**, *11* (8), 2092. https://doi.org/10.3390/biomedicines11082092.
30. Besedovsky, L.; Lange, T.; Born, J. Sleep and Immune Function. *Pflüg. Arch. - Eur. J. Physiol*. **2012**, *463* (1), 121–137. https://doi.org/10.1007/s00424-011-1044-0.
31. Asif, N.; Iqbal, R.; Nazir, C. F. Human Immune System during Sleep. *Am. J. Clin. Exp. Immunol*. 2017, *6* (6), 92–96.
32. DiSabato, D. J.; Quan, N.; Godbout, J. P. Neuroinflammation: The Devil Is in the Details. **2016**. https://doi.org/10.1111/jnc.13607.
33. Rasch, B.; Born, J. About Sleep's Role in Memory. *Physiol. Rev*. 2013, *93* (2), 681–766. https://doi.org/10.1152/physrev.00032.2012.
34. Querfurth, H. W.; LaFerla, F. M. Alzheimer's Disease. The New England Journal of Medicine. https://doi.org/10.1056/NEJMra0909142.
35. Scheltens, P.; Blennow, K.; Breteler, M. M. B.; Strooper, B. de; Frisoni, G. B.; Salloway, S.; Flier, W. M. V. der. Alzheimer's Disease. *The Lancet* **2016**, *388* (10043), 505–517. https://doi.org/10.1016/S0140-6736(15)01124-1.
36. Xia, X.; Jiang, Q.; McDermott, J.; Han, J. J. Aging and Alzheimer's Disease: Comparison and Associations from Molecular to System Level. *Aging Cell* **2018**, *17* (5), e12802. https://doi.org/10.1111/acel.12802.
37. Shokri-Kojori, E.; Wang, G.-J.; Wiers, C. E.; Demiral, S. B.; Guo, M.; Kim, S. W.; Lindgren, E.; Ramirez, V.; Zehra, A.; Freeman, C.; Miller, G.; Manza, P.; Srivastava, T.; De Santi, S.; Tomasi, D.; Benveniste, H.; Volkow, N. D. β-Amyloid Accumulation in the Human Brain after One Night of Sleep Deprivation. *Proc. Natl. Acad. Sci. U. S. A*. **2018**, *115* (17), 4483–4488. https://doi.org/10.1073/pnas.1721694115.
38. Spira, A. P.; Gamaldo, A. A.; An, Y.; Wu, M. N.; Simonsick, E. M.; Bilgel, M.; Zhou, Y.; Wong, D. F.; Ferrucci, L.; Resnick, S. M. Self-Reported Sleep and β-Amyloid Deposition in Communi-

DOI: 10.36838/IJHSR812.22

ty-Dwelling Older Adults. *JAMA Neurol.* **2013**, *70* (12), 1537–1543. https://doi.org/10.1001/jamaneurol.2013.4258.
39. Roh, J. H.; Jiang, H.; Finn, M. B.; Stewart, F. R.; Mahan, T. E.; Cirrito, J. R.; Heda, A.; Snider, B. J.; Li, M.; Yanagisawa, M.; de Lecea, L.; Holtzman, D. M. Potential Role of Orexin and Sleep Modulation in the Pathogenesis of Alzheimer's Disease. *J. Exp. Med.* **2014**, *211* (13), 2487–2496. https://doi.org/10.1084/jem.20141788.
40. Liu, S.; Liu, Y.; Hao, W.; Wolf, L.; Kiliaan, A. J.; Penke, B.; Rübe, C. E.; Walter, J.; Heneka, M. T.; Hartmann, T.; Menger, M. D.; Fassbender, K. TLR2 Is a Primary Receptor for Alzheimer's Amyloid β Peptide to Trigger Neuroinflammatory Activation. *J. Immunol. Baltim. Md 1950* **2012**, *188* (3), 1098–1107. https://doi.org/10.4049/jimmunol.1101121.
41. Zhang, W.; Xiao, D.; Mao, Q.; Xia, H. Role of Neuroinflammation in Neurodegeneration Development. *Signal Transduct. Target. Ther.* **2023**, *8* (1), 267. https://doi.org/10.1038/s41392-023-01486-5.
42. van den Berg, M. M. J.; Krauskopf, J.; Ramaekers, J. G.; Kleinjans, J. C. S.; Prickaerts, J.; Briedé, J. J. Circulating microRNAs as Potential Biomarkers for Psychiatric and Neurodegenerative Disorders. *Prog. Neurobiol.* **2020**, *185*, 101732. https://doi.org/10.1016/j.pneurobio.2019.101732.
43. Zhang, L.; Grip, A.; Hjelmqvist, D.; Benedict, C.; Brandão, L. E. M.; Cedernaes, J. Acute Sleep Loss Increases Circulating Morning Levels of Two MicroRNAs Implicated in Neurodegenerative Disease in Healthy Young Men. *J. Cell. Mol. Med.* **2025**, *29* (7), e70523. https://doi.org/10.1111/jcmm.70523.
44. Lv, Y.-N.; Cui, Y.; Zhang, B.; Huang, S.-M. Sleep Deficiency Promotes Alzheimer's Disease Development and Progression. *Front. Neurol.* **2022**, *13*. https://doi.org/10.3389/fneur.2022.1053942.
45. da Fonseca, A. C. C.; Matias, D.; Garcia, C.; Amaral, R.; Geraldo, L. H.; Freitas, C.; Lima, F. R. S. The Impact of Microglial Activation on Blood-Brain Barrier in Brain Diseases. *Front. Cell. Neurosci.* **2014**, *8*, 362. https://doi.org/10.3389/fncel.2014.00362.
46. Kadry, H.; Noorani, B.; Cucullo, L. A Blood–Brain Barrier Overview on Structure, Function, Impairment, and Biomarkers of Integrity. *Fluids Barriers CNS* **2020**, *17* (1), 69. https://doi.org/10.1186/s12987-020-00230-3.
47. Dotiwala, A. K.; McCausland, C.; Samra, N. S. Anatomy, Head and Neck: Blood Brain Barrier. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), **2025**.
48. Brenhouse, H. C.; Schwarz, J. M. Immunoadolescence: Neuroimmune Development and Adolescent Behavior. *Neurosci. Biobehav. Rev.* **2016**, *70*, 288–299. https://doi.org/10.1016/j.neubiorev.2016.05.035.
49. Preston, A. R.; Eichenbaum, H. Interplay of Hippocampus and Prefrontal Cortex in Memory. *Curr. Biol. CB* **2013**, *23* (17), R764–R773. https://doi.org/10.1016/j.cub.2013.05.041.
50. Rao, Y. L.; Ganaraja, B.; Murlimanju, B. V.; Joy, T.; Krishnamurthy, A.; Agrawal, A. Hippocampus and Its Involvement in Alzheimer's Disease: A Review. *3 Biotech* **2022**, *12* (2), 55. https://doi.org/10.1007/s13205-022-03123-4.
51. Cai, Y.; Zhang, Y.; Leng, S.; Ma, Y.; Jiang, Q.; Wen, Q.; Ju, S.; Hu, J. The Relationship between Inflammation, Impaired Glymphatic System, and Neurodegenerative Disorders: A Vicious Cycle. *Neurobiol. Dis.* **2024**, *192*, 106426. https://doi.org/10.1016/j.nbd.2024.106426.
52. Wang, C.; Holtzman, D. M. Bidirectional Relationship between Sleep and Alzheimer's Disease: Role of Amyloid, Tau, and Other Factors. *Neuropsychopharmacology* **2020**, *45* (1), 104–120. https://doi.org/10.1038/s41386-019-0478-5.
53. Krishnamurthy, H. K.; Jayaraman, V.; Krishna, K.; Wang, T.; Bei, K.; Changalath, C.; Rajasekaran, J. J. An Overview of the Genes and Biomarkers in Alzheimer's Disease. *Ageing Res. Rev.* **2025**, *104*, 102599. https://doi.org/10.1016/j.arr.2024.102599.

Author

Mingwei Ma is a high school junior at Phillips Exeter Academy in New Hampshire interested in the intersection of neuroscience, public health, and biomedical engineering.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

The Impact of E-cigarette Use on Seizure Frequency in Vapers with and Without a History of Epilepsy

Sarayu Gadam

Braintree High School, 128 Town St, Braintree, MA 02184, USA, sarayu.gadam@gmail.com
Mentor: Professor Virgel Torremocha, Dr. Kathryn Wilwohl

ABSTRACT: E-cigarettes/electronic nicotine delivery systems (ENDS), though initially introduced as possible smoking cessation devices for preexisting smokers, have rapidly increased in use over the last few years, attracting youth with aggressive market tactics. Vaping is becoming an increasingly prevalent phenomenon, particularly among younger populations. Though cigarette use has somewhat declined in the United States in the last few years because of awareness and education campaigns, the teen vaping crisis has become more burdensome, especially as teenagers lack insight into the detrimental effects. E-cigarettes mimic regular cigarettes and usually contain nicotine with various flavors, and are used through inhalation. Despite the absence of carcinogenic tars, the use of e-cigarettes still has detrimental effects on several organ systems, such as the lungs and the heart, contributing to various short- and long-term side effects, which are still poorly understood. While using e-cigarettes, nicotine can disperse more quickly than in a regular cigarette due to bypassing first-pass metabolism, which is a phenomenon where a drug is metabolized at a certain location before entering circulation. Bypassing first-pass metabolism leads to higher concentrations of nicotine being delivered. This could lead to acute poisoning with nicotine inhalation toxicity, causing neuronal excitability and, in turn, lowering the seizure threshold. This secondary literature review investigates the effects of vaping on various organ systems, but especially the brain, and its potential for reducing the seizure threshold. This review suggests that e-cigarettes may be related to seizures in someone who has epilepsy. Since vaping is rapidly becoming a public health concern, this information underscores the importance of providing education to adolescents and cautioning them regarding the effects of e-cigarette use. Awareness campaigns and bans on promotions for e-cigarettes could help address the problem of consumption, shedding light on its possible detriments.

KEYWORDS: Vaping, Nicotine Toxicity, Epilepsy, Seizure Threshold, Adolescent Health.

Introduction

The days of regular cigarettes are declining, and many young individuals are drifting toward the trendy electronic cigarettes, also known as vapes or e-cigarettes. In 2020, the Centers for Disease Control and Prevention (CDC) reported that over 480,000 Americans and over 7,000,000 people worldwide had died from various smoking-related illnesses.[1] E-cigarettes were first made in 1965, in China, and patented in 2003.[2] According to the US Food and Drug Administration (FDA), e-cigarettes use a substance called "e-liquid," which is composed of nicotine "derived from tobacco, as well as flavorings, propylene glycol, vegetable glycerin, etc."[3] In addition to nicotine, e-cigarettes may also have ingredients like tetrahydrocannabinol (THC), cannabidiol, and butane hash oil, known as dabs,[4] along with some toxic components, such as formaldehyde, acetaldehyde, acetone, and metals, and volatile organic compounds (VOCs).[5] Each device consists of a rechargeable lithium-ion battery, an electrical heating coil, and liquid in a disposable cartridge or refillable tank, which have undergone modifications over the years. The electric heating coil in the device turns the e-liquid into an aerosol that is inhaled or 'vaped' through the mouthpiece.[6] Nicotine concentration ranges widely, anywhere from 0mg/mL (0%) to 59 mg/mL (5%), and may be more than that of a regular cigarette.[7] The most popular small disposable E-liquid cartridge system is a single JUUL, which contains 5% nicotine in each one, equivalent to one pack of cigarettes, which can give peak nicotine concentration rapidly with sweet flavors, making it even more addictive.[7]

Rapid delivery of the aerosolized nicotine and lack of the usual tobacco aroma made this an attractive option among American adolescents, in addition to fancy delivery systems with various additives. In 2020, the FDA banned cartridge-based e-cigarettes from the market.[8] This did not deter most vapers, as they were still able to acquire them, and many types of e-cigarettes evaded this category.[8] Seizures, as explained by the Mayo Clinic, are when the brain has abrupt and unrestrained outbursts of "electrical activity."[13] Epilepsy is a chronic disease in which people have recurring seizures of various causes. It has been observed that the use of cigarettes is connected to higher breakthrough seizure rates in people with epilepsy.[14] Due to cigarettes increasing seizure frequency, there is the question of whether e-cigarettes do the same. Though the underlying pathophysiology is not completely understood, researchers believe that nicotine modulates acetylcholine receptors and, in turn, contributes to the release of the excitatory neurotransmitter, glutamate.[14] Through the activation of nicotinic acetylcholine receptors (nAChRs) by nicotine, it is thought that there is a large, spontaneous release of glutamate.[15]

DOI: 10.36838/IJHSR812.30

In 2024, 1.63 million students in the United States used e-cigarettes.[32] Between November 2016 and August 2019, it was noted that the total ENDS sales increased by nearly 300 percent,[9] especially among middle and high school students. In the United States, 7.7% of students participate in vaping.[10] By education level, 4.6% of middle school students and 10% of high school students use e-cigarettes.[10] These numbers may be underestimating the true percentage of students who vape due to many students' unwillingness to report the truth. As seen in Figure 1, the prevalence of nicotine-based e-cigarettes in 2017 for those in 8th, 10th, and 12th grade was between 3.5% and 11%. In 2018 (Figure 2), the prevalence of nicotine-based e-cigarette use in grades 8, 10, and 12 almost doubled, as it was reported between 6.1% and 20.9%. There has been a steep increase in the percentage of teenagers who are using nicotine-based e-cigarettes, with the prevalence almost doubling from 2017 to 2018. Many people, especially teenagers, believe that vaping is a healthy alternative to smoking, but this is an incorrect assumption. Smoking cigarettes contributes to most cases of lung cancer and COPD, as stated by the American Lung Association. This does not imply that vaping is without physiological effects; flavors in e-cigarettes can lead to "unknown inflammation"[11] in the lungs, and the development of the brain can be affected, in addition to several other multi-system symptoms.[11]

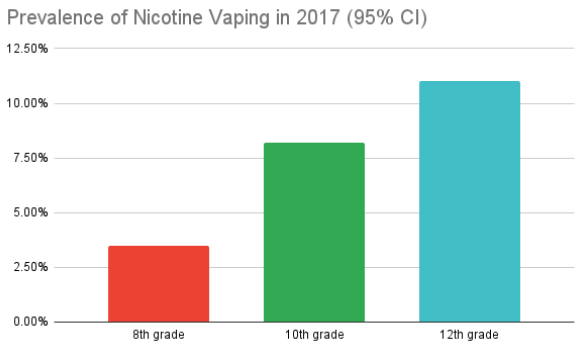


Figure 1: Prevalence of Nicotine Vaping in 2017. The prevalence of nicotine vaping in 2017 in grades 8, 10, and 12 ranged from 3.5% to 11%. Most students who vaped nicotine were in high school, 10th and 11th grade; 8.2% and 11% of students, respectively. Data gathered from Miech *et al.*, 2019.12 Created using Google Sheets.

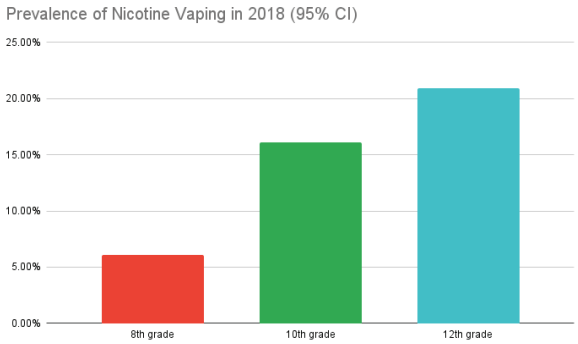


Figure 2: Prevalence of Nicotine Vaping in 2018. The prevalence of nicotine vaping in 2018 in grades 8, 10, and 12 ranged from 6.1% to 20.9%. Most students who vaped nicotine were in high school, 10th and 11th grade; 16.1% and 20.9% of students, respectively. The percentage of students who engaged in the practice nearly doubled from 2017. Data gathered from Miech *et al.*, 2019.12 Created using Google Sheets.

Through this paper, the effects of e-cigarette use on the brain, and more specifically on seizures, will be explored. The findings from this will possibly aid efforts for more education on the consequences of vaping.

Methodology

The key objective of this research paper is to determine the repercussions of e-cigarette use, specifically on seizure potentiation, its frequency, and the proximity to the use of vaping products. This paper is a secondary literature review, which is being conducted through the analysis of many case reports, journal articles, and review studies. To conduct this review, many articles from credible sources, including peer-reviewed original research articles on the health effects of e-cigarettes, e-cigarette use data, effects of vaping on the brain, and the frequency of seizures, were analyzed. A majority of these articles were found through PubMed, Google Scholar, and the Mayo Clinic. Studies published between 2015 and 2024 were considered for inclusion. Overall, a total of 17 peer-reviewed studies and sources were analyzed. A qualitative style of review was used to determine whether vaping increases seizure frequency in those who use e-cigarettes. After this detailed review of various studies, conclusions were made regarding the effect of vaping on the number of seizures, along with other physiologic harmful effects. The following search terms were used: "e-cigarettes," "vaping," "seizures," "epilepsy," "nicotine and brain activity," "nicotine-induced seizures," "neurological effects of vaping," "e-cigarettes and neurological health," and "adolescent vaping and seizures." Through this process of literature review, the ways nicotine affects the brain, and in turn, seizures were researched. Due to this being a literature review, the only materials used were online; no physical resources were used, resulting in no ethical considerations to be observed. Many articles and sources from various parts of the world were used, reducing the risk of research bias.

Discussion

Types of Seizures:

Seizures can be triggered or spontaneous. Spontaneous seizures without any clear identifiable metabolic, or toxic cause, may be related to an underlying genetic cause or structural abnormality. The 2017 International League against epilepsy (ILAE) proposed etiology-specific syndromes contributing to seizures and epilepsy.[16] According to the ILAE report, epilepsy or an epileptic seizure is a neurological disorder that can be associated with abnormal, excessive, or synchronous electrical activity of the brain, causing recurrence of seizures. This can present either as a focal or generalized event, with associated epileptiform abnormalities noted on the EEG. Age of onset can be variable, with a peak noted during the early years of life and another peak again noted in the elderly. Focal seizures are usually related to a focal lesion of the brain, like a stroke, a tumor, an intrinsic brain malformation, or a genetic disorder like tuberous sclerosis. An uncontrolled ongoing seizure lasting longer than 5 minutes is termed "Status epilepticus." This is a neurological emergency. In cases of "Status epilepticus,"

either a prolonged seizure or recurrent seizures without return to the neurologic baseline between seizures may occur.

Although on several occasions, one may not find a clear etiology on imaging studies for the seizure, there may be underlying epileptiform discharges on the EEG. This tendency to have a recurrence of seizures may also be associated with certain triggers. Seizure triggers are causes that can lower the seizure threshold in patients with a tendency to have epilepsy. This can include infections, electrolyte disturbances, metabolic derangements, sleep deprivation, stress, certain medications, and various drugs such as alcohol, cocaine, nicotine, newer synthetic hallucinogens like synthetic cannabinoids and spice. Newer synthetic agents pose challenges for diagnosis, as these are usually not seen in the standard lab testing.

Effects of E-Cigarettes on the Body:

In addition to being a contributor to abuse potential and dependence, nicotine, through vaping, has been seen to have a significant effect on various organ systems, most importantly the lungs. Vapers have differently altered airway epithelia than smokers. It was found that vaping causes significant changes to the airway epithelial proteome that are distinct from the changes seen in smokers, including an altered physical appearance of the airways with erythema and increased friability.[18] Airway epithelial cells play a role in managing the volume and composition of airway surface liquid (ASL), secreting mucins and cytokines, and also providing a barrier against harmful pathogens.[17] The vitamin E acetate with THC in e-liquid was believed to be the cause of the outbreak of acute lung injury related to vaping.[19] Although seeing the exact role of nicotine in this requires more studies, in those who had never smoked and had used an e-cigarette, the gene expression in their lungs changed much more than in those who had used a Propylene Glycol or Vegetable Glycerin (PG/VG) based e-cigarette. The other components of e-cigarettes, including nicotine, have been shown to affect the immune system. This could lead to more susceptibility to pulmonary diseases such as emphysema and bronchiectasis. Nicotine-containing e-liquids are one of the most important contributors to e-cigarette or vaping product use-associated lung injury (EVALI).[17]

Nicotine also has direct effects on the cardiovascular system, via activation of the sympathetic nervous system, given binding of nicotine to α3β4 receptors, triggering release of catecholamines (Epinephrine, Norepinephrine). This, in turn, activates the beta receptors in the heart, causing increased heart rate, workload, and cardiac contractility. In the long term, this over-stimulation of the sympathetic nervous system leads to cardiac remodeling, which promotes the development of heart failure and increases arrhythmogenesis. The use of e-cigarettes could also lead to changes in the heart. Increased arterial stiffness, vascular endothelial dysfunction, increased angiogenesis, cardiorenal fibrosis, and increased atherosclerotic plaque formation were found when observing the effect of chronic exposure to aerosols in animal models.[17] These mechanisms contribute to increased heart rate along with elevated systolic and diastolic blood pressure.

Other components of e-cigarettes have also been observed to cause issues in the body. Flavorings in e-cigarettes, such as sweeteners and chemical flavorants, might account for cardiopulmonary toxicity, asthma, and airway remodeling.

Effects of Nicotine on the Brain:

Nicotine is a key ingredient that is often found in e-cigarettes. The Food and Drug Administration (FDA) received about 117 cases of seizures and other neurological disorders in relation to vaping from April 3, 2019, to June 30, 2019, especially among adolescents.

Inhaled aerosolized nicotine bypasses the first-pass metabolism, allowing it to be quickly absorbed into various organ systems, including the brain. In contrast, only 30% nicotine gets into the systemic circulation through the oral or intraperitoneal route.[20] Nicotine has various complex interactions with neural networks, especially involving mesolimbic reward circuitry in the ventral tegmental area (VTA), striatum, prefrontal cortex, and other limbic pathways. Because of involvement in these reward circuits, substance abuse and addiction problems are primary concerns. Nicotine affects the developing brain of an adolescent more than that of an adult,[21] through specific receptors called nicotinic cholinergic receptors (nAChRs), which are widely distributed in abundance throughout the brain.[22,23] Various animal and human studies have shown that nicotine negatively impacts memory storage and other cognitive functions performed by the hippocampus, thus reducing teen attention, performance, and working memory.[24] Nicotine is reported to increase levels of glutamate in the hippocampus by decreasing glutamate uptake. Additional mechanisms of nicotine induce seizures by attenuation of excitatory amino acid transporter type 3 (EAAT3), thus affecting the clearance of glutamate, which is an excitatory neurotransmitter.[25] As per the Tobacco-Induced Neurotoxicity of Adolescent Cognitive Development (TINACD) theory, which was proposed in 2008 by deBry and Tiffany, smoking nicotine showed aberrant detrimental effects in emotional regulation by reducing the inhibitory control, thus making adolescents more impulsive and inattentive. The amygdala is the primary site in the brain for emotional processing and regulation, along with memory consolidation through its connections with the hippocampus. α7-containing nicotinic acetylcholine receptors (α-nAChRs) are highly expressed in the amygdala, which is an area also involved in seizure generation and epileptogenesis. Nicotine-induced neuronal activation is noted in the amygdala along with other areas involved in the limbic pathways.[26]

Review of Case Reports - Seizures and Vaping:

A case report by Liu-Zarzuela and Sun from the University of Texas focuses on a 20-year-old man who has had recurrent seizures every time he vapes for the past 5 years. The first time he used an e-cigarette, when he was fifteen, he experienced a tonic-clonic (grand mal) seizure.[27] At the age of 19, he used an e-cigarette for the second time, and shortly after, the patient experienced a seizure and had to be hospitalized. This led to him refraining from vaping. Three months later, he used an e-cigarette once again, and within minutes, he experienced a

DOI: 10.36838/IJHSR812.30

generalized tonic-clonic seizure. His labs and scans were all unremarkable and normal, and he was put on levetiracetam (500 mg). The patient quit vaping and started using a nicotine patch. The patches allow for the release of nicotine to be slower, possibly contributing to the lack of seizures after the patient ceases vaping. He had no history of epilepsy, no family history of seizures, nor any risks for them, leading to the conclusion that his seizures were caused by vaping. Nicotine toxicity has been seen in many young people due to e-cigarettes. The authors note that this could lead to interruptions in the neuronal excitability, lowering the "seizure threshold". Exogenous nicotine leads to nAChRs being released into parts of the brain such as the neocortex, amygdala, hippocampus, and thalamus, possibly encouraging seizures.[26] The disruption caused by nicotine leads to overactivity in the brain, encouraging seizures. Nicotine has also been seen to release glutamate, which also excites brain cells, which may trigger seizures. The authors acknowledge that to get a clear link on the relationship between seizures and vaping, more investigation is needed, but it is an important phenomenon to question in terms of public health awareness.[27]

E-cigarettes are popular among youth as the manufacturers make them quite appealing to them with various flavors. These flavorants make adolescents misjudge the nicotine exposure that they are experiencing and their effect on the brain.

One of the case reports by Tatum and Oster follows an eighteen-year-old female without a history of epilepsy. The patient's first seizure occurred in 2019, thirty minutes after she had vaped using a "Juul" e-cigarette, which contains nicotine. During her seizure, she toppled over and hit her head, leading to unconsciousness. She was taken to the emergency room and was quickly discharged with a prescription of levetiracetam (500mg) after her EEG monitoring was reportedly normal without any epileptiform activity. Six weeks later, the patient experienced another seizure one hour after she used a vape with nicotine in it. However, it is important to note that the patient had discontinued levetiracetam by herself two weeks prior to the second seizure episode. The patient endured two more seizures a couple of weeks after the second one, again, within one hour of using the nicotine-based e-cigarette. Her last seizure occurred in early-mid 2020, and she had no other instances of seizures after stopping vaping. In this patient without a prior history of seizure disorder, it was seen that vaping could have been a possible trigger for her seizures, since all of them occurred within an hour of using an e-cigarette.[29]

Based on the case report from Tufts Medical Center in Boston, MA, there is objective evidence of seizure recurrence after adequate control in a young 32-year-old male patient with refractory bitemporal epilepsy in the setting of data available from the responsive neurostimulation (RNS) device (Figure 3). He underwent electroencephalography (EEG) monitoring starting in 2011. By 2017, he was living his life comfortably, was working, and was no longer enduring auras, among other seizure-related advancements. Suddenly, in 2019, there was an increase in ictal and interictal activity that was noted on the RNS recordings, and this was when the patient had started to vape. This report further helps to be aware of the risk associated with the usage of e-cigarettes with tobacco and cannabinoids, resulting in breakthrough seizures.[15]

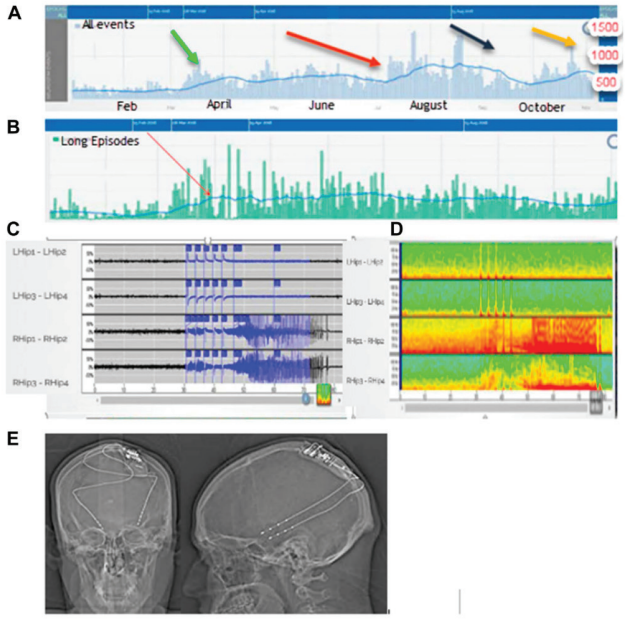


Figure 3: Responsive neurostimulator (Neuropace) Data. RNS device showing the patient's seizure data. Adapted from Oster *et al.* (2021).[15]
A. Data showing interictal discharges and seizures, with worsening especially when associated with vaping (long red arrow)
B. A tally of long episodes of ictal & interictal activity correlating with vaping (thin arrow)
C; D. Recordings of ictal activity; the same data presented in a compressed spectral array.
E. X-ray showing the impulse generator of the RNS device in the posterior head region and the bitemporally placed leads, which record data and stimulate to terminate the electrographic seizure.

In the case report by Tatum and Oster, the focus was on a patient who did not have a history of epilepsy. But in this report by Oster and others, the focus was on a patient with a history of epilepsy. In both works, findings suggest a correlation between vaping and increased seizure frequency.

Lastly, a case was reported regarding a 16-year-old girl with a diagnosis of idiopathic generalized epilepsy at the age of 9. Her semiology included eye rolling, eye fluttering, at times, progressing into generalized tonic-clonic. EEG showing generalized spike and wave discharges, suggesting an underlying generalized epilepsy disorder. Over the years, she tried various anti-seizure regimens, and finally, she was well controlled on lamotrigine monotherapy. After being seizure-free for 7 months, she had recurrent breakthrough seizures over a few days, a total of 5, without any clear trigger. Later, she was found to be vaping in association with these seizures, using a commercial device with nicotine juice. She reported usage of this spearmint, fruit punch, or watermelon flavored vapes, 4 months prior to these breakthrough seizures. She confirmed the usage of e-cigarettes prior to four out of the five reported seizures, but denied any other illicit drug use or alcohol.[29]

E-cigarettes have a plethora of side effects. In the cases observed in the paper, there is seen to be somewhat of a correlation between e-cigarettes and seizure frequency. In most cases, seizures occur within minutes to an hour after vaping.

Through more recognition of the effect of vaping on seizures, the quality of life could be greatly improved. As previously noted, vaping is not healthier than smoking–vaping can lead to various health problems, such as lung disease or heart disease.[31] Inclusion of this repercussion of e-cigarettes could lead to fewer people partaking in the harmful practice, so that they can lead a healthier, more fulfilling life.

Case	Patient Demographics	E-Cigarette Use	Seizure Occurrence	Intervention	Outcome
Case 1	20 year old male, no history of epilepsy	Occasional use of E-cigarette	Seizure experienced minutes after vaping	He was put on levetiracetam and started using a nicotine patch	Did not report to have any more seizures.
Case 2	18 year old female, no history of epilepsy	Used a "Juul" e-cigarette occasionally	Seizures occurred ~30 minutes-1 hour after vaping	She was put on levetiracetam and quit vaping	Did not report to have any more seizures.
Case 3	32 year old male, history of refractory bitemporal epolepsy	Started to use an E-cigarette in 2019	Increase in ictal and interictal activity was noted in RNS activity	Was undergoing EEG monitoring and had an RNS device	N/A
Case 4	16 year old female, history of idiopathic generalized epilepsy	Started to use an E-cigarette with nicotine juice	⅘ of her seizures occured after the use of an E-cigarette	N/A	N/A

Figure 4: Summary of case studies. Overall, a link can be seen between e-cigarette use and seizures.

Conclusion

More research is necessary to establish a definitive causal effect between seizures and vaping. Healthcare professionals also should have a high index of clinical suspicion when an adolescent presents with a new-onset seizure or concerning neurological symptoms, to get a detailed history about vaping products and report these events to the Safety Reporting Portal (SRP). More intense regulatory measures on e-cigarettes should be in place, such as strict law enforcement to curb illegal sales, a promotion ban, a ban on flavoring to be less enticing to teenagers, and possibly increasing the federal ENDS taxes. It is also critical that more education and awareness campaigns be undertaken regarding the consequences of e-cigarettes and their toxic additives.

The purpose of this research was to determine the relation between e-cigarette use and seizures. This was conducted as a secondary literature review through the analysis of multiple primary source articles and studies to conclude. Due to e-cigarettes only recently being popularized, there has not been much exploration of the connection between e-cigarettes and seizure frequency. With the study of long-term impacts of vaping on neurological health, valuable findings would come about. Despite the benefits of using articles, this also poses a limitation: no firsthand studies were conducted for this paper. Through the examination of many articles, a link between e-cigarette use and seizure frequency can be suggested.

Acknowledgments

I would like to showcase my gratitude for the support and mentorship provided by Dr. Kathryn Wilwohl, Professor Virgel Torremocha, and Gifted Gabber School.

References

1. CDC. (2020, April 28). *Tobacco-Related Mortality*. Archive.cdc.gov. https://archive.cdc.gov/#/details?url=https://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/tobacco_related_mortality/index.htm
2. Lichtenberg, K. E-Cigarettes: Current Evidence and Policy. *Missouri Medicine* **2017**, *114* (5), 335–338.
3. FDA. E-Cigarettes, Vapes, and Other Electronic Nicotine Delivery Systems (ENDS). *FDA* **2024**.
4. Layden, J. E.; Ghinai, I.; Pray, I.; Kimball, A.; Layer, M.; Tenforde, M.; Navon, L.; Hoots, B.; Salvatore, P. P.; Elderbrook, M.; Haupt, T.; Kanne, J.; Patel, M. T.; Saathoff-Huber, L.; King, B. A.; Schier, J. G.; Mikosz, C. A.; Meiman, J. Pulmonary Illness Related to E-Cigarette Use in Illinois and Wisconsin — Preliminary Report. *New England Journal of Medicine* **2019**, *382* (10). https://doi.org/10.1056/nejmoa1911614.
5. Breland, A.; Soule, E.; Lopez, A.; Ramôa, C.; El-Hellani, A.; Eissenberg, T. Electronic Cigarettes: What Are They and What Do They Do? *Annals of the New York Academy of Sciences* **2016**, *1394* (1), 5–30. https://doi.org/10.1111/nyas.12977.
6. CDC. "About E-Cigarettes (Vapes)." *Smoking and Tobacco Use*, 9 May 2024, www.cdc.gov/tobacco/e-cigarettes/about.html. m
7. Raymond, B. H.; Collette-Merrill, K.; Harrison, R. G.; Jarvis, S.; Rasmussen, R. J. The Nicotine Content of a Sample of E-Cigarette Liquid Manufactured in the United States. *Journal of Addiction Medicine* **2018**, *12* (2), 127–131. https://doi.org/10.1097/adm.0000000000000376.
8. American Academy of Pediatrics. (2020). *Flavored E-Cigarette and Tobacco Products*. Aap.org. https://www.aap.org/en/advocacy/state-advocacy/flavored-e-cigarette-and-tobacco-products/?srsltid=AfmBOor410XE4ApxXndyXUCAWoHISGDvAJr2w_bhmZ3u2Kiy-htQFElK
9. Ali, F. R. M. E-Cigarette Unit Sales, by Product and Flavor Type — United States, 2014–2020. *MMWR. Morbidity and Mortality Weekly Report* **2020**, *69*. https://doi.org/10.15585/mmwr.mm6937e2.
10. CDC. (2024, October 17). *E-Cigarette use among youth*. Smoking and Tobacco Use; CDC. https://www.cdc.gov/tobacco/e-cigarettes/youth.html
11. King, G. (2021, April 5). *Doctor's warning about dangers of vaping*. Mayo Clinic Health System; Mayo Clinic Health System. https://www.mayoclinichealthsystem.org/hometown-health/speaking-of-health/a-doctors-warning-about-the-dangers-of-vaping
12. Miech, R.; Johnston, L.; O'Malley, P. M.; Bachman, J. G.; Patrick, M. E. Adolescent Vaping and Nicotine Use in 2017–2018 — U.S. National Estimates. *New England Journal of Medicine* **2019**, *380* (2), 192–193. https://doi.org/10.1056/nejmc1814130.
13. Mayo Clinic. (2023, September 2). *Seizures - Symptoms and Causes*. Mayo Clinic; Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/seizure/symptoms-causes/syc-20365711
14. Zhong, R.; Li, Z.; Zhang, X.; Chen, Q.; Lin, W. Current Cigarette Smoking Is Associated with a High Seizure Frequency and Anxiety Symptoms in People with Epilepsy. *Frontiers in Neurology* **2022**, *13*. https://doi.org/10.3389/fneur.2022.834694.
15. Oster, J. M.; Tatum, P.; Monigan, C.; Kryzanski, J. Seizures Noted by Responsive Neurostimulation from E-Cigarette Use (Vaping). *Journal of Clinical Neurophysiology* **2021**, *39* (1), 1–3. https://doi.org/10.1097/wnp.0000000000000866.
16. Specchio, N.; Wirrell, E. C.; Scheffer, I. E.; Nabbout, R.; Riney, K.; Samia, P.; Guerreiro, M.; Gwer, S.; Zuberi, S. M.; Wilmshurst, J. M.; Yozawitz, E.; Pressler, R.; Hirsch, E.; Wiebe, S.; Cross, H. J.; Perucca, E.; Moshé, S. L.; Tinuper, P.; Auvin, S. International League against Epilepsy Classification and Definition of Epilepsy

DOI: 10.36838/IJHSR812.30

Syndromes with Onset in Childhood: Position Paper by the ILAE Task Force on Nosology and Definitions. *Epilepsia* **2022**, *63* (6). https://doi.org/10.1111/epi.17240.
17. Herman, M.; Tarran, R. E-Cigarettes, Nicotine, the Lung and the Brain: Multi-Level Cascading Pathophysiology. *The Journal of Physiology* **2020**, *598* (22). https://doi.org/10.1113/jp278388.
18. Ghosh, A.; Coakley, R. C.; Mascenik, T.; Rowell, T. R.; Davis, E. S.; Rogers, K.; Webster, M. J.; Dang, H.; Herring, L. E.; Sassano, M. F.; Livraghi-Butrico, A.; Van Buren, S. K.; Graves, L. M.; Herman, M. A.; Randell, S. H.; Alexis, N. E.; Tarran, R. Chronic E-Cigarette Exposure Alters the Human Bronchial Epithelial Proteome. *American Journal of Respiratory and Critical Care Medicine* **2018**, *198* (1), 67–76. https://doi.org/10.1164/rccm.201710-2033oc.
19. CDC. "Smoking and Tobacco Use; Electronic Cigarettes." *Centers for Disease Control and Prevention*, 3 Aug. 2021, archive.cdc.gov/www_cdc_gov/tobacco/basic_information/e-cigarettes/severe-lung-disease.html.
20. Shao, X. M.; Fang, Z. T. Severe Acute Toxicity of Inhaled Nicotine and E-Cigarettes: Seizures and Cardiac Arrhythmia. *CHEST* **2020**, *157* (3), 506–508. https://doi.org/10.1016/j.chest.2019.10.008.
21. Azam, L.; Chen, Y.; Leslie, F. M. Developmental Regulation of Nicotinic Acetylcholine Receptors within Midbrain Dopamine Neurons. *Neuroscience* **2007**, *144* (4), 1347–1360. https://doi.org/10.1016/j.neuroscience.2006.11.011.
22. Feduccia, A. A.; Chatterjee, S.; Bartlett, S. E. Neuronal Nicotinic Acetylcholine Receptors: Neuroplastic Changes Underlying Alcohol and Nicotine Addictions. *Frontiers in Molecular Neuroscience* **2012**, *5*. https://doi.org/10.3389/fnmol.2012.00083.
23. Zoli, M.; Pistillo, F.; Gotti, C. Diversity of Native Nicotinic Receptor Subtypes in Mammalian Brain. *Neuropharmacology* **2015**, *96* (96), 302–311. https://doi.org/10.1016/j.neuropharm.2014.11.003.
24. Jacobsen, L. K.; Krystal, J. H.; Mencl, W. E.; Westerveld, M.; Frost, S. J.; Pugh, K. R. Effects of Smoking and Smoking Abstinence on Cognition in Adolescent Tobacco Smokers. *Biological Psychiatry* **2005**, *57* (1), 56–66. https://doi.org/10.1016/j.biopsych.2004.10.022.
25. Jacobsen, L. K.; Pugh, K. R.; Constable, R. T.; Westerveld, M.; Mencl, W. E. Functional Correlates of Verbal Memory Deficits Emerging during Nicotine Withdrawal in Abstinent Adolescent Cannabis Users. *Biological Psychiatry* **2007**, *61* (1), 31–40. https://doi.org/10.1016/j.biopsych.2006.02.014.
26. Hea Jun Yoon; Young Jin Lim; Zuo, Z.; Hur, W.; Do, S. Nicotine Decreases the Activity of Glutamate Transporter Type 3. *Toxicology Letters* **2014**, *225* (1), 147–152. https://doi.org/10.1016/j.toxlet.2013.12.002.
27. Leslie, F. M. Unique, Long-Term Effects of Nicotine on Adolescent Brain. *Pharmacology Biochemistry and Behavior* **2020**, *197* (173010), 173010. https://doi.org/10.1016/j.pbb.2020.173010.
28. Liu-Zarzuela, J. A.; Sun, R. Three Seizures Provoked by E-Cigarette Use in a Five-Year Period: A Case Report. *Cureus* **2022**. https://doi.org/10.7759/cureus.27616.
29. Tatum, P. S.; Oster, J. M. Seizures and Vaping—a Report of Two Cases. *Military Medicine* **2021**, *188* (5-6). https://doi.org/10.1093/milmed/usab270.
30. Wharton, J. D.; Kozek, L. K.; Carson, R. P. Increased Seizure Frequency Temporally Related to Vaping: Where There's Vapor, There's Seizures? *Pediatric Neurology* **2020**, *104* (P66-67), 66–67. https://doi.org/10.1016/j.pediatrneurol.2019.10.006.
31. American Lung Association. (2018). *Health Risks of E-Cigarettes and Vaping*. Lung.org. https://lung.org/quit-smoking/e-cigarettes-vaping/impact-of-e-cigarettes-on-lung
32. Centers for Disease Control and Prevention. *E-Cigarette use among youth*. Smoking and Tobacco Use. https://www.cdc.gov/tobacco/e-cigarettes/youth.html.

Author

Sarayu Gadam is a high school junior who is passionate about neuroscience. She is interested in epilepsy as well as its impacts on youth. Sarayu hopes to contribute to the field and pursue a career in it.

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Evaluating Diagnostic Tools for Frontotemporal Dementia: Current Limitations and Emerging Advances

Anusha Manda

Dripping Springs High School, 940 US-290 West, Dripping Springs, TX 78620, USA; manda.anu07@gmail.com

ABSTRACT: Frontotemporal dementia (FTD) is associated with one of the highest patient burden rates among all dementia types, typically affecting individuals between the ages of 40 and 65. Diagnosing FTD is challenging due to its symptomatic overlap with other neurodegenerative disorders, such as Alzheimer's Disease (AD). Clinical data show that only one-third of suspected FTD cases are ultimately confirmed, while almost half of misdiagnoses stem from misinterpreted neuroimaging. Current diagnostic methods lack FTD-specific biomarkers, instead relying on indicators shared with AD, compromising accuracy. This review synthesizes current limitations in primary diagnostic approaches for FTD, highlighting how integrative methods in neuroimaging and cerebrospinal fluid biomarkers, alongside advancements in genetic screening, may substantially improve diagnostic precision. Genetic testing effectively identifies pathogenic variants, which can be enhanced by incorporating modifier genes that influence disease onset and progression. Neuroimaging tools like MRI and FDG-PET provide structural and metabolic information about the brain, and their limitations can be alleviated through artificial intelligence and multimodal integration. Cerebrospinal fluid biomarkers—including neurofilament light chain, t-tau, and inflammation-related markers—support differential diagnosis. Multi-protein panels demonstrate high specificity with an AUC ranging from 0.91 to 0.96 in distinguishing FTD. Together, these findings support early detection and address a critical diagnostic gap.

KEYWORDS: Behavioral and Social Sciences, Neuroscience, Frontotemporal Dementia (FTD), Biomarkers, Diagnostic Tools.

Introduction

Frontotemporal dementia (FTD) remains one of the most frequently misdiagnosed neurodegenerative disorders. This is largely due to overlapping behavioral symptoms and a limited number of reliable biomarkers. This diagnostic limitation contributes not only to delayed treatment but also to prolonged psychological and emotional burden for patients and their families. The difficulty in distinguishing the primary causes of FTD through current biomarkers represents a critical area of research focused on improving biomarker accuracy. FTD refers to a group of neurodegenerative disorders that primarily cause damage to the neurons in the frontal and temporal lobes. This degeneration ultimately leads to the atrophy of nerve cells.[1] This targeted damage results in two major variants: behavioral variant FTD (bvFTD) and primary progressive aphasia (PPA), as illustrated in Figure 1.

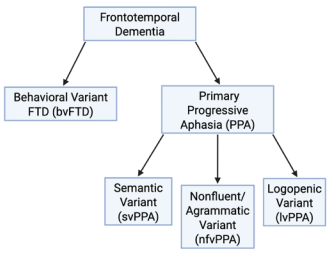


Figure 1: Major Variants that Make Up FTD. This diagram presents the major clinical subtypes of FTD, organized into behavioral and language-based categories. Each subtype is associated with distinct patterns of neurodegeneration and symptoms. Information gathered from the National Institute on Aging and created using Biorender.

Generally, symptoms of FTD can be associated with the region of the brain that has experienced atrophy. Although that is the primary cause of FTD symptoms, the neurodegeneration itself is driven by underlying causes such as protein accumulation and genetic mutations. Determining reliable biomarkers that can pinpoint the primary underlying causes of FTD remains a complex and ongoing area of research. This requires relying on diagnosis based on personal history and behavioral symptoms, which can result in symptoms being overlooked or misdiagnosed. For instance, behavioral variant FTD—the most common variant of FTD, is frequently misdiagnosed for psychiatric and other neurological disorders due to its symptoms overlapping. This pattern persists across other subtypes of FTD, where there is a lack of identified biomarkers that can provide a near-definite diagnosis and prevent misinterpretation of FTD.

A 2025 study by the University of Queensland found that of the 100 patients used in their study, only 34 of them were found to be true-positive cases of FTD.[2] It was later concluded that misinterpretation of nuclear imaging was the cause of about 50% of all falsely diagnosed cases. FTD is one of the most common types of young-onset dementia, typically affecting individuals between the ages of 40 and 65. It accounts for about 5-15% of all dementia cases.[3] FTD's nature to affect a younger population compared to other dementias often contributes to the high rates of misdiagnosis since it falls out of the age range that dementia is usually associated with.

DOI: 10.36838/IJHSR812.36

Another study conducted in Brazil, by two researchers, examined the diagnostic outcomes across 30 patients, which included 10 with FTD, 10 with Alzheimer's disease, and 10 with no neurodegenerative disorders. During the span of 12 months, it was found that FTD patients exhibited the highest misdiagnosis rate, often diagnosed with psychiatric disorders such as depression, bipolar disorder, schizophrenia, and obsessive-compulsive disorder, especially common for the behavioral variant.[4] These data can be linked to FTD cases being more likely to be misdiagnosed as a psychiatric disorder if the patient has a personal history of psychiatric disorders.

Several biomarkers used in diagnosis are not exclusive to FTD but are shared across multiple diseases. This can impede an accurate diagnosis by making the misinterpretation of biomarkers harder to avoid. Research conducted by the University of Washington in 2025 analyzed over 10,000 samples and found that 996 proteins are shared across Alzheimer's, Parkinson's, and FTD.[5] Protein accumulation, along with genetic mutations in FTD, is challenging to identify, and various studies urge researchers to focus on distinguishing biomarkers specifically for FTD.

Frontotemporal dementia, like most neurodegenerative diseases, lacks a definitive cure, with current therapies focused primarily on symptom management. The delay in diagnosis prevents individuals with FTD from accessing appropriate treatment for several years. Without an accurate diagnosis, patients may receive ineffective treatment while the disease continues to progress. This outcome is preventable—however, doing so requires a deeper understanding of FTD and its diagnostic indicators. This study addresses this need by evaluating current diagnostic tools and proposing strategies to overcome their limitations, with the goal of identifying reliable FTD-specific biomarkers.

Methodology

This study is a literature review focused on the most prevalent diagnostic tests used in FTD diagnosis to identify strengths and limitations based on various research articles. After identifying a gray area in diagnosis, this research analyzed emerging evidence on potential biomarkers that display the potential to enhance diagnostic accuracy. To conduct this research, this paper solely used online and published sources from scientific databases such as the National Library of Medicine and the National Institute of Health. The selection criteria required that all sources be peer-reviewed when applicable and published within approximately the last ten years, generally between 2015 and 2025. In addition to peer-reviewed studies, reputable scientific and medical websites such as the Wisconsin Alzheimer's Institute were included when they provided foundational or contextual information necessary to understand diagnostic procedures or biomarker mechanisms. Using these criteria, a total of 55 sources were reviewed. Searches were conducted using terms such as "FTD emerging biomarkers 2020-2025," "FTD brain scans compared to AD," "emerging advances in FTD diagnosis," and similar variations to capture both established and developing diagnostic approaches. First, research papers containing current diagnostic methods for FTD were gathered to understand the procedure for the diagnostic test and to then recognize the limitations of what each test could detect. Afterwards, papers that focused on investigating biomarkers that may eliminate the limitation identified through experiments and trials were analyzed to see if the results supported the use of the biomarker for the diagnosis of frontotemporal dementia. The primary goal of this paper is to highlight biomarkers and diagnostic tests that provide a high diagnostic accuracy for FTD. After an in-depth analysis of current diagnosis methods, this review was able to find promising biomarkers that require further research to be able to implement them in current diagnosis. As this study relied solely on literature-based sources and did not involve any physical tools, materials, or human participants, no ethical considerations were required.

Discussion

Frontotemporal Dementia:

Frontotemporal Dementia (FTD) is a term used to describe a group of neurodegenerative disorders that belong to a broader category of frontotemporal lobar degeneration (FTLD). Within FTD, different variants arise, which can be characterized by symptoms including difficulty in emotional and behavioral control. In addition to psychological abnormalities, FTD may result in executive dysfunction. This includes difficulty in movement or walking and trouble verbally communicating. These various symptoms are the result of damaged neurons in the brain, which gradually lead to shrinkage in the area in which the damage occurred.[6]

Degeneration caused by FTD primarily takes place in the frontal and temporal lobes of the brain. The lobe of the brain that is affected first or most is dependent on which symptoms appear first. If a common pattern of difficulty in decision making or using appropriate behavior in a specific setting is found, then it can be correlated to degeneration in the frontal lobe. The same can be observed with degeneration in the temporal lobe, which is associated with symptoms like language comprehension and processing auditory information.[1]

The cause of FTD is not yet identified, but there is evidence of underlying causes that can trigger progressive degeneration of neurons. One of the major contributors to atrophy is protein aggregation—a common characteristic in neurodegenerative diseases.[7] An error in protein misfolding is a common occurrence within cells, and it is often managed through cellular mechanisms that have evolved to ensure proper folding. These cellular mechanisms include a variety of molecular chaperones that facilitate either refolding misfolded proteins or redirecting proteins that cannot be refolded to ubiquitin and autophagy, which are cellular mechanisms that are able to degrade the misfolded protein.[8] If these mechanisms fail to eliminate the misfolded proteins, over time, the proteins will abnormally cluster and accumulate in the frontal and temporal lobes of the brain, which will lead to neurotoxicity and neurodegeneration.[9] The specific proteins that are associated with the development of FTD through protein aggregation are tau pathology and TDP-43 proteinopathies. The mislocalization

and buildup of these proteins have been linked to neuronal atrophy.[10, 11]

Around 30% of all FTD cases have a strong family history. This heritability is attributed to the genetic mutations that are passed down in an autosomal dominant pattern. The mutated genes that have been identified to cause FTD are *MAPT*, *GRN*, and *C9orf72*. These genes encode for cellular functions in the nervous system and contain the instructions for coding specific proteins. These proteins include tau and TDP-43: a mutation in these genes often results in improper protein formation, hence becoming an underlying cause for FTD. Tau pathology is the result of a mutation in the *MAPT* gene, while TDP-43 proteinopathies are the result of mutations in the *GRN* and *C9orf72* genes. Besides disrupting protein synthesis, these three genes can also interfere with cellular mechanisms that monitor protein folding and eliminate any improperly folded proteins.[12]

Prevalence & Demographics:

Over 10,000 papers dedicated to FTD research have been published between 2000 and 2022. These papers have been published in over 900 academic journals from over 80 countries/regions.[13] These articles touched upon various topics with a trend of research in the pathological mechanisms of FTD. Even though FTD is classified as a rare disease with a global prevalence rate of 15-22 per 100,000 among individuals aged 45 to 64, FTD remains a disease worth researching.[14] Studying FTD unlocks a broader path of understanding mechanisms like protein aggregation, tauopathy, and TDP-43 pathology. Additionally, FTD has a strong link with mutations in the *MAPT*, *GRN*, and *C9orf72* genes, which are prevalent in other neurodegenerative diseases—for example, Progressive Supranuclear Palsy (PSP), Corticobasal Degeneration (CBD), Parkinson's Disease (PD), and Amyotrophic Lateral Sclerosis (ALS).[15] Apart from how many people this neurodegenerative disease affects, it is important to understand how this disease affects the quality of life for someone who struggles with FTD.

While prevalence remains low, the age distribution is notable as approximately 60% of all FTD cases occur between the ages of 45 and 60. This dementia tends to occur at a younger age compared to most dementias and is also known as a young-onset dementia.[6] Due to this disease affecting the individuals earlier than most neurodegenerative diseases, it contributes to a wide range of challenges across physical, financial, emotional, and psychological domains. These challenges can be experienced both by the individual with FTD, who often experiences behavioral and psychological symptoms, and by caregivers who manage the resulting demands of care. To quantify these domains, clinical assessments such as the Neuropsychiatric Inventory (NPI), which measure patient neuropsychiatric symptoms severity, and the Zarit Burden Interview (ZBI), which measures caregiver burden, are commonly used.

Scales for Patient Symptom Severity & Caregiver Burden	Description	Scoring Index	FTD	AD
Neuropsychiatric Inventory	Measures the presence, frequence, and severity of behavioral and psychological symptoms in dementia patients	Total score range: 0 to 144 Higher scores indicate higher symptom severity	58.0 ± 19.3	3.6 ± 4.7
Zarit Burden Inventory	Assessing caregivers' emotional, physical, and social burden	Total score range: 0 to 48 0-10: no to mild burden, 10-20: mild to moderate burden, > 20: high burden	23.62 ± 15.9	12.26 ± 9.74

	FTD	AD
Financial Patient Burden (annual direct cost in USD)	$47,916	$28,078

Aside from direct cost, there are indirect costs such as loss of productivity and income for the patient and caregiver. This may significantly add to the annual cost in addition to the direct cost.

Figure 2: Comparison of Patient Symptom Severity and Caregiver Burden Between FTD and AD. When these three aspects of burden were compared—patient symptom severity (NPI), caregiver burden (ZBI), and financial burden—were compared between FTD and AD, individuals with FTD and their caregivers experienced significantly greater overall burden. Information gathered from Cummings, Bahia & Viana, Wisconsin Alzheimer's Institute, Liu *et al.*, Galvin *et al.*, and Nandi *et al.* Created using Biorender.

Cummings provides a comprehensive overview and description of the NPI, an assessment developed to assess behavioral and psychological symptoms of dementia. The NPI includes multiple variations, such as the original full version, which consists of all twelve domains, and the questionnaire version, a shortened, self-administered version.[16] Bahia & Viana conducted research that suggests that the original version of the NPI was used for both FTD and AD patients due to the number of domains evaluated; the comparative results are displayed in Figure 2.[17] In addition to the NPI, caregiver burden was assessed using the ZBI, which also includes different variations of the assessment.[18] When evaluating the caregiver burden for FTD and AD, Liu *et al.* likely used the short 12-item ZBI version, as indicated by the scoring range.[19] The last aspect used to compare patient burden between FTD and AD in Figure 2 is the financial burden in USD. The annual direct costs in USD for FTD and AD were sourced from Galvin *et al.* and Nandi *et al.*, respectively.[20, 21] Collectively, these three aspects demonstrate the challenges associated with a young-onset neurodegenerative disease.

A dementia is classified as a young-onset dementia if the onset of symptoms occurs before the age of 65. The symptoms of young-onset dementia can vary from those seen initially in older adults. These symptoms include behavior, language, personality change, and executive dysfunction. These symptoms can be commonly found in FTD variants such as bvFTD. The early development of dementia is linked to the heritability of dementia. This is demonstrated in FTD as it is more likely to be inherited than other dementia types, highlighting how genetic mutations have a strong influence on its development. Genetic mutations can trigger disease-causing processes like protein aggregation to occur at a younger age. Although FTD development happens before most dementias, it does not necessarily make it easier to detect. In fact, there remain various hurdles in concluding an accurate diagnosis.[22] These difficulties are derived from FTD's overlapping symptoms with psychiatric disorders or AD. Since several of the initial symptoms are

DOI: 10.36838/IJHSR812.36

behaviorally related changes, it can be easily misinterpreted. Moreover, there is no definitive test for FTD, and it takes various diagnostic tests to come to a diagnosis. Interpreting the results of multiple diagnostic tests can be challenging, often making it difficult to arrive at an accurate diagnosis. However, the available diagnostic tests can detect biomarkers indicative of FTD, aiding in the identification of its primary causes.[23]

Current Limitations and Emerging Advances in Diagnostic Tools:

Genetic Testing:

Genetic testing is a well-established and highly accurate diagnostic test, especially for neurodegenerative disorders, with underlying causes that include genetic factors like FTD. As medicine advanced, the complexity of genetic testing grew as its scope for identifying genetic mutations spread from single gene analysis to examining the entire genome.[24] Genetic testing's ability to detect pathogenic variants plays a critical role in accurate diagnosis. Genetic mutations are a prominent factor, occurring in nearly 40% of all identified cases of FTD, and have the potential to provide an accurate diagnosis.[25]

Typically, genetic testing is recommended when a family history of a specific disease is found in the patient's records. A strong family history of a disease can indicate that a genetic cause is triggering the development of the disease. Genetic testing may also be recommended if a clinician identifies a trend in the patient's symptoms or from previous diagnostic test results. Aside from clinician recommendations, it can additionally be used as predictive testing. If the patient is aware of a family history of FTD or related disorders like ALS, it may become a way of testing and diagnosing a patient before the symptoms develop. An early diagnosis of neurodegenerative disorders is the most desirable outcome, as symptoms can be immediately treated.

Genetic tests analyze deoxyribonucleic acid (DNA), which is acquired from samples of blood, hair, skin, amniotic fluid, or other tissues. The DNA is further examined in a laboratory where it is inspected for changes in chromosomes, DNA, and protein.[26] In the case of FTD, the most common mutations found from genetic tests occur in the *MAPT*, *GRN*, and *C9orf72* genes. More rare genetic mutations include *TARDBP*, *VCP*, *FUS*, *CHMP2B*, *SQSTM1*, *UBQLN1*, or *TBK1* genes.[25]

	Protein & Function	Prevalence in FTD	Penetrance	Disease Duration
C9orf72	**C9orf72** Lysosomal homeostasis	20-25% of familial FTD & 6-8% of sporadic FTD	~0% at age 35 years, 50% at age 58 years, & near 100% at age 80 years	6.4 years on average
GRN	**Progranulin** Lysosomal homeostasis; inflammation	5-25%of familial FTD & 5% of sporadic FTD	Over 60% affected by age 60, rising to 90% by 75 and over 95% by 70	3-12 years
MAPT	**Microtubule-associated protein tau** Microtubule stabilization and assembly	5-20% of familial FTD & 2% of sporadic FTD	Typically a fully penetrant disorder, with rare cases of reduced penetrance in some families with specific *MAPT* variants	9.3 years on average

Figure 3: Summary of Mutated Genes Directly Linked to FTD. The table depicts each gene and the protein associated with the gene, its prevalence in FTD, the proportion of individuals with FTD who will show the associated signs and symptoms, and the duration of the FTD for an individual with that mutated gene. Information gathered from Sirkis *et al.*, Zampatti *et al.*, Gossye *et al.*, Hsiung *et al.*, Rohrer *et al.* Created using Biorender.

Sirkis *et al.* emphasizes FTD's heritability and provides a summary of the major mutated genes in FTD including each gene's associated protein and function, as well as the prevalence rates that were used in Figure 3.[27] Information on penetration and disease duration for *C9orf72* was gathered from Zampatti *et al.* and Gossye *et al.* respectively; both articles investigated the genetic complexity of FTD, particularly in relation to *C9orf72*.[28, 29] Similarly, the penetration and disease duration estimated for *GRN* were derived from Hsiung *et al.*, and for *MAPT* from Rohrer *et al.*[30, 31]

For over a decade, mutations in genes such as *C9orf72*, *GRN*, and *MAPT* have served as key diagnostic markers for FTD and have been identified through genetic testing. However, this diagnostic test does not complete the full picture of a patient's genetic risk. A positive genetic test result indicates that the individual carries the genetic mutation in their DNA. It is important to note that a positive test result may indicate the risk of developing the disease, but it is still possible that the individual may never actually develop the disease. If the individual has already developed FTD and the diagnostic test is being used as a means of diagnosis, it would be able to confirm that the individual has FTD, but it would provide limited information on how the disease would affect the individual.[32]

In order to enhance diagnosis results, it is important to take a step further than identifying the genetic mutation. There have been several cases where the genetic mutation in a gene has been established, but the age of onset and symptoms can vary. This becomes a more prevalent problem when several diseases share the same genetic mutations. Although genetic testing is an accurate diagnostic test, it is crucial to integrate other biomarkers and tests when intending to diagnose a patient and ensure the individual gets access to proper treatment.

Modifier Genes:

Modifier genes are defined as genetic loci that influence the expression of a primary disease-causing mutation, affecting parameters of a disease such as age at onset, severity of disease, and the duration of a disease.[33] This biomarker holds significant relevance for neurodegenerative diseases, as its integration into existing genetic testing could potentially enhance the ability to predict an individual's disease progression. Modifier genes have been identified in FTD, causing mutated genes which include *C9orf72*, *GRN*, and *MAPT*.

These common FTD-linked mutated genes demonstrate notable variability when it comes to expression in the individual. This variability can be attributed to genetic modifiers. For example, this can be observed in the *GRN* gene, which produces proteins called progranulin—a critical protein for the survival of nerve cells in the brain.[34] The FTD-causing gene mutation that occurs in this gene is known as a haploinsufficiency mutation, which suggests that an insufficiency of progranulin protein leads to neurodegeneration. Although the haploinsufficiency mutation itself drives degeneration in the brain, gene modifiers such as *SORT1* significantly impact the onset and severity of FTD, as variation near *SORT1* affects the production of progranulin. Modifier genes have also been identified for *C9orf72*, which has been mechanically linked

to *TMEM106B*—another modifier linked to progranulin, which influences lysosomal size and number. A specific variant of *TMEM106B—rs1990622 G* allele has been found to accelerate FTD's onset in *C9orf72* carriers. In contrast, the same gene modifier has also been found to affect *GRN*, but instead of accelerating the disease onset, the *rs1990622 G* allele is able to delay the disease.[35] There is not yet an identified gene modifier for the *MAPT* gene, like there is for *GRN* or *C9orf72* genes. Although there has been research done that found *TMEM106B* may have a link with tauopathies, which may suggest that *TMEM106B* could potentially be linked to *MAPT* as a modifier gene.[36]

Modifier genes clearly influence how FTD develops, which reinforces why it is essential to consider integrating the detection of modifier genes into genetic screening. Currently, there are established tests able to examine vast amounts of genomic data, including whole genome sequencing (WGS) and genome-wide association studies (GWAS). While researchers have been able to confirm that these tests can be utilized to detect modifier genes, they are typically used in research settings rather than as a diagnostic test.[37] Although modifier genes are more difficult to identify—since they do not directly cause the disease—it is still a promising emerging diagnostic test with potential to enhance diagnostic results as the field continues to evolve.

Neuroimaging:

Accurate diagnosis of FTD heavily relies on neuroimaging techniques that are able to identify patterns in neurodegeneration. Several neuroimaging techniques are utilized in diagnosis, one of the most widely used being magnetic resonance imaging (MRI). MRI scans—specifically structural MRIs—are commonly used due to their ability to analyze and present brain atrophy, which can potentially be used to recognize atrophy patterns and link those patterns to a disease.[38] For example, bvFTD typically exhibits atrophy on the frontotemporal lobes, with the right hemisphere often displaying more extensive atrophy than the left side.[39] For more specific results, clinicians often rely on volumetric MRIs—a type of structural MRI—which provide an analysis of the volume of a particular brain region. To analyze the volumes, the scans taken are compared to a database of healthy brains of individuals with similar age and sex to pinpoint changes in volume. This is an effective tool for monitoring neurodegenerative diseases and potentially detecting minor changes in brain volume that could otherwise be missed.[40]

Additionally, fluorodeoxyglucose positron emission tomography (FDG-PET) serves as a neuroimaging technique used in FTD diagnosis, by characterizing patterns of neurodegenerative and hypometabolism.[41] It is generally used when an MRI scan is inconclusive to make a clearer distinction by increasing sensitivity/specificity or providing more information on whether the disease is or is not present in an individual. In the early stages of dementia, FDG-PET outperforms MRIs in sensitivity, while MRIs outperform FDG-PET in specificity.

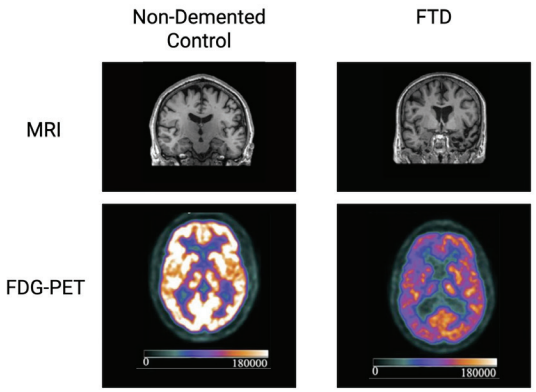


Figure 4: Comparison of MRI and FDG-PET Scans for Non-Demented Control and FTD Brain. The images captured using an MRI scan demonstrate how a brain looks after it has experienced atrophy due to FTD, while the images captured using an FDG-PET scan show the metabolic levels between the non-demented control and the individual with FTD. The color scale located at the bottom of the FDG-PET images indicates that the colors closer to 180000 convey a higher metabolic activity, whereas the colors closer to 0 imply a lower metabolic activity. Adapted from Chouliaras and O'Brien, and Ward *et al.* Created using Biorender.

Figure 4 compares the MRI scans from a FTD patient and a non-demented control, adapted from Figure 1 in Chouliaras and O'Brien's study, which explores how advanced brain imaging methods, such as MRI, can improve early detection and classification of dementia types.[42] These images highlight atrophy patterns that may aid in identifying or diagnosing FTD. Additionally, Figure 4 includes FDG-PET scans of a FTD patient and a control, adapted from Figure 3 in Ward *et al.*, providing a side-by-side comparison of two neuroimaging techniques.[39] This comparison demonstrates how combining biomarkers such as brain atrophy and metabolic changes may contribute to higher diagnostic accuracy.

However, it is important to note the risk modifiers that may alter findings of FDG-PET scans, such as traumatic brain injuries, diabetes mellitus, autoimmune disorders, and a family history of FTD. Studies have found that these diseases and illnesses may cause asymmetrical hypometabolism patterns. Therefore, it is crucial to take into account the individual's history of risk modifiers or other prior diseases or illnesses.[39]

Multimodal Imaging:

Structural MRIs along with FDG-PET scans are widely used and an established diagnostic tool for neurodegenerative diseases, including FTD. Combining the abilities of both neuroimaging techniques has shown significant improvements in diagnosis, as it provides a greater scale of biomarkers to examine when determining what neurodegenerative disease is being represented.[43] It is given that every scan provides a limited view of the brain: multimodal imaging is a technique that is able to overcome this gap by merging the strengths of multiple scans into one. This approach was demonstrated in a recent study using integrated PET/MR imaging to assess patients with FTD. Structural MRI was used to measure gray matter volume (GMV), while FDG-PET provided data on glucose metabolism, which was expressed as standardized uptake value ratios (SUVR). By acquiring both scans simultaneously, re-

DOI: 10.36838/IJHSR812.36

searchers were able to ensure precise spatial alignment, allowing for direct comparison across identical brain regions. Quantitative metrics, which refer to the measurable data extracted from each type of brain scan, were analyzed using voxel-wise analysis and region-of-interest (ROI) methods. These analysis tools essentially allowed the researchers to examine the brain in small 3D units called voxels, compare each voxel across subjects in order to detect subtle differences, and then zoom out to focus on larger name-based areas of the brain, such as the frontal lobe. Lastly, to verify anatomical consistency across scans and subjects, the study used the Automated Anatomical Labeling (AAL) atlas. This functioned as a detailed map of the brain, which divided the brain into named regions, each with clear boundaries. Multimodaling allows for a detailed analysis of the combination of data acquired from both scans.[44] This method is not flawless, but there are promising emerging fields that investigate how to further close the gaps of diagnosis, such as the use of artificial intelligence to enhance the accuracy of multimodal.

Artificial Intelligence:

Recent research by Wahul *et al.* demonstrates that fusion of neuroimaging driven by artificial intelligence significantly improves early detection and subtype classification of neurodegenerative diseases such as AD and FTD.[45] The review highlights how multimodal outperforms unimodal, combined with the use of deep learning frameworks, achieving diagnostic accuracies exceeding 90%. For example, when an electroencephalogram (EEG) machine learning data model was used, it reported an 81.1% accuracy in differentiating FTD from AD. Another aspect to consider when implementing artificial intelligence is its subset, machine learning. A 2025 study investigated how AI-aided automated brain volume analysis could enhance clinical assessment of FTD, including the classification of specific subtypes within FTD. The researchers started by gathering 104 areas of patients' normalized regional brain volumes using VUNO-Med DeepBrain, an AI-based brain volumetrics software.[46] Afterwards, a comparative analysis was completed to explore the different distributions in regional volume to identify FTD subtypes based on atrophy patterns. The researchers then input the brain volume measurements for healthy controls and FTD into an algorithm called XGBoost. This algorithm was used to distinguish between the subtypes of FTD. This algorithm resulted in significant differences in atrophy patterns to be observed for subtypes such as the behavioral and semantic variants. The classification performance for XGBoost yielded an overall 95% confidence interval in accuracy, with the lowest percentage of accuracy for differentiating between subtypes and normal control at 79-81%.[47] Moreover, recent multimodal neuroimaging research demonstrates that AI-enhanced fusion of MRI, PET, and cognitive datasets yields substantially higher diagnostic precision for differentiating AD and FTD, with deep-learning-based cortical mapping outperforming traditional statistical pipelines across multiple cohorts.[48] Complementary evidence from a 2024 systematic review further confirms that machine learning architectures—particularly those integrating structural, functional, and clinical biomarkers—consistently elevate early-stage classification accuracy for neurodegenerative disorders, reinforcing the clinical utility of AI-driven diagnostic frameworks.[49] These results suggest that the usage of artificial intelligence could be beneficial in not just distinguishing subtypes but also between neurodegenerative diseases, including FTD and AD.

Fluid Biomarkers:

Cerebrospinal fluid (CSF) is a clear fluid that surrounds the brain and spinal cord, resulting in one of the most effective fluid biomarkers for detecting biochemical changes in brain tissue.[50] This fluid is collected through lumbar puncture, also referred to as a spinal tap. The procedure consists of a spinal needle injection between a patient's vertebrae, passing through the skin and tissue.[51] After collecting samples of a patient's CSF, it is sent to a lab to examine the proteins found in the fluid. These proteins are present in the CSF due to axonal damage or cell death, which causes the release of proteins from the neuron into the surrounding cerebrospinal fluid.[52] Depending on the neurodegenerative disorder, this protein and its concentration in the fluid may vary. In FTD cases, proteins such as amyloid-beta 42 (Aß42), t-tau, which serves as a general marker for neuronal damage, and p-tau, a more specific marker of tau pathology, have been widely studied. Researchers have used these proteins to differentiate FTD and AD, as CSF Aß42 levels were typically found to be lower, whereas t-tau/Aß42 and p-tau/Aß42 ratios were higher in patients with AD compared with those with FTD. However, while these markers provide valuable insight into neuronal damage, there are indicators that they are unreliable for FTD diagnosis.[53]

In addition to Aß42 proteins, neurofilament light chain (NfL) proteins have been proven to serve as a potential biomarker for FTD. When these proteins were researched in relation to each other and compared to determine the overall accuracy or AUC for differentiating FTD from AD, Nfl alone reported 0.67 AUC, t-tau at 0.89 AUC, and t-tau/NfL ratio with an AUC of 0.95. Despite the low AUC for Nfl alone, it was noted that the Nfl proteins were able to accurately identify neurodegenerative processes that were the underlying causes of cognitive symptoms. The results of these two studies indicate that Nfl, Aß42, and tau pathology are not reliable biomarkers when considered individually. When combined, however, they improve the accuracy of differentiating FTD from AD, primarily because they reflect the general neurodegenerative process. Since their utility lies in identifying neurodegeneration broadly rather than disease-specific mechanisms, they are better suited for differentiation than for positively diagnosing FTD, as demonstrated by Giuffrè *et al.*[54]

Hok-A-Hin *et al.* investigate this gap in diagnosis by analyzing the CSF proteome on a large scale to identify a biomarker that can be utilized specifically for FTD diagnosis. The study included CSF samples from patients with FTD, AD, and cognitively normal individuals.[55] A total of 665 CSF proteins were analyzed using proximity extension assays (PEA) and data from four cohort groups, which led to the development of custom multiplex panels with significant diagnostic capabilities. When a classification analysis followed by internal

cross-validation was performed to discriminate FTD from healthy controls, this resulted in the creation of two different types of panels: the first being a 14-protein panel that distinguishes FTD from healthy controls with a 0.96 AUC and a 13-protein panel that distinguishes FTD from AD with a 0.91 AUC, which was a subset of the 14-protein panel. These panels included proteins involved in inflammation, oxidative stress, and synaptic remodeling. Furthermore, the panels were able to outperform or match NfL, which was used as a benchmark biomarker, in identifying FTD itself, with the 14-protein panel having an AUC of 0.96 and the 13-protein panel with an AUC of 0.91. Although NfL was confirmed to be a stronger marker of general neurodegeneration, especially useful for distinguishing FTD from controls, there are still limitations, and the multi-protein panels offer a greater specificity.

Biomarker/Panel	Biomarker Type	Overall Accuracy (AUC) for Differentiating FTD from AD	Role in Differentiation
NfL	Single biomarker	0.67	Limited ability to distinguish FTD from AD
t-tau	Single biomarker	0.89	Reflects neuronal injury; provides strong differentiation between FTD and AD
t-tau/NfL	Combined ratio	0.95	Stronger differentiation when biomarkers are combined
13-protein panel	Multi-protein panel	0.91	High specificity for FTD vs AD
14-protein panel	Multi-protein panel	0.96	Strongest differentiation; separates FTD from controls

Figure 5: Overall Accuracy (AUC) of CSF Biomarkers and Multi-Protein Panels for Differentiating FTD and AD. This table presents the AUC values reported for individual CSF biomarkers, combined biomarkers, and multi-protein panels in distinguishing FTD from AD. It highlights the relative diagnostic performance of each biomarker, ranging from single-protein measures to multiplex panels with higher specificity. Information gathered from Giuffrè *et al.* and Hok-A-Hin *et al.* Created using Biorender.

While Aß42, t-tau, and p-tau are prominent biomarkers for AD, their inclusion in FTD diagnosis has been proven to provide valuable insight into differential diagnosis.[55] When examining NfL, there is strong evidence that it has the potential to serve as a sensitive marker of axonal damage and general neurodegeneration, as it results in elevated NfL levels. However, there are still noticeable limitations in each protein: this is where multi-protein panels could potentially be used in order to combine the benefits of multiple proteins. In order for the panel to be effective in both differentiating and diagnosing, it is crucial to include proteins that are specific to FTD pathology. Based on current research, the ideal multi-protein panel for FTD would include proteins involved in inflammation, oxidative stress, synaptic remodeling, neurotransmission, NfL, and t-tau. In current published research, panels that contain these proteins have achieved AUC values ranging from 0.91 to 0.96 in distinguishing FTD from AD and provide a wealth of information about how FTD affects a certain individual's brain biology, disease progression, and underlying pathology.[55] Incorporating such panels will not only improve the accuracy of FTD detection but also facilitate earlier diagnosis, potentially enabling individuals to access supportive care sooner and qualify for emerging clinical trials aiming to evaluate targeted therapies.

Conclusion

After analyzing the most effective diagnostic tools for FTD, it remains clear that FTD diagnosis is a complex challenge due to its heterogeneity and overlap with other neurodegenerative disorders. Current research and literature highlight the importance of implementing multiple diagnostic tools, such as genetic testing, neuroimaging, and body fluid biomarkers, to improve accuracy and potentially enable earlier detection. Genetic testing is highly reliable as it provides a specificity in identifying pathogenic variants, specifically in *C9orf72*, *GRN*, and *MAPT* genes, while modifier genes such as *TMEM106B* and *SORT1* offer greater insight into an individual's variability in disease onset and progression. Neuroimaging techniques, including structural and volumetric MRI, as well as FDG-PET scans, reveal brain atrophy and hypermetabolism patterns. These patterns can be more effectively interpreted through multimodal fusion frameworks enhanced by artificial intelligence, which allow for deeper interpretation of brain scans, improving diagnostic accuracy. Meanwhile, cerebrospinal fluid biomarkers such as NfL, t-tau, and proteins involved in inflammation, oxidative stress, synaptic remodeling, and neurotransmission contribute to differential diagnosis. Regarding the limitations of the proteins found in cerebrospinal fluid to function as a single marker for identifying FTD itself, it highlights how multi-protein panels may be the solution, as these panels provide a wide range of proteins that are able to measure a variety of FTD characteristics, exceeding a diagnostic accuracy of 90%. Together, these different diagnostic tools and approaches reveal a shift towards a more precise diagnosis of FTD, with the potential to lead to earlier disease intervention and possible personalized care.

Several emerging fields of diagnosis further expand on the diagnostic methods mentioned, but the application and implementation of these tools in diagnosis are crucial to improving FTD diagnosis. Adopting modifier genes into genetic screening, multimodal imaging into neuroimaging, and multi-protein panels into cerebrospinal fluid testing can all be beneficial in increasing diagnostic accuracy. Although these tools help cover gaps in diagnosis, they still have their limitations, which is why it is pivotal to continue identifying proteins or biomarkers that are FTD-specific. Further research into genes, brain atrophy and hypometabolism patterns, and proteins, which are the underlying causes of neurodegeneration, can substantially improve the identification and diagnosis of FTD.

Acknowledgments

I am deeply grateful to Professor Jobin Varkey for his expert guidance in neuroscience and to Professor Virgel Torremocha for his creative insight and unwavering support. Their encouragement and mentorship not only guided my research but also inspired me to explore the field of neuroscience more deeply.

References

1. 1. National Institute on Aging. *What are frontotemporal disorders? Causes, symptoms, and treatment.* https://www.nia.nih.gov/health/frontotemporal-disorders/what-are-frontotemporal-dis-

DOI: 10.36838/IJHSR812.36

orders-causes-symptoms-and-treatment (updated Jan 22, 2025; accessed July 2025).
2. Flavell, J.; Ahern, E. G. M.; Logan, B.; Shaw, T. B.; Adam, R. J.; McElligott, C. A. T.; Nestor, P. J. Factors associated with true-positive and false-positive diagnoses of behavioural variant frontotemporal dementia in 100 consecutive referrals from specialist physicians. *Eur. J. Neurol*. **2025**, *32*(1). https://doi.org/10.1111/ene.70036.
3. Weiler, N. *Lifestyle Choices Could Slow Familial Frontotemporal Dementia.* https://www.ucsf.edu/news/2020/01/416391/lifestyle-choices-could-slow-familial-frontotemporal-dementia (published Jan 7, 2020; accessed July 2025).
4. Beber, B. C.; Chaves, M. L. F. Evaluation of patients with behavioral and cognitive complaints: misdiagnosis in frontotemporal dementia and Alzheimer's disease. *Dement. Neuropsychol.* **2013**, *7*(1), 60—65. https://doi.org/10.1590/S1980-57642013DN70100010.
5. Ali, M.; Erabadda, B.; Chen, Y.; Xu, Y.; Gong, K.; Liu, M.; Binette, A. P.; Timsina, J.; Western, D.; Yang, C.; Heo, G.; Vogel, J. W.; Tijms, B. M.; Krish, V.; Imam, F.; Hansson, O.; Winchester, L.; Cruchaga, C. Shared and Disease-Specific Pathways in Frontotemporal Dementia and Alzheimer's and Parkinson's Diseases. *Nat. Med*. **2025**, *31*, 2567-2577. https://doi.org/10.1038/s41591-025-03833-1.
6. National Institute of Neurological Disorders and Stroke. *Frontotemporal dementia and other frontotemporal disorders.* https://www.ninds.nih.gov/health-information/disorders/frontotemporal-dementia-and-other-frontotemporal-disorders#toc-introduction-to-frontotemporal-dementia (accessed Jul 2025).
7. National Institute on Aging. *What is frontotemporal dementia?* https://www.alzheimers.gov/alzheimers-dementias/frontotemporal-dementia (updated Aug 27, 2025; accessed July 2025).
8. Houck, S. A.; Singh, S.; Cyr, D. M. Cellular responses to misfolded proteins and protein aggregates. In *Ubiquitin Family Modifiers and the Proteasome*; Dohmen, R., Scheffner, M., Eds.; *Methods in Molecular Biology*; Humana Press: Totowa, NJ, **2012**; Vol. *832*, pp 455—461. https://doi.org/10.1007/978-1-61779-474-2_32.
9. Takalo, M.; Salminen, A.; Soininen, H.; Hiltunen, M.; Haapasalo, A. Protein Aggregation and Degradation Mechanisms in Neurodegenerative Diseases. *Am. J. Neurodegener. Dis*. **2013**, *2*(1), 1. https://pmc.ncbi.nlm.nih.gov/articles/PMC3601466/.
10. Samudra, N.; Lane-Donovan, C.; VandeVrede, L.; Boxer, A. L. Tau Pathology in Neurodegenerative Disease: Disease Mechanisms and Therapeutic Avenues. *J. Clin. Invest*. **2023**, *133*(12), e168553. https://doi.org/10.1172/JCI168553.
11. Ducharme, S.; Yolande A.L. Pijnenburg; Rohrer, J. D.; Huey, E. D.; Finger, E.; Tatton, N. Identifying and Diagnosing TDP-43 Neurodegenerative Diseases in Psychiatry. *Am. J. Geriatr. Psychiatry* **2024**, *32*(1), 98—113. https://doi.org/10.1016/j.jagp.2023.08.017.
12. Greaves, C. V.; Rohrer, J. D. An Update on Genetic Frontotemporal Dementia. *J. Neurol*. **2019**, *266*(8), 2075—2086. https://doi.org/10.1007/s00415-019-09363-4.
13. Chen, X.; Chen, Y.; Ni, B.; Huang, C. Research Trends and Hotspots for Frontotemporal Dementia from 2000 to 2022: A Bibliometric Analysis. *Front. Neurol*. **2024**, *15*. https://doi.org/10.3389/fneur.2024.1399600.
14. Fast Facts about Frontotemporal Degeneration. https://www.theaftd.org/wp-content/uploads/2009/02/Fast-Facts-Final-6-11.pdf (published 2011; accessed Jul 2025)
15. Guven, G.; Lohmann, E.; Bras, J.; Gibbs, J. R.; Gurvit, H.; Bilgic, B.; Hanagasi, H.; Rizzu, P.; Heutink, P.; Emre, M.; Erginel-Unaltuna, N.; Just, W.; Hardy, J.; Singleton, A.; Guerreiro, R. Mutation Frequency of the Major Frontotemporal Dementia Genes, MAPT, GRN and C9ORF72 in a Turkish Cohort of Dementia Patients. *PLoS ONE* **2016**, *11*(9), e0162592. https://doi.org/10.1371/journal.pone.0162592.
16. Cummings, J. L. Neuropsychiatric Inventory (NPI) Comprehensive Assessment of Psychopathology in Patients with Dementia. https://www.psychdb.com/_media/geri/dementia/npi-original.pdf (published 1994; accessed Oct 2025)
17. Santoro Bahia, V.; Viana, R. Accuracy of Neuropsychological Tests and the Neuropsychiatric Inventory in Differential Diagnosis between Frontotemporal Dementia and Alzheimer's Disease. *Dement. Neuropsychol.* **2009**, *3*(4), 332—336. https://www.demneuropsy.org/wp-content/uploads/2023/06/v3n4a12-ing.pdf.
18. Wisconsin Alzheimer's Institute. Zarit Burden Interview Assessing Caregiver Burden; 2021. https://wai.wisc.edu/wp-content/uploads/sites/1129/2021/11/Zarit-Caregiver-Burden-Assessment-Instruments.pdf.
19. Liu, S.; Liu, J.; Wang, X.-D.; Shi, Z.; Zhou, Y.; Li, J.; Yu, T.; Ji, Y. Caregiver Burden, Sleep Quality, Depression, and Anxiety in Dementia Caregivers: A Comparison of Frontotemporal Lobar Degeneration, Dementia with Lewy Bodies, and Alzheimer's Disease. *Int. Psychogeriatr*. **2017**, *30*(8), 1131—1138. https://doi.org/10.1017/s1041610217002630.
20. Galvin, J. E.; Howard, D. H.; Denny, S. S.; Dickinson, S.; Tatton, N. The Social and Economic Burden of Frontotemporal Degeneration. *Neurology* **2017**, *89*(20), 2049—2056. https://doi.org/10.1212/wnl.0000000000004614.
21. Nandi, A.; Counts, N.; Bröker, J.; Malik, S.; Chen, S.; Han, R.; Klusty, J.; Seligman, B.; Tortorice, D.; Vigo, D.; Bloom, D. E. Cost of Care for Alzheimer's Disease and Related Dementias in the United States: 2016 to 2060. *npj Aging* ***2024***, *10*(1), 1—8. https://doi.org/10.1038/s41514-024-00136-6.
22. Loi, S. M.; Cations, M.; Velakoulis, D. Young-Onset Dementia Diagnosis, Management and Care: A Narrative Review. *Med. Jo. Aust*. **2023**, *218*(4), 182—189. https://doi.org/10.5694/mja2.51849.
23. Mollah, S. A.; Nayak, A.; Barhai, S.; Maity, U. A Comprehensive Review on Frontotemporal Dementia: Its Impact on Language, Speech and Behavior. *Dement. Neuropsychol*. **2024**, *18*, e20230072. https://doi.org/10.1590/1980-5764-DN-2023-0072.
24. Roberts, J. S.; Patterson, A. K.; Uhlmann, W. R. Genetic Testing for Neurodegenerative Diseases: Ethical and Health Communication Challenges. *Neurobio. Dis*. **2020**, *141*, 104871. https://doi.org/10.1016/j.nbd.2020.104871.
25. Genetic Testing and Counseling in FTD - FTD Disorders Registry. *FTD Disorders Registry*. https://ftdregistry.org/press/genetic-testing-and-counseling-in-ftd/ (accessed August 2025).
26. MedlinePlus. How is genetic testing done? medlineplus.gov. https://medlineplus.gov/genetics/understanding/testing/procedure/ (accessed August 2025).
27. Sirkis, D. W.; Geier, E. G.; Bonham, L. W.; Karch, C. M.; Yokoyama, J. S. Recent Advances in the Genetics of Frontotemporal Dementia. *Curr. Genet. Med. Rep*. **2019**, *7*(1), 41—52. https://doi.org/10.1007/s40142-019-0160-6.
28. Zampatti, S.; Peconi, C.; Campopiano, R.; Gambardella, S.; Caltagirone, C.; Giardina, E. C9orf72-Related Neurodegenerative Diseases: From Clinical Diagnosis to Therapeutic Strategies. *Front. Aging Neurosci*. **2022**, *14*. https://doi.org/10.3389/fnagi.2022.907122.
29. Gossye, H.; Engelborghs, S.; Van Broeckhoven, C.; van der Zee, J. C9orf72-Related Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. *PubMed*, December 17, **2020**. https://www.ncbi.nlm.nih.gov/books/NBK268647/.
30. Hsiung, G.-Y. R.; Feldman, H. H. GRN Frontotemporal Dementia. *PubMed*, February 6, **2020**. https://www.ncbi.nlm.nih.gov/books/NBK1371/.

31. Rohrer, J.; Ryan, B.; Ahmed, R. MAPT-Related Frontotemporal Dementia. *Nih.gov*, August 18, **2022**. https://www.ncbi.nlm.nih.gov/books/NBK1505/#ftdp-17.Penetrance.
32. Gossye, H.; Van Broeckhoven, C.; Engelborghs, S. The Use of Biomarkers and Genetic Screening to Diagnose Frontotemporal Dementia: Evidence and Clinical Implications. *Front. Neurosci*. **2019**, *13*. https://doi.org/10.3389/fnins.2019.00757.
33. Lamar, K.-M.; McNally, E. M. Genetic Modifiers for Neuromuscular Diseases. *J. Neuromuscul. Dis*. **2014**, *1*(1), 3—13. https://doi.org/10.3233/jnd-140023.
34. GRN gene: MedlinePlus Genetics. medlineplus.gov. https://medlineplus.gov/genetics/gene/grn/ (accessed August 2025)
35. Jain, N.; Chen-Plotkin, A. S. Genetic Modifiers in Neurodegeneration. *Curr. Genet. Med. Rep.* **2018**, *6*(1), 11—19. https://doi.org/10.1007/s40142-018-0133-1.
36. Edwards, G. A.; Wood, C. A.; He, Y.; Nguyen, Q.; Kim, P. J.; Gomez-Gutierrez, R.; Park, K.-W.; Xu, Y.; Zurhellen, C.; Al-Ramahi, I.; Jankowsky, J. L. TMEM106B Coding Variant Is Protective and Deletion Detrimental in a Mouse Model of Tauopathy. *Acta. Neuropathol*. **2024**, *147*(1), 61. https://doi.org/10.1007/s00401-024-02701-5.
37. Ciani, M.; Benussi, L.; Bonvicini, C.; Ghidoni, R. Genome Wide Association Study and Next Generation Sequencing: A Glimmer of Light toward New Possible Horizons in Frontotemporal Dementia Research. Front. Neurosci. **2019**, *13*. https://doi.org/10.3389/fnins.2019.00506.
38. Huang, P.; Zhang, M. Magnetic Resonance Imaging Studies of Neurodegenerative Disease: From Methods to Translational Research. *Neurosci. Bull.* **2022**, *39*(1), 99—112. https://doi.org/10.1007/s12264-022-00905-x.
39. Ward, J.; Ly, M.; Raji, C. A. Brain PET Imaging. *PET Clin*. **2023**, *18*(1), 123—133. https://pmc.ncbi.nlm.nih.gov/articles/PMC9884902/
40. Meysami, S.; Raji, C. A.; Mendez, M. F. Quantified Brain Magnetic Resonance Imaging Volumes Differentiates Behavioral Variant Frontotemporal Dementia from Early-Onset Alzheimer's Disease. *J. Alzheimer's Dis*. **2022**, *89*(1), 1—9. https://doi.org/10.3233/jad-215667.
41. Whitwell, J. L. Neuroimaging across the FTD Spectrum. *Prog. Mol. Biol. Transl. Sci*. **2019**, *165*, 187–223. https://doi.org/10.1016/bs.pmbts.2019.05.009.
42. Chouliaras, L.; O'Brien, J. T. The Use of Neuroimaging Techniques in the Early and Differential Diagnosis of Dementia. *Mol. Psychiatry* **2023**, *28*. https://doi.org/10.1038/s41380-023-02215-8
43. Radder, N.; Sonar, S.; Nanivadekar, A.; Radder, S. Synergy in Neuroimaging: PET-CT and MRI Fusion for Enhanced Characterization of Brain Pathology. *Cureus* **2024**, *16*, https://doi.org/10.7759/cureus.74353.
44. Bi, S.; Chen, Z.; Tao, W.; Yan, S.; Cui, B.; Yang, H.; Lu, J. Multimodal Imaging for Diagnosis of Frontotemporal Dementia using Integrated PET/MR. *J. Nucl. Med*. **2024**, *65*(Suppl. 2), 241923. https://jnm.snmjournals.org/content/65/supplement_2/241923.
45. Wahul, R. M.; Ambadekar, S.; Dhanvijay, D. M.; Dhanvijay, M. M.; Dudhedia, M. A.; Varsha Gaikwad; Bhavana Kanawade; Pansare, J. R.; Balaji Bodkhe; Gawande, S. H. Multimodal Approaches and AI-Driven Innovations in Dementia Diagnosis: A Systematic Review. *Discov. Artif. Intell*. **2025**, *5*(1). https://doi.org/10.1007/s44163-025-00358-x.
46. US Health Connect. AI-Based Brain Volumetric Software, VUNO Med-DeepBrain, Receives FDA 510(k) Clearance. *Practical Neurology*, October 31, **2023**. https://practicalneurology.com/news/ai-based-brain-volumetric-software-vuno-med-deepbrain-receives-fda-510k-clearance/2470328/.
47. Cho, H.; Park, M.; Lee, S. H.; Jung, W.; Kim, D.; Kim, J.; Lee, Y. Automated Brain Volumetry Analysis for Differential Diagnosis of Frontotemporal Dementia Subtypes. *Alzheimer's Dement*. **2024**, *20*(S2). https://doi.org/10.1002/alz.086728.
48. Pérez-Millan, A.; Contador, J.; Juncà-Parella, J.; Bosch, B.; Borrell, L.; Tort-Merino, A.; Falgàs, N.; Borrego-Écija, S.; Bargalló, N.; Rami, L.; Balasa, M.; Lladó, A.; Sánchez-Valle, R.; Sala-Llonch, R. Classifying Alzheimer's disease and frontotemporal dementia using machine learning with cross-sectional and longitudinal magnetic resonance imaging data. *Hum. Brain Mapp*. 2023, *44*, 26205. https://doi.org/10.1002/hbm.26205.
49. Ranjan, S.; Tripathi, A.; Shende, H.; Badal, R.; Kumar, A.; Yadav, P.; Joshi, D.; Kumar, L. Deep learning-based classification of dementia using image representation of subcortical signals. *BMC Med. Inform. Decis. Mak.* 2025, *25* (1). https://doi.org/10.1186/s12911-025-02924-w.
50. Swift, I. J.; Sogorb-Esteve, A.; Heller, C.; Synofzik, M.; Otto, M.; Graff, C.; Galimberti, D.; Todd, E.; Heslegrave, A. J.; van der Ende, E. L.; Van Swieten, J. C.; Zetterberg, H.; Rohrer, J. D. Fluid Biomarkers in Frontotemporal Dementia: Past, Present and Future. *J. Neurol. Neurosurg. Psychiatry* **2020**, *92*(2), 204—215. https://doi.org/10.1136/jnnp-2020-323520.
51. Jane, L. A.; Wray, A. A. Lumbar Puncture. *PubMed*, June 24, **2023**. https://www.ncbi.nlm.nih.gov/books/NBK557553/.
52. Machacek, M.; Garcia-Montoya, E.; McColgan, P.; Sanwald-Ducray, P.; Mazer, N. A. NfL Concentration in CSF Is a Quantitative Marker of the Rate of Neurodegeneration in Aging and Huntington's Disease: A Semi-Mechanistic Model-Based Analysis. *Front. Neurosci*. **2024**, *18*. https://doi.org/10.3389/fnins.2024.1420198.
53. Casoli, T.; Paolini, S.; Fabbietti, P.; Fattoretti, P.; Paciaroni, L.; Fabi, K.; Gobbi, B.; Galeazzi, R.; Rossi, R.; Lattanzio, F.; Pelliccioni, G. Cerebrospinal Fluid Biomarkers and Cognitive Status in Differential Diagnosis of Frontotemporal Dementia and Alzheimer's Disease. *J. Int. Med. Res*. **2019**, *47*(10), 4968—4980. https://doi.org/10.1177/0300060519860951.
54. Giuffrè, G. M.; Quaranta, D.; Costantini, E. M.; Citro, S.; Martellacci, N.; De Ninno, G.; Vita, M. G.; Guglielmi, V.; Rossini, P. M.; Calabresi, P.; Marra, C. Cerebrospinal Fluid Neurofilament Light Chain and Total-Tau as Biomarkers of Neurodegeneration in Alzheimer's Disease and Frontotemporal Dementia. *Neurobiol. Dis*. **2023**, *186*, 106267. https://doi.org/10.1016/j.nbd.2023.106267.
55. Hok-A-Hin, Y. S.; Vermunt, L.; Peeters, C. F. W.; Emma; de, C. M.; Meeter, L. H.; Houwer, J. de; Harro Seelaar; John; Hu, W. T.; Lleó, A.; Alcolea, D.; Sebastiaan Engelborghs; Sieben, A.; Chen-Plotkin, A.; Irwin, D. J.; van; Pijnenburg, Y. A. L.; Teunissen, C. E.; Campo, M. del. Large-Scale CSF Proteome Profiling Identifies Biomarkers for Accurate Diagnosis of Frontotemporal Dementia. *Mol. Neurodegener*. **2025**, *20*(1). https://doi.org/10.1186/s13024-025-00882-5.

Author

Anusha Manda is a 10th-grade student at Dripping Springs High School in Dripping Springs, Texas. She is passionate about neuroscience and intends to pursue it as her university major. Anusha is particularly interested in exploring the complexities of the human brain and plans to continue researching various domains of neuroscience.

DOI: 10.36838/IJHSR812.36

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

The Impact of News Sentiment and the COVID-19 Pandemic on Bitcoin Volatility, Return, and Liquidity

Elif Yıldız

Robert College, Arnavutköy, Kuruçeşme St. No:87, Beşiktaş, Istanbul, 34345, Turkey; elf.yldz208@gmail.com

ABSTRACT: The purpose of this paper is to investigate the effects of both macroeconomic, political, and social news, as well as the COVID-19 pandemic, on Bitcoin's volatility, return, and liquidity using high-frequency data from January 1, 2018, to July 17, 2025. The paper utilizes 263,946 15-minute observations to reduce microstructure noise and capture responses to different types of information. The results indicate that the COVID-19 pandemic has a significant impact on Bitcoin's volatility and liquidity, with no statistically significant effect on its average return. Many different types of news are statistically significantly related to volatility and liquidity, but not to return. These findings underscore the pivotal role of information flow in shaping cryptocurrency market behavior, offering valuable insights for investors, policymakers, and managers.

KEYWORDS: Economics, Finance, Bitcoin, Sentiment, COVID-19.

Introduction

Cryptocurrencies have gained popularity as an alternative investment tool over the past decade due to their transparency, security, and low-cost, cross-border transactions. Bitcoin, the most influential cryptocurrency, shapes investor sentiment, drives overall market trends, and has gained legitimacy through adoption by major corporations and institutional investors.[1]

Volatility is a key indicator in financial markets, playing a crucial role in assessing market risk, guiding investment strategies, and pricing financial derivatives. For traders and investors, understanding volatility is vital for effective risk management and identifying potential opportunities in diverse market conditions.[2]

Bitcoin's volatility is a defining characteristic that both attracts and challenges investors. Its sharp price fluctuations create significant opportunities for short-term traders, while also posing substantial risks for long-term holders. As the largest and most influential cryptocurrency, Bitcoin's volatility often sets the tone for the entire crypto market, influencing investor sentiment and capital flows. Additionally, its price swings reflect the asset's sensitivity to macroeconomic events, regulatory developments, and shifts in market adoption, making volatility analysis essential for understanding and navigating the cryptocurrency ecosystem.[3,5] That's why this paper focuses on the relationship between Bitcoin and news sentiment to give valuable insight into today's rapidly evolving financial world, contributing to a better understanding of market behavior and investor decision-making.

The results show that the COVID-19 pandemic significantly increased bitcoin volatility and liquidity, but had no effect on average return. Google trends data are positively related to both volatility and liquidity but not to return, suggesting that heightened investor attention increases market activity and instability without affecting price direction. News sentiment is positively associated with volatility but negatively related to liquidity, indicating that negative sentiment may trigger greater trading participation. Event-based regression indicates that political, macroeconomic, and social events have a significant effect on Bitcoin market dynamics. These findings highlight the critical role of information flow in shaping cryptocurrency market behavior.

This study contributes to the growing literature on volatility dynamics of digital asset markets by providing comprehensive high-frequency data analytics of how macroeconomic, political, and social news affect Bitcoin's key market indicators. While the prior research has often relied on daily (lower-frequency) data, this paper employs intraday data, enabling more precise characterization of short-term market reactions. Our study takes into account multiple dimensions for information flow, including Google Trends as a proxy for investor attention, news sentiment as a proxy for investor sentiment, and political, macroeconomic, and social news for broader societal attention. These findings have practical implications for traders, investors, and policymakers in understanding volatility patterns.

Literature Review

Various methodologies have been used and applied to investigate the effect of news on Bitcoin price such as regression analysis (Corbet *et al.*[1]; Sapkota *et al.*[2]), HAR-RV (Sapkota *et al.*[2]), AR-CGARCH (Vidal-Tomás and Ibañez[3]), NARDL (Zhu *et al.*[4]), NLP (Karalevicius *et al.*[5]), DCC-GARCH (Mariana *et al.*[7]), TVP-VAR (Elsayed *et al.*[10]), ANN (Arratia *et al.*[11]), Jump Diffusion Model (Philippas *et al.*[12]), Wavelet transformation (Aysan *et al.*[16]), and Quantile Regression (Naifar and Altamimi[18]).

The variables used in the literature often varied, including Bitcoin (Corbet *et al.*[1]), macroeconomic indicators (Corbet *et al.*[1]), news (Sapkota *et al.*[2]), (Zhu *et al.*[4]), S&P 500 (Conlon and McGee[6]), Stock Twits (Bouteska *et al.*[8]), VIX (Elsayed *et al.*[10]), ETH (Oraştean *et al.*[13]), oil prices fake news index (Zhang *et al.*[15]), sentiment news (Aysan *et al.*[16]), COVID-19-related

DOI: 10.36838/IJHSR812.45

variables (Derbali *et al.*[17]), and Google Trends (Aslanidis *et al.*[9]), Gold (Derbali *et al.*[17]).

The studies show that Bitcoin reacts differently depending on the type of news. Corbet *et al.*[1] found that Bitcoin responds to unemployment and durable good news but ignores GDP and inflation. Vidal-Tomás and Ibañez[3] showed that Bitcoin reacts mainly to its own events, not economic news. Sapkota *et al.*[2] and Karalevicius *et al.*[5] highlighted that media sentiment strongly influences Bitcoin volatility.

During the COVID-19 pandemic, findings were mixed. Conlon and McGee[6] stated that Bitcoin did not function as a haven. At the same time, Sapkota *et al.*[2] found that Bitcoin and Ethereum exhibited nonstationary price behavior, with their prices rising as COVID-19 cases and deaths increased. In contrast, Mariana *et al.*[7] provided evidence of Bitcoin's short-term haven properties. Zhu *et al.*[4] suggested that gold effectively hedges negative pandemic news in the short term. In addition, Zhang *et al.*[15], Derbali *et al.*[17], and Naifar and Altamimi[18] found that COVID-19 media coverage drove short-term volatility in crude oil, gold, and Bitcoin, with risk spreading mainly to gold and Bitcoin. At the same time, investor sentiment continued to show short-term Bitcoin returns but weakened during COVID-19 (Aysan *et al.*[16]).

Finally, studies using Google Trends show that search interest is related to price change, confirming the impact of investor attention (Aslanidis *et al.*[9]; Philippas *et al.*[12]). Oraştean *et al.*[13] found that increased search interest often precedes price rises, trading volume spikes, and sentiment shifts. At the same time, Arratia *et al.*[11] argue that it is not consistently reliable, yet Google Trends can still reflect demand.

Unlike prior studies relying on daily data, this paper uses high-frequency intraday data, providing a valuable perspective on news sentiment and the Bitcoin relationship.

Methods

Data:

The data consists of intraday data on the closing price of Bitcoin from 01/01/2018 to 17/07/2025. The data obtained from the Kaggle platform resulted in 263,946 observations. To reduce microstructure noise bias, we sample intra-daily returns at a 15-minute frequency.

Google Trends data taken from the Trends web page[20] as an index showing how frequently the keyword "bitcoin" has been searched from 2018 to 2025. It is used as a proxy for investor attention.

The Daily News Sentiment Index data is taken from the Federal Reserve Bank of San Francisco19 as an index showing the high-frequency measure of economic sentiment based on lexical analysis of economics-related news articles. The index is described in Buckman, Shapiro, Sudhof, and Wilson and is based on the methodology developed in Shapiro, Sudhof, and Wilson. It is used as a proxy for investor sentiment to see the effect of many different information inflows.

The political, macroeconomic, and social indices have been constructed by the present paper, based on the events listed in the Appendix. These indices are dummy variables that take the value of 1 on a news day and 0 otherwise.

Model:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

Where Y_t is the dependent variable for observation t, which refers to realized volatility, average daily return, and average daily liquidity. β_0 is the constant term, representing the expected value of the dependent variable when the independent variable is 0. β_1 is the coefficient for the independent variable, which is a binary dummy variable for the COVID-19 pandemic, which takes 1 during the COVID-19 pandemic and 0 otherwise. The coefficient shows how much the dependent variable changes when the corresponding independent variable changes by 1 unit. ε_t is the error, which represents the difference between the actual value and the predicted value from the model.

Results and Discussion

Descriptive Statistics:

Table 1a: Descriptive Statistics for Total Period. This table summarises the descriptive statistics of the total sample. The data indicate that the average returns are low and that the realised volatility is highly dispersed. This indicates that the market is sensitive to periodic shocks over the long run.

Variable	Obs	Mean	Std. Dev.	Min	Max
avgreturn2	2753	.001	.042	-.211	1.203
rvnew	2753	13.264	25.434	.044	464.287
liquidity	2753	5.999	.816	3.392	8.844
NewsSentiment	2752	-.052	.179	-.666	.3

Table 1b: Descriptive Statistics During the COVID-19 Period. This table summarises the descriptive statistics for the pandemic period. The results show that realised volatility reaches its highest average value, while News Sentiment declines to its lowest level. This pattern indicates increased market uncertainty, highlighting the significant stress experienced by financial markets during this period.

Variable	Obs	Mean	Std. Dev.	Min	Max
avgreturn2	839	.002	.04	-.174	.186
rvnew	839	17.474	28.851	.486	464.287
liquidity	839	6.264	.486	4.897	8.007
NewsSentiment	839	-.141	.225	-.666	.198

Table 1c: Descriptive Statistics During Non-COVID-19 Period. This table summarises the descriptive statistics for the non-COVID-19 period. Compared with the COVID-19 period, realised volatility declines, while News Sentiment remains close to a neutral level. This indicates a more stable market environment.

Variable	Obs	Mean	Std. Dev.	Min	Max
avgreturn2	1914	.001	.044	-.211	1.203
rvnew	1914	11.419	23.554	.044	424.254
liquidity	1914	5.883	.9	3.392	8.844
NewsSentiment	1913	-.013	.137	-.43	.3

Tables 1a, 1b, and 1c provide descriptive statistics for average return, realized volatility, liquidity, and news sentiment for the total period, the COVID-19 period, and the Non-COVID-19 period, respectively. The average return remains the same for all periods. There is a slight increase during the COVID-19 period. However, realized volatility is notably higher during the COVID-19 period, confirming the market uncertainty. Liquidity also increases during the COVID-19 period, indicating heightened trading activity and market participation. The news sentiment was negative during the COVID-19 pandemic; this confirms that the pandemic increased uncertainty and pessimism in the news.

COVID-19 Analysis:

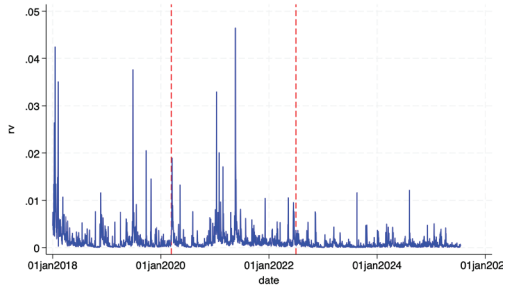


Figure 1: RV during the COVID-19 pandemic. Figure 1 shows Bitcoin's volatility before, during, and after the COVID-19 pandemic, with red dashed lines representing the beginning and end of the pandemic. Compared to non-COVID times, realised volatility increased significantly. Volatility decreases and stabilises around the end of the pandemic, indicating market normalisation.

Figure 1 shows the realized volatility of Bitcoin before, during, and after the COVID-19 period. Red dashed lines indicate the start and the end of COVID-19. The figure clearly illustrates a significant spike in realized volatility during the COVID-19 period compared to non-COVID-19 periods. Especially at the end of the pandemic, volatility remained relatively low and stable, indicating standard market conditions.

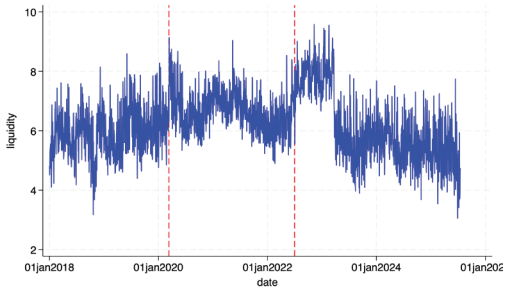


Figure 2: Average liquidity during the COVID-19 pandemic. Figure 2 shows market liquidity before, during, and after the COVID-19 pandemic. Results show that pandemic liquidity increased significantly compared to pre-COVID. Liquidity remains high following COVID-19, suggesting persistent trading activity despite market uncertainty.

Figure 2 indicates market liquidity before, during, and after the COVID-19 period. The graph indicates that market liquidity increased notably during the COVID-19 period compared to the pre-COVID period. Interestingly, liquidity remained relatively high immediately after the COVID-19 period.

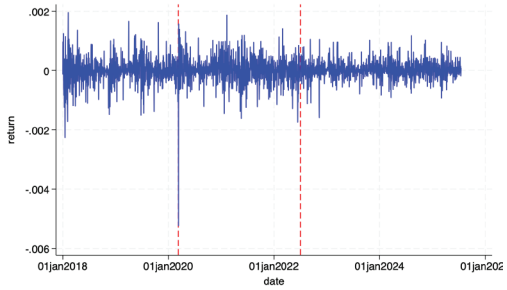
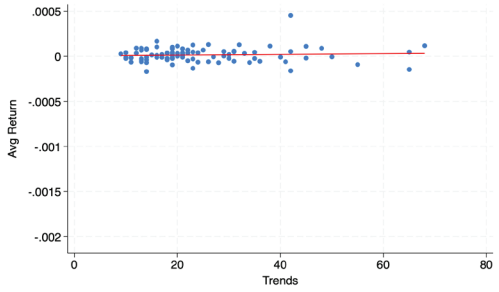


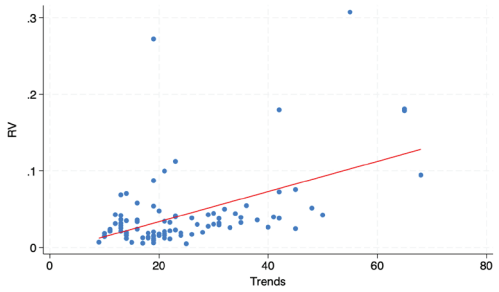
Figure 3: Average return during the COVID-19 pandemic. Figure 3 shows average returns before, during, and after COVID-19. There is no clear difference between the pre-COVID and COVID-19 periods. However, average returns decline noticeably in the post-pandemic period, indicating weaker market performance after the pandemic period.

Figure 3 displays the average return before, during, and after the COVID-19 period. Especially at the beginning of COVID-19, there was a dramatic, huge negative response. Moreover, there appears to be no significant difference between before the pandemic and during the pandemic period. However, returns have noticeably decreased in the post-pandemic period.

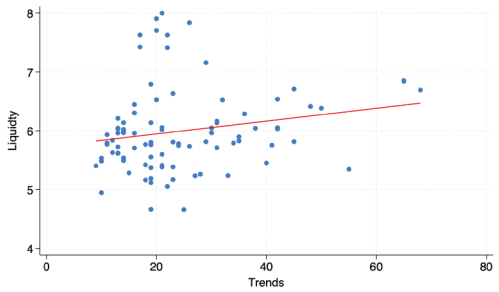
Bitcoin Trends Analysis:



Panel a: Return vs Trends.



Panel b: RV vs Trends.

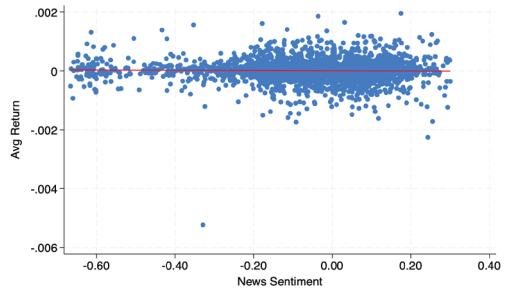


Panel c: Liquidity vs Trends.

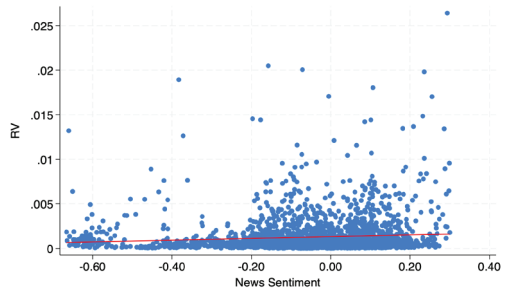
Figure 4: Relationship between average return, realized volatility, and liquidity with the Bitcoin trend. Figure 4 shows scatter plots of average return, realised volatility, and liquidity against Bitcoin trends. The results indicate an overall positive relationship, suggesting that stronger Bitcoin trends are associated with higher realised volatility and increased liquidity.

Figure 4 presents the scatter plot illustrating the relationship between average return, realized volatility, and liquidity with the Bitcoin trend. The figure shows an overall positive correlation among these variables, suggesting that Bitcoin trends have a positive effect on realized volatility and liquidity.

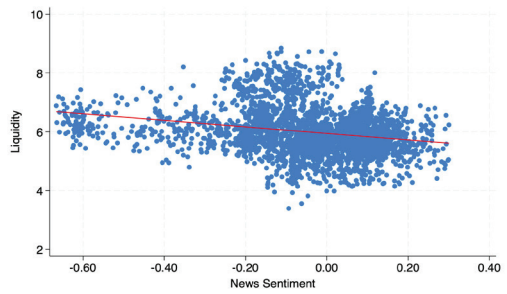
DOI: 10.36838/IJHSR812.45



Panel a: Return vs News Sentiment.



Panel b: RV vs News Sentiment.



Panel c: Liquidity vs News Sentiment.

Figure 5: Relationship between average return, realized volatility, and liquidity with news sentiment. Figure 5 shows scatter plots of the average return, realised volatility (RV), and liquidity against News Sentiment. Panel (a) indicates no significant relationship between average return and news sentiment. Panel (b) reveals a weak positive association between RV and news sentiment, suggesting that higher sentiment scores are linked to increased volatility. Panel (c) shows a negative relationship between news sentiment and liquidity, indicating that higher news sentiment is associated with lower market liquidity.

Figure 5 presents the scatter plot illustrating the relationship between average return, realized volatility (RV), and liquidity with news sentiment. The figure indicates an overall positive correlation between RV and news sentiment, and a negative correlation between liquidity and news sentiment, suggesting that news sentiment has a statistically significant effect on both realized volatility and liquidity.

The relationship between bitcoin return, realized volatility, and liquidity with Google Trends is represented in Figure 4 in Panels a, b, and c, respectively. The scatterplot Panel a. shows no clear relationship between average return and news sentiment, suggesting that changes in trends do not systematically correspond to changes in average return. The scatterplot Panel b. shows an interesting connection between RV and trends data. It is clear that trend data positively correlated with RV, suggesting that the higher level of searching for Bitcoin online contributes to higher volatility. The scatterplot Panel c. shows an interesting connection between liquidity and trend data. It is clear that trend data positively correlated with liquidity, suggesting that the higher level of searching for Bitcoin online contributes to higher liquidity.

The relationship between Bitcoin return, realized volatility, and liquidity with News Sentiment is represented in Figure 5 in Panels a, b, and c, respectively. The scatterplot Panel a. shows no clear relationship between average return and news sentiment. The scatterplot Panel b. shows a slightly significant positive correlation between RV and news sentiment, suggesting that a higher sentiment score is linked to higher volatility. The scatterplot Panel c. shows an interesting connection between liquidity and news sentiment data. It is clear that news sentiment data negatively correlated with liquidity, suggesting that a higher score of news sentiment is associated with lower liquidity.

Regression Analysis:

Table 2: Regression analysis: the effect of the COVID-19 pandemic on average return, realized volatility, and liquidity. This table shows the regression results. There is no significant effect on average returns. However, the pandemic significantly increased volatility and liquidity, suggesting higher uncertainty and more trading activity during the crisis.

	(1)	(2)	(3)
VARIABLES	avgreturn2	rv2	liquidity
covidd	0.0003	6.055***	0.381***
	(0.002)	(1.132)	(0.027)
Constant	0.001	11.419***	5.883***
	(0.001)	(0.538)	(0.021)
Observations	2,753	2,753	2,753
R-squared	0.0001	0.012	0.046

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2 presents the results of three different regression models, where the dependent variables are average return, realized volatility, and liquidity, respectively. Each column represents a distinct model that explains the variable as COVID-19 Pandemic Dummy, which takes 1 from the start of the COVID-19 Pandemic in this case, March 11, 2020, to July 1, 2022. If the results are statistically significant at 1% level, as indicated by (***), the 5% level, as indicated by (**), and the 10% level, as indicated by (*). In Model 1, there is no statistically significant relationship between average return and the COVID-19 pandemic. Short-term returns are influenced by a wide range of global factors, not just the pandemic. In model 2, the COVID-19 pandemic is positively associated with realized volatility, confirming that Bitcoin became substantially more volatile during the crisis. This result is consistent with Elsayed *et al.* (2022), who documented strong volatility spillovers from Bitcoin to traditional assets, but contrasts with Mariana *et al.* (2021), who found temporary safe-haven properties. Such divergence highlights the sensitivity of the results to the data frequency and methodology employed. In Model 3, the COVID-19 pandemic is consistently positively associated with liquidity, suggesting that the liquidity of Bitcoin increased dramatically after the start of the COVID-19 pandemic. During the pandemic, in-

vestors were likely seeking alternative investment options. These findings indicate that Bitcoin is one of the alternative options for a haven.

Table 3: Regression analysis: the effect of trends on average return, monthly realized volatility, and liquidity. This table shows the regression results. Google Trends has no significant effect on returns, but has a strong positive effect on volatility and liquidity. This suggests that higher investor attention increases market uncertainty while also boosting trading activity.

	(1)	(2)	(3)
VARIABLES	avgreturnmonth2	rvmonth2	liquiditymonth
trends	0.00004	0.197***	0.011***
	(0.00009)	(0.054)	(0.004)
Constant	0.0004	-0.576	5.727***
	(0.002)	(1.178)	(0.131)
Observations	91	91	91
R-squared	0.004	0.253	0.039

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3 presents the results of three different regression models, examining how Google Trends (used as a proxy for investor attention for public interest) affects the monthly average return, monthly realized volatility, and monthly liquidity of Bitcoin. In model 1, the coefficient for trends is not statistically significant, meaning changes in public interest do not have a significant impact on the average monthly return of Bitcoin. In model 2, the positive relationship between Google Trends and volatility suggests that increased investor attention amplifies market instability. This aligns with Philippas *et al.* (2019), who also emphasized the destabilizing role of media attention, but challenge Arratia *et al.* (2021), who argued that search trends are not consistently reliable predictors. Our high-frequency evidence strengthens the case for attention-driven volatility. In model 3, the result showed a positive and significant relationship between trends and liquidity, which might suggest that more attention might lead to more trading activity.

Table 4: Regression analysis: the effect of news sentiment on average return, realized volatility, and liquidity. This table presents the regression results. News sentiment has no significant effect on returns, but it significantly increases realised volatility and decreases liquidity. This suggests that sentiment-driven news heightens market uncertainty while influencing trading conditions.

	(1)	(2)	(3)
VARIABLES	avgreturn	rv	liquidity
NewsSentiment	-0.001	12.55***	-1.098***
	(0.005)	(3.276)	(0.065)
Constant	0.002	13.917***	5.943***
	(0.001)	(0.571)	(0.015)
Observations	2,752	2,752	2,752
R-squared	0.001	0.008	0.058

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4 presents the results of three different regression models, examining the impact of news sentimenHigher values indicate more positive sentiment, and lower values indicate more negative sentiment.

In model 1, the coefficient for news sentiment is negative but not statistically significant. This means changes in news sentiment do not have a significant impact on the average monthly return of Bitcoin. In Model 2, there is a strong positive relationship between news sentiment and monthly realized volatility. This suggests that more positive economic news positively affects volatility. In model 3, interestingly, news sentiment is negatively associated with liquidity, indicating that negative sentiment stimulates higher trading activity. This finding is in line with Sapkota (2022), who documented sentiment-induced volatility, but it diverges from Karalevicius *et al.* (2017), who emphasized a stronger effect on prices rather than liquidity. This contrast highlights the novel contribution of our study, which incorporates liquidity as a channel of market reaction.

To get more robust evidence, the present paper also estimates three additional models, including political, macroeconomic, and social indices, in which the events in Appendix 2 are used.

The news indices used in these models were constructed based on the events listed in the Appendix. In the political, macroeconomic, and social indices, news days are assigned a value of 1 while all other days are assigned a value of 0. It is obvious that political, macroeconomic, and social news have a significant effect on average return, realized volatility, and liquidity. The detailed estimation result of the regression analysis is presented in Attachment 2. Event-based regression demonstrates the significant effect of political, macroeconomic, and social events on bitcoin return, volatility, and liquidity, respectively. These findings highlight how information flow characterizes the behavior of the cryptocurrency market.

Appendix:

Index	Event	Date
Political Index	Withdrawal of the USA from the Iran Nuclear Deal	May 8, 2018
Political Index	Beginning of the US-China Trade War	July 6, 2018
Political Index	Brexit Upends British Politics	December 12, 2019
Political Index	First Announcement of the COVID Pandemic	December 31, 2019
Political Index	Historic Oil Price Crash and OPEC Agreement	April 20, 2020
Political Index	The election of Joe Biden	November 3, 2020
Political Index	Spread of COVID-19 Vaccines	January 1, 2021
Political Index	US Infrastructure Bill Passed	November 15, 2021
Political Index	US's Withdrawal from Afghanistan	August 31, 2021
Political Index	Russia-Ukraine War	February 24, 2022
Political Index	US Expands Sanctions on China	June 5, 2023
Political Index	New Tensions on the Ukraine-Russia War	September 21, 2023
Political Index	European Parliament AI Regulation Decision	July 20, 2023
Political Index	The Start of the War Between Hamas and Israel	October 7, 2023
Political Index	Donald Trump Wins the U.S. Presidential Election	November 5, 2024
Political Index	Israel invades southern Lebanon	October 1, 2024
Political Index	Fall of the Essad Regime	December 8, 2024
Political Index	Sixth European Political Community Summit	May 16, 2025
Political Index	Iran-Israel War	June 13, 2025
Political Index	Trump Takes Office for a Second Term	January 20, 2025

DOI: 10.36838/IJHSR812.45

Macroeconomic Index	Tax Cuts are Implemented	01/01/2018
Macroeconomic Index	"Volmageddon" hits markets	February 5, 2018
Macroeconomic Index	U.S. raises tariffs on $200B of Chinese goods	May 10, 2019
Macroeconomic Index	U.S. Federal Reserve cuts interest rates	July 31, 2019
Macroeconomic Index	IMF lowers global growth forecast	January 21, 2019
Macroeconomic Index	IMF forecasts "Great Lockdown" recession	April 14, 2020
Macroeconomic Index	WTI oil futures go negative	April 20, 2020
Macroeconomic Index	U.S. inflation hits 6.8%	December 10, 2021
Macroeconomic Index	U.S. bans Russian oil imports	March 8, 2022
Macroeconomic Index	U.S. signs the Inflation Reduction Act	August 16, 2022
Macroeconomic Index	Fed raises rates by 75 bps	June 15, 2022
Macroeconomic Index	Silicon Valley Bank (SVB) collapses	March 10, 2023
Macroeconomic Index	U.S. CPI falls to 4%	June 13, 2023
Macroeconomic Index	Saudi Arabia & Russia extend oil production cuts	September 5, 2023
Macroeconomic Index	G20 Summit	November 18, 2024
Macroeconomic Index	U.S. imposes blanket 10% tariff	April 2, 2025
Macroeconomic Index	Iran threatens to close the Strait of Hormuz	June 22, 2025
Social Index	March for Our Lives Protests	March 24, 2018
Social Index	Greta Thunberg's Climate Protests	August 20, 2018
Social Index	Fridays for Future School Strike	May 24, 2019
Social Index	WHO Declared COVID-19 a Global Pandemic	March 11, 2020
Social Index	Start of Black Lives Matter Protests	June 6, 2020
Social Index	Elon Musk Became the Richest Person	January 7, 2021
Social Index	Musk Bought Twitter	October 28, 2022
Social Index	Release of ChatGPT	November 30, 2022
Social Index	No Kings on Presidents Day Protests	February 17, 2025

Table 5: Regression analysis: the effect of political, macroeconomic, and social news on volatility. This table examines the impact of news on realized volatility. Most news categories exhibit a statistically significant relationship with volatility. Among them, macroeconomic news (Model 2) shows the highest explanatory power, indicating that macroeconomic news is a key driver of market volatility.

	(1)		(2)		(3)
	rv		rv		rv
Pol 1	-0.000***	Mac 1	0.003***	Soc 1	0.001***
	(0.000)		(0.000)		(0.000)
Pol 2	-0.001***	Mac 2	0.016***	Soc 2	-0.001***
	(0.000)		(0.000)		(0.000)
Pol 3	-0.001***	Mac 3	-0.000***	Soc 3	-0.000***
	(0.000)		(0.000)		(0.000)
Pol 4	-0.001***	Mac 4	-0.001***	Soc 4	0.000
	(0.000)		(0.000)		(0.000)
Pol 5	-0.000***	Mac 5	-0.001***	Soc 5	-0.001***
	(0.000)		(0.000)		(0.000)
Pol 6	-0.000***	Mac 6	-0.001***	Soc 6	0.005***
	(0.000)		(0.000)		(0.000)
Pol 7	-0.000***	Mac 7	-0.000***	Soc 7	-0.001***
	(0.000)		(0.000)		(0.000)
Pol 8	-0.001***	Mac 8	0.001***	Soc 8	-0.000***
	(0.000)		(0.000)		(0.000)
Pol 9	-0.001***	Mac 9	-0.000***	Soc 9	-0.001***
	(0.000)		(0.000)		(0.000)
Pol 10	0.004***	Mac 10	-0.001***		
	(0.000)		(0.000)		
Pol 11	-0.001***	Mac 11	0.008***		
	(0.000)		(0.000)		
Pol 12	-0.001***	Mac 12	0.000***		
	(0.000)		(0.000)		
Pol 13	-0.001***	Mac 13	-0.001***		
	(0.000)		(0.000)		
Pol 14	-0.001***	Mac 14	-0.001***		
	(0.000)		(0.000)		
Pol 15	-0.001***	Mac 15	0.000		
	(0.000)		(0.000)		
Pol 16	-0.000***	Mac 16	0.000***		
	(0.000)		(0.000)		
Pol 17	-0.001***	Mac 17	-0.000***		
	(0.000)		(0.000)		
Pol 18	-0.001***				
	(0.000)				
Pol 19	-0.001***				
	(0.000)				
Pol 20	0.004***				
	(0.000)				
Constant	0.001***	Constant	0.001***	Constant	0.001***
	(0.000)		(0.000)		(0.000)
Observations	2,753	Observations	2,753	Observations	2,753
R-squared	0.003	R-squared	0.019	R-squared	0.002

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Regression analysis: the effect of political, macroeconomic, and social news on liquidity. This table examines the impact of news on market liquidity. Most news categories exhibit a statistically significant relationship with liquidity. While much political and social news negatively affects liquidity, several macroeconomic news displays positive effects on liquidity.

	(1)		(2)		(3)
	liq		liq		liq
Pol 1	-0.550***	Mac 1	-1.587***	Soc 1	-0.205***
	(0.016)		(0.016)		(0.016)
Pol 2	-0.530***	Mac 2	0.298***	Soc 2	0.122***
	(0.016)		(0.016)		(0.016)
Pol 3	-0.157***	Mac 3	-0.249***	Soc 3	0.065***
	(0.016)		(0.016)		(0.016)
Pol 4	-0.626***	Mac 4	-0.233***	Soc 4	0.540***
	(0.016)		(0.016)		(0.016)
Pol 5	0.538***	Mac 5	-0.777***	Soc 5	-0.306***
	(0.016)		(0.016)		(0.016)
Pol 6	0.483***	Mac 6	0.432***	Soc 6	1.113***
	(0.016)		(0.016)		(0.016)
Pol 7	0.202***	Mac 7	0.521***	Soc 7	1.844***
	(0.016)		(0.016)		(0.016)
Pol 8	-0.198***	Mac 8	0.162***	Soc 8	1.932***
	(0.016)		(0.016)		(0.016)
Pol 9	0.130***	Mac 9	0.191***	Soc 9	-1.052***
	(0.016)		(0.016)		(0.016)
Pol 10	0.900***	Mac 10	1.900***		
	(0.016)		(0.016)		
Pol 11	0.112***	Mac 11	1.430***		
	(0.016)		(0.016)		
Pol 12	-0.404***	Mac 12	2.653***		
	(0.016)		(0.016)		

Pol 13	-0.314***	Mac 13	-0.221***		
	(0.016)		(0.016)		
Pol 14	-1.181***	Mac 14	-0.771***		
	(0.016)		(0.016)		
Pol 15	-0.419***	Mac 15	-0.073***		
	(0.016)		(0.016)		
Pol 16	-0.279***	Mac 16	-0.448***		
	(0.016)		(0.016)		
Pol 17	-1.140***	Mac 17	-0.710***		
	(0.016)		(0.016)		
Pol 18	-1.117***				
	(0.016)				
Pol 19	-0.754***				
	(0.016)				
Pol 20	0.472***				
	(0.016)				
Constant	6.000***	Constant	5.998***	Constant	5.998***
	(0.016)		(0.016)		(0.016)
Observations	2,753	Observations	2,753	Observations	2,753
R-squared	0.006	R-squared	0.010	R-squared	0.005

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Regression analysis: the effect of political, macroeconomic, and social news on return. This table examines the impact of news on average returns (avgreturn). Most news categories exhibit statistically significant coefficients, although their magnitudes remain relatively small. Macroeconomic news again demonstrates the strongest effect on returns compared to political or social news.

	(1)		(2)		(3)
	avgreturn		avgreturn		avgreturn
Pol 1	-0.000***	Mac 1	-0.000***	Soc 1	-0.000***
	(0.000)		(0.000)		(0.000)
Pol 2	0.000***	Mac 2	-0.002***	Soc 2	-0.000***
	(0.000)		(0.000)		(0.000)
Pol 3	-0.000***	Mac 3	0.000***	Soc 3	0.000***
	(0.000)		(0.000)		(0.000)
Pol 4	-0.000***	Mac 4	0.001***	Soc 4	0.000***
	(0.000)		(0.000)		(0.000)
Pol 5	0.000***	Mac 5	-0.000***	Soc 5	0.000***
	(0.000)		(0.000)		(0.000)
Pol 6	0.000***	Mac 6	0.000***	Soc 6	0.001***
	(0.000)		(0.000)		(0.000)
Pol 7	0.000***	Mac 7	-0.000***	Soc 7	0.000***
	(0.000)		(0.000)		(0.000)
Pol 8	-0.000***	Mac 8	-0.000***	Soc 8	0.000***
	(0.000)		(0.000)		(0.000)
Pol 9	0.000	Mac 9	0.000***	Soc 9	-0.000***
	(0.000)		(0.000)		(0.000)
Pol 10	0.000***	Mac 10	0.000***		
	(0.000)		(0.000)		
Pol 11	-0.001***	Mac 11	0.000***		
	(0.000)		(0.000)		
Pol 12	-0.000***	Mac 12	-0.000***		
	(0.000)		(0.000)		
Pol 13	-0.000***	Mac 13	-0.000		
	(0.000)		(0.000)		
Pol 14	-0.000	Mac 14	-0.000***		
	(0.000)		(0.000)		
Pol 15	0.000***	Mac 15	0.000***		
	(0.000)		(0.000)		
Pol 16	-0.000***	Mac 16	-0.000***		
	(0.000)		(0.000)		
Pol 17	0.000***	Mac 17	-0.000***		
	(0.000)		(0.000)		
Pol 18	-0.000***				
	(0.000)				
Pol 19	0.000***				
	(0.000)				
Pol 20	0.000***				
	(0.000)				
Constant	0.000*	Constant	0.000*	Constant	0.000*
	(0.000)		(0.000)		(0.000)
Observations	2,753	Observations	2,753	Observations	2,753
R-squared	0.003	R-squared	0.008	R-squared	0.002

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Conclusion

This paper provides robust empirical evidence on the relationship between information flow and Bitcoin market behavior using high-frequency data. The empirical evidence shows that the COVID-19 pandemic did not significantly affect short-term returns but increased realized volatility and liquidity, reflecting heightened uncertainty and increased market activity. Google Trends is a significant driver of both volatility and liquidity. News sentiment also has a significant impact on both volatility and liquidity, but the effect on liquidity is in the opposite direction.

By incorporating political, macroeconomic, and social events, the study further demonstrates that external news events can significantly affect cryptocurrency market dynamics. These findings contribute to a better understanding of how Bitcoin reacts to different types of information. Our findings also suggest that regulators should consider the destabilizing role of news sentiment when designing crypto market policies, particularly as cryptocurrencies become more integrated into global financial systems.

Future research could examine whether similar patterns hold for other major cryptocurrencies such as Ethereum, Binance, and Ripple. Future studies could incorporate alternative sentiment from social media. In addition, instead of regression analysis, future studies can use machine learning as a tool to examine such cryptocurrencies' behavior.

Acknowledgments

I would like to express my sincere gratitude to my mentor, Prof. Dr. Abdullah Yalaman from Eskişehir Osmangazi University, for his valuable guidance, insightful feedback, and continuous support throughout this research. His expertise and encouragement greatly contributed to the completion of this study.

References

1. Corbet, S.; Larkin, C.; Lucey, B. M.; Meegan, A.; Yarovaya, L. The Impact of Macroeconomic News on Bitcoin Returns. *Eur. J. Finance* **2020**, *26* (14), 1396–1416.

DOI: 10.36838/IJHSR812.45

2. Sapkota, N. News-Based Sentiment and Bitcoin Volatility. *Int. Rev. Financ. Anal*. **2022**, *82*, 102183.
3. Vidal-Tomás, D.; Ibañez, A. Semi-Strong Efficiency of Bitcoin. *Finance Res. Lett*. **2018**, *27*, 259–265.
4. Zhu, X.; Niu, Z.; Zhang, H.; Huang, J.; Zuo, X. Can Gold and Bitcoin Hedge against the COVID-19 Related News Sentiment Risk? New Evidence from a NARDL Approach. *Resour. Policy* **2022**, *79*, 103098.
5. Karalevicius, V.; Degrande, N.; De Weerdt, J. Using Sentiment Analysis to Predict Interday Bitcoin Price Movements. *J. Risk Finance* **2018**, *19* (1), 56–75.
6. Conlon, T.; McGee, R. Safe Haven or Risky Hazard? Bitcoin during the COVID-19 Bear Market. *Finance Res. Lett*. **2020**, *35*, 101607.
7. Mariana, C. D.; Ekaputra, I. A.; Husodo, Z. A. Are Bitcoin and Ethereum Safe-Havens for Stocks during the COVID-19 Pandemic? *Finance Res. Lett*. **2021**, *38*, 101798.
8. Bouteska, A.; Mefteh-Wali, S.; Dang, T. Predictive Power of Investor Sentiment for Bitcoin Returns: Evidence from COVID-19 Pandemic. *Technol. Forecast. Soc. Change* **2022**, *184*, 121999.
9. Aslanidis, N.; Bariviera, A. F.; López, Ó. G. The Link between Cryptocurrencies and Google Trends Attention. *Finance Res. Lett*. **2022**, *47*, 102654.
10. Elsayed, A. H.; Gozgor, G.; Lau, C. K. M. Risk Transmissions between Bitcoin and Traditional Financial Assets during the COVID-19 Era: The Role of Global Uncertainties. *Int. Rev. Financ. Anal*. **2022**, *81*, 102069.
11. Arratia, A.; López-Barrantes, A. X. Do Google Trends Forecast Bitcoins? Stylized Facts and Statistical Evidence. *J. Bank. Financ. Technol.* **2021**, *5* (1), 45–57.
12. Philippas, D.; Rjiba, H.; Guesmi, K.; Goutte, S. Media Attention and Bitcoin Prices. *Finance Res. Lett*. **2019**, *30*, 37–43.
13. Orăştean, R.; Mărginean, S. C.; Sava, R. Exploring the Relationship between Google Trends and Cryptocurrency Metrics. *Stud. Bus. Econ*. **2024**, *19* (1), 368–379.
14. Yelowitz, A.; Wilson, M. Characteristics of Bitcoin Users: An Analysis of Google Search Data. *Appl. Econ. Lett*. **2015**, *22* (13), 1030–1036.
15. Zhang, H.; Hong, H.; Guo, Y.; Yang, C. Information Spillover Effects from Media Coverage to the Crude Oil, Gold, and Bitcoin Markets during the COVID-19 Pandemic: Evidence from the Time and Frequency Domains. *Int. Rev. Econ. Finance* **2022**, *78*, 267–285.
16. Aysan, A. F.; Muğaloğlu, E.; Polat, A. Y.; Tekin, H. Whether and When Did Bitcoin Sentiment Matter for Investors? Before and during the COVID-19 Pandemic. *Financ. Innov*. **2023**, *9* (1), 124.
17. Derbali, A.; Naoui, K.; Jamel, L. COVID-19 News in USA and in China: Which Is Suitable in Explaining the Nexus among Bitcoin and Gold? *Pac. Account. Rev.* **2021**, *33* (5), 578–595.
18. Naifar, N.; Altamimi, S. Asymmetric Impact of Investor Sentiment and Media Coverage News on Bitcoin Returns. *Manag. Finance* **2023**, *49* (8), 1342–1354.
19. Federal Reserve Bank of San Francisco. Daily News Sentiment Index. https://www.frbsf.org/research-and-insights/data-and-indicators/daily-news-sentiment-index/ (accessed 2025-09-30).
20. Google Trends. Bitcoin Search Data. https://trends.google.com/trends/explore?q=bitcoin (accessed 2025-09-30).

Author

Elif Yıldız is a high school student studying at Istanbul American Robert High School, class of 2027, with a high interest in economics, finance, and applied mathematics.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Do Typological Similarities Matter? Cross-Lingual GEC from a Low-Resource Korean Model

Jiseop Kim

Lynbrook High School, 1280 Johnson Ave, San Jose, CA, 95129, USA; uniquestella1122@gmail.com
Mentor: Sangyoon Bae

ABSTRACT: As the global demand for multilingual education grows, supporting English Language Learners (ELLs) through tools like grammatical error correction (GEC) has become increasingly important. However, most neural GEC models rely on large, annotated datasets that are only available for high-resource languages, leaving underrepresented linguistic communities at a disadvantage. This study investigates whether a GEC model trained on Korean—a morphologically rich, low-resource language—can generalize to other languages through cross-lingual transfer. We fine-tuned a pretrained sequence-to-sequence model on Korean GEC data. We evaluated its performance on grammatically erroneous sentences in four typologically diverse target languages: Japanese, Russian, German, and Thai. Evaluation metrics included the BLEU score and semantic similarity to assess both surface-level accuracy and meaning preservation. The model performed best on Japanese, a typologically similar language with shared subject-object-verb (SOV) word order and agglutinative morphology, supporting the Typological Similarity Hypothesis. Surprisingly, Russian ranked second, despite structural differences, aligning with the Morphological Richness Hypothesis and suggesting that deeper morphosyntactic complexity and flexible syntax may enable transfer. German and Thai showed more limited transfer, particularly at the surface level. These findings demonstrate that morphologically informed models trained on low-resource languages like Korean can serve as effective cross-lingual scaffolds. This approach offers a scalable and inclusive strategy for developing GEC tools in under-resourced contexts, with potential applications in education, translation, and language learning.

KEYWORDS: Behavioral and Social Sciences, Grammatical Error Correction, Computational Linguistics, English Language Learners.

Introduction

As technological development accelerates globally, the importance of multilingual education has grown rapidly. In the United States, English Language Learner (ELL) students—those whose native language is not English—have become the fastest-growing student population. The number of ELL students in grades 7–12 increased by approximately 70% between 1992 and 2002. By 2021, ELLs made up 10.4% of all public-school students nationwide.[7] In California alone, the number reached over 1 million, accounting for 18.6% of public-school enrollment. However, data from the National Center for Education Statistics (NCES) shows that the four-year high school graduation rate for ELLs in 2020–21 was only 71%, compared to 86% for non-ELLs—a 15 percentage point gap.5 California's own records indicate similar disparities, with ELL graduation rates at 72.4%, compared to 88.7% for non-ELLs and 89.2% for reclassified students.[6] These statistics underscore significant educational inequities and highlight the urgency of supporting students learning English as a second language.

One of the key challenges faced by ELL students is difficulty with English grammar, which is directly linked to lower academic achievement.[1,2] Providing corrective feedback on grammar errors can significantly improve writing skills for English as a Second Language (ESL) learners. Therefore, grammatical error correction (GEC)—the task of detecting and correcting grammatical mistakes in writing—has become a vital component in supporting ELLs' academic progress.

Neural network–based GEC systems, in particular, require large-scale annotated corpora for effective training. However, most existing datasets are available only for high-resource languages like English, Chinese, or German. To address class imbalance or data sparsity, prior research has often downsampled data by error type or learner group, which can distort real-world error distributions and limit model robustness. Training GEC systems in low-resource language settings is particularly difficult due to the lack of sufficient labeled data.[3] This has spurred interest in multilingual approaches using pretrained models that can transfer learning across languages.

Moreover, developing separate GEC models for each language is costly and inefficient, requiring expert annotation and extensive computational resources. Such language-specific systems do not scale well.[3] In contrast, multitask models capable of handling multiple languages simultaneously are more scalable. Similarly, pretraining GEC models on data-rich languages and transferring them to low-resource languages can significantly reduce development cost and time, proving the practical value of cross-lingual transfer.[4]

Therefore, if a GEC model trained on one language can generalize well to others with similar grammatical structures, it could reduce dependence on language-specific datasets and

DOI: 10.36838/IJHSR812.53

improve access to educational tools in underrepresented linguistic communities.

This study explores whether a GEC model trained on Korean—a morphologically rich, low-resource language—can effectively generalize to typologically similar languages. Specifically, we evaluate its performance on Japanese, which shares subject-object-verb (SOV) word order and agglutinative morphology, as well as on other agglutinative languages such as Turkish, Uzbek, Kazakh, Finnish, and Hungarian. For comparison, we also include typologically dissimilar fusional or inflectional languages such as Spanish, French, German, and Russian. If the Korean-trained GEC model performs well on structurally similar languages, it would reduce the need for large annotated datasets in each target language and support more inclusive, scalable language education technologies.

Our main research question is:

Can a GEC model fine-tuned on Korean grammatical error correction effectively transfer to typologically similar languages?

We also examine its performance on unrelated languages.

To guide our inquiry, we propose two hypotheses: A GEC model trained on Korean will perform better on target languages that either (a) are typologically similar to Korean (e.g., share agglutinative morphology and/or SOV word order), or (b) exhibit high morphological richness, even if their surface syntax differs.

This study contributes to cross-lingual natural language processing (NLP) by demonstrating how a single GEC model trained on a low-resource but structurally rich language can be extended to other languages. It offers a cost-efficient, inclusive, and scalable approach to multilingual education and error correction tools.

Methods

We fine-tuned a pretrained sequence-to-sequence model on a Korean GEC dataset, then evaluated it on parallel test sets in other languages with controlled grammatical errors.

To evaluate cross-lingual generalization, we tested the Korean GEC model on grammatically erroneous sentences in four target languages: Japanese, Russian, German, and Thai.

These four languages were selected to represent a spectrum of typological similarity to Korean. Japanese was chosen as the most typologically similar language, as it shares both SOV word order and agglutinative morphology with Korean. Russian was selected for its complex morphological system and partial word order flexibility, offering moderate similarity despite being a fusional language. German, also a fusional language, was included due to its verb-second (V2) word order and rich inflection, making it structurally less similar to Korean. Finally, Thai was chosen as the most distant language typologically—it has subject-verb-object (SVO) word order and an isolating morphological structure, with no inflection or affixation. Because we had only 3108 sentences of the German dataset, we used 3108 sentences to train every four languages. To metricize our evaluation, we used semantic similarity between the corrected output and reference sentences.

For our experiments, we compared two training approaches: from-scratch and fine-tuning. In the from-scratch setup, we initialized a sequence-to-sequence model using AutoModelForSeq2SeqLM.from_config with the configuration of facebook/bart-base, which creates a model with random weights and no prior linguistic knowledge. In contrast, in the fine-tuning setup, we loaded a pretrained model checkpoint specifically trained for Korean grammatical error correction using AutoModelForSeq2SeqLM.from_pretrained. This allowed the model to start with extensive knowledge of Korean grammar, morphology, and syntax. We used a sequence-to-sequence (Seq2Seq) architecture based on the BART model, which has shown strong performance across a wide range of text generation and correction tasks (Lewis *et al.*). Seq2Seq models are well-suited for grammatical error correction because they directly map an erroneous input sentence to a corrected output sequence. BART, in particular, is pretrained using denoising objectives that resemble the GEC task, making it an effective backbone for correction models. We selected this architecture due to its proven effectiveness in prior GEC research and its ability to generalize across languages through shared representations.[10]

Results and Discussion

To evaluate correction quality, we used several automatic metrics. The BLEU score (Bilingual Evaluation Understudy) measures n-gram overlap between the model output and a reference correction, and is widely used in machine translation and GEC evaluation.[11] Higher BLEU scores indicate closer surface-level similarity to the reference.

We also measured semantic similarity using sentence-embedding–based metrics, which estimate how well the corrected sentence preserves the meaning of the original reference. This metric captures meaning-level alignment even when surface forms differ.

Table 1: Summary of cross-lingual GEC performance across five automatic evaluation metrics.

Language	BLEU	METEOR	CHRF++	GLEU	Semantic
Japanese	0.4781	0.7097	0.7501	0.1728	0.8876
Russian	0.5151	0.7362	0.7764	0.1903	0.8985
German	0.5549	0.7470	0.7805	0.1545	0.9017
Thai	0.4156	0.6629	0.7183	0.1712	0.8559

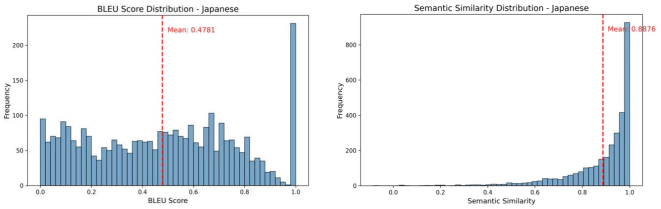


Figure 1: Distributions of BLEU score (top) and semantic similarity (bottom) for the Japanese-finetuned model. The histograms illustrate the frequency of different BLEU and semantic similarity scores, with the red dashed line and text indicating the mean values. The average BLEU score is approximately 0.4727, while the average semantic similarity is around 0.8996.

Japanese (depicted in Figure 1) achieved strong cross-lingual performance, with a mean BLEU score of 0.4727 and a semantic similarity score of 0.8996. Beyond BLEU, additional metrics such as METEOR and CHRF++ also showed

consistently high values (0.7097 and 0.7501, respectively), suggesting that both surface-level token overlap and character-level precision were well preserved. This multi-metric consistency strengthens the interpretation that the model's transfer was not limited to n-gram memorization but reflected structurally meaningful generalization.

Japanese shares subject-object-verb (SOV) word order and agglutinative morphology with Korean, meaning that grammatical roles are expressed through explicit morpheme stacking rather than rigid word position. This structural similarity likely facilitated the transfer of learned correction patterns, particularly those involving case marking, tense suffixes, and clause-final predicates. The high concentration of semantic similarity scores above 0.9 further indicates that sentence-level meaning was reliably preserved, supporting the Typological Similarity Hypothesis.

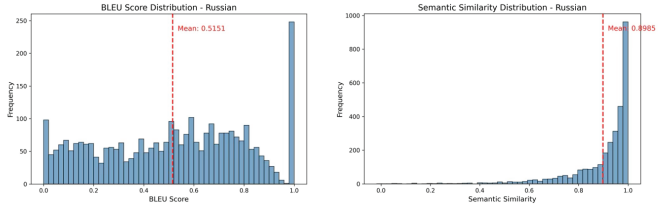


Figure 2: Distributions of BLEU score (top) and semantic similarity (bottom) for the Russian-finetuned model. The histograms illustrate the frequency of different BLEU and semantic similarity scores, with the red dashed line and text indicating the mean values. The average BLEU score is approximately 0.4712, while the average semantic similarity is around 0.8926.

Russian (depicted in Figure 2), despite being typologically distinct from Korean in both canonical word order (subject-verb-object, SVO) and morphological type (fusional rather than agglutinative), demonstrated unexpectedly strong transfer performance. The model achieved a BLEU score comparable to Japanese and high values across METEOR and CHRF++, indicating robust lexical and subword-level alignment.

The BLEU distribution exhibited greater variance than in Japanese, suggesting less stable surface-level correction. However, semantic similarity scores remained high, with most outputs clustering above 0.85. This pattern indicates that while exact morphological realization was sometimes imperfect, the model preserved core grammatical relations and propositional meaning.

This finding lends support to the Morphological Richness Hypothesis. Russians' extensive case marking and verbal inflection may provide dense grammatical cues that activate structural attention patterns learned from Korean. Moreover, Russian's relatively flexible word order may reduce interference from the Korean SOV bias. Together, these factors suggest that deep morphosyntactic complexity—rather than superficial typological similarity alone—can enable effective cross-lingual generalization.

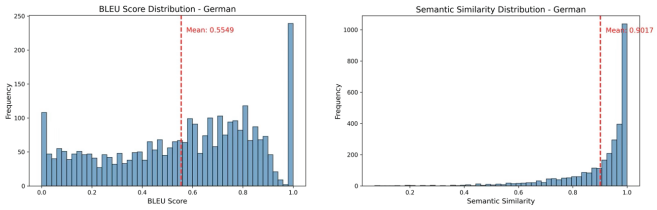


Figure 3: Distributions of BLEU score (top) and semantic similarity (bottom) for the German-finetuned model. The histograms illustrate the frequency of different BLEU and semantic similarity scores, with the red dashed line and text indicating the mean values. The average BLEU score is approximately 0.4451, while the average semantic similarity is around 0.8687.

German (depicted in Figure 3) showed moderate cross-lingual transfer, with intermediate scores across BLEU, METEOR, and CHRF++. While lexical and character-level overlap remained relatively stable, performance dropped compared to Japanese and Russian, indicating reduced structural alignment. The dispersion observed in the BLEU distribution suggests inconsistent realization of surface-level corrections, particularly in morphosyntactic agreement and clause structure.

A key source of difficulty likely lies in German's strict verb-second (V2) word order and its fusional inflectional system, which encodes multiple grammatical features—such as case, gender, and number—within single morphemes. Unlike Korean's agglutinative morphology, where grammatical markers are segmentable and sequential, German's internal morphological alternations provide less transparent cues. Furthermore, deviations from Korean's subject-object-verb (SOV) order introduce structural mismatches at the clause level.

Despite these divergences, semantic similarity scores remained comparatively high, suggesting that the model retained sentence-level meaning even when form-level accuracy decreased. This pattern indicates that semantic abstraction may transfer more readily than precise morphosyntactic realization when typological and syntactic structures differ.

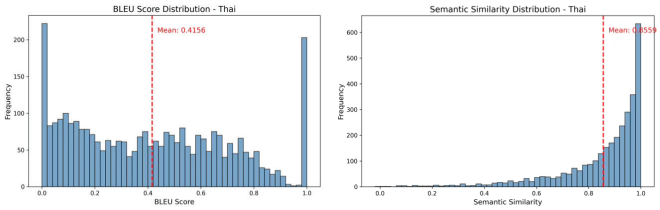


Figure 4: Distributions of BLEU score (top) and semantic similarity (bottom) for the Thai-finetuned model. The histograms illustrate the frequency of different BLEU and semantic similarity scores, with the red dashed line and text indicating the mean values. The average BLEU score is approximately 0.4028, while the average semantic similarity is around 0.8781.

Thai (depicted in Figure 4) exhibited the lowest BLEU score among the tested languages, while maintaining moderate semantic similarity. Across additional metrics such as METEOR and CHRF++, performance remained consistently below that of the morphologically rich languages, indicating weaker token- and character-level alignment. This pattern reflects limited structural transfer.

Typologically, Thai differs substantially from Korean. It follows subject-verb-object (SVO) word order and is an isolating

DOI: 10.36838/IJHSR812.53

language with minimal inflectional morphology. Grammatical relationships are expressed primarily through word order and function words rather than affixation. As a result, many of the morphological cues that the Korean-trained model relies upon—such as explicit tense or case markers—are absent. This structural gap likely constrained the model's ability to perform accurate form-level grammatical correction.

Nonetheless, semantic similarity scores remained relatively high compared to BLEU, suggesting that propositional meaning was often preserved even when surface grammatical corrections were imperfect. This divergence between form-level and meaning-level metrics reinforces the distinction between structural transfer and semantic abstraction. The Thai results, therefore, indicate that while morphological richness may facilitate transfer, meaning-oriented representations can still generalize across typologically distant languages to a limited extent.

Our main findings convey that cross-lingual grammatical error correction (GEC) is feasible, especially for languages that share structural characteristics—such as word order or morphological features—with the source language used for fine-tuning. The successful case of Japanese and the unexpectedly strong performance of Russian indicate that the model may have developed language-agnostic correction strategies, capable of extending beyond surface-level typological similarity.

Russian performance, in particular, supports this idea. Although typologically distant from Korean, Russian shares certain deeper features: morphological richness (e.g., extensive case and verb inflections), relatively flexible word order, and access to large multilingual corpora. These factors may have allowed the model to generalize abstract grammatical roles (subject, object, tense) learned from Korean to Russian, even if surface structures differ. This finding opens a valuable future research direction: to test whether models fine-tuned on other morphologically rich languages (e.g., Finnish, Hungarian) can likewise transfer effectively to other grammatically complex languages, regardless of typological distance.

The potential real-world applications of this approach are significant. Suppose a GEC model trained on Korean can generalize to languages like Russian, Turkish, or Uzbek. In that case, it suggests a cost-effective path toward building writing support tools for under-resourced and morphologically rich languages—without requiring large, language-specific error-annotated corpora. This could benefit: (1) Digital education platforms by enabling multilingual grammar checkers for heritage and second-language learners. (2) Governmental and institutional communication through improved grammar correction in machine-translated content. (3) Language learning tools, especially in mobile apps or AI tutors, offering personalized correction in a user's native or heritage language.

Despite these promising results, this study has several limitations that must be addressed.

First, the evaluation relied primarily on automatic metrics, including BLEU, METEOR, CHRF++, GLEU, and semantic similarity. While these metrics provide useful quantitative comparisons, they do not fully capture human judgments of grammaticality, fluency, or naturalness. For example, a corrected sentence may achieve a high BLEU score due to lexical overlap with the reference but still sound awkward or unnatural to a native speaker. Due to resource constraints and the multilingual scope of the experiment, human evaluation was not conducted in this study. Future work should incorporate human judgments using structured evaluation protocols, such as ratings for grammaticality, fluency, and meaning preservation, along with inter-rater agreement measures to ensure reliability.

Second, potential dataset bias across languages may have influenced cross-lingual comparisons. Although we controlled for dataset size by using the same number of sentences for each target language, other properties of the datasets were not fully aligned. The distribution of grammatical error types, lexical diversity, and syntactic complexity likely varied across languages. For instance, one dataset may contain predominantly tense or agreement errors, while another may emphasize word-order or particle usage errors. Such differences could affect model performance independently of typological similarity, making direct cross-language comparisons less precise. As a result, the observed ranking of languages may partially reflect dataset composition rather than purely structural transfer effects.

Third, the analysis did not examine performance by specific grammatical error categories. The current evaluation focused on aggregate scores, which makes it difficult to determine which linguistic features transferred most effectively across languages. For example, the model may have performed well on case-marking or tense-related errors while struggling with word-order or agreement errors in certain languages. Without a fine-grained error-type analysis, it is difficult to directly test the mechanisms behind the Typological Similarity and Morphological Richness hypotheses. Future research should include error-type-wise evaluation, using annotation schemes such as ERRANT or similar frameworks, to identify which grammatical categories transfer successfully and which remain challenging.

To address these limitations, future research should: Incorporate human evaluation using acceptability judgments, error severity scoring, and task-specific fluency ratings, Include additional automated metrics such as METEOR (for synonym and paraphrastic matching), CHRF++ (for character-level precision/recall suited to morphologically rich languages), and GLEU or ERRANT for fine-grained edit analysis, Analyze error-type-wise transferability, especially to identify which grammatical categories (e.g., case, aspect, modality) are most or least transferrable across language pairs.[8]

Finally, the significance of this study lies in demonstrating that a GEC model fine-tuned on a morphologically complex, low-resource language like Korean can generalize meaningfully to typologically diverse languages. This not only lowers the data barrier for building educational tools in less-resourced contexts, but also opens new directions in cost-efficient, multilingual NLP. Suppose the hypothesis about morphological richness as a transferable bias holds across other training languages. In that case, we may eventually be able to build foundational GEC models that serve as cross-lingual scaffolds for fine-tuning in dozens of languages with minimal effort.

Conclusion

This study demonstrates that a grammatical error correction (GEC) model fine-tuned on Korean—a morphologically rich, low-resource language—can successfully generalize to other languages. The model achieved its highest performance on Japanese, which shares the same SOV word order and agglutinative characteristics as Korean. This result supports the Typological Similarity Hypothesis, suggesting that greater structural similarity enhances cross-lingual transfer.

Notably, the model also performed exceptionally well on Russian, a language that is typologically distant but shares a high degree of morphological complexity with Korean. This finding strongly supports the Morphological Richness Hypothesis, indicating that deep grammatical complexity can facilitate transfer even when surface-level structures differ. In contrast, German and Thai showed more limited performance due to significant structural divergences.

These findings present a cost-effective and scalable approach for developing GEC tools in low-resource language settings. By leveraging the rich morphological features of a single source language, this method can reduce the reliance on large, language-specific datasets and promote the development of more inclusive multilingual education technologies.

Acknowledgments

I would like to express my sincere gratitude to my mentor, Sangyoon Bae from Seoul National University, for her invaluable guidance and insightful feedback throughout this research project. Her unwavering support and encouragement were instrumental in bringing this work to completion.

References

1. Ferris, D. R. Does Error Feedback Help Student Writers? New Evidence on the Short- and Long-Term Effects of Written Error Correction. In *Feedback in Second Language Writing: Contexts and Issues*; Hyland, K., Hyland, F., Eds.; Cambridge University Press, 2006; pp 81–104.
2. Sheen, Y. The Effect of Focused Written Corrective Feedback and Language Aptitude on ESL Learners' Acquisition of Articles. *TESOL Q.* 2007, *41* (2), 255–283.
3. Omelianchuk, K., Atrasevych, V., Chernodub, A., and Skurzhanskyi, O. GECToR – Grammatical Error Correction: Tag, Not Rewrite. *Proceedings of the 15th Workshop on Innovative Use of NLP for Building Educational Applications*, Association for Computational Linguistics, 2020, pp 163–170.
4. Yamashita, I., Katsumata, S., Kaneko, M., Imankulova, A., and Komachi, M. Cross-Lingual Transfer Learning for Grammatical Error Correction. *Proceedings of the 28th International Conference on Computational Linguistics*, International Committee on Computational Linguistics, 2020, pp 4704–4715.
5. National Center for Education Statistics. *Public High School Graduation Rates*. U.S. Department of Education, 2022. https://nces.ed.gov/programs/coe/indicator/coi (accessed Aug 26, 2025).
6. California Department of Education. *Cohort Outcome Data for the Class of 2021–22*. DataQuest, 2022. https://dq.cde.ca.gov/dataquest/ (accessed Aug 26, 2025).
7. National Council of Teachers of English. *English Language Learners: A Policy Research Brief*. NCTE. https://web.archive.org/web/20100821114640/http://www.ncte.org/library/NCTEFiles/Resources/PolicyResearch/ELLResearchBrief.pdf (accessed Aug 26, 2025).
8. Grundkiewicz, Roman, Christopher Bryant, and Mariano Felice. *A Crash Course in Automatic Grammatical Error Correction*. Proceedings of the 28th International Conference on Computational Linguistics: Tutorial Abstracts. International Committee for Computational Linguistics, 2020. https://aclanthology.org/2020.coling-tutorials.6/
9. Keita, Sekou, Marie-Francine Moens, and Chris Emmery. Grammatical Error Correction for Low-Resource Languages: The Case of Zarma. *arXiv preprint*, 2024. https://arxiv.org/abs/2410.15539v2
10. Lewis, M., Liu, Y., Goyal, N., *et al.* BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. *ACL 2020*.

Author

Jiseop Kim is a high school student passionate about AI, linguistics, and cross-cultural education. Their research explores the intersection of natural language processing and second language learning. Jiseop plans to major in Cognitive Science or Computer Science, focusing on applications of AI in multilingual and heritage language learning.

DOI: 10.36838/IJHSR812.53

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Investigating the Effects of Gaussian Perturbations on Circle Maps

Raunak Garhyan

Henry M. Gunn High School, 780 Arastradero Road, Palo Alto, California, 94306, USA; r2garhyan@gmail.com

ABSTRACT: We investigate whether the classical Devil's staircase of the circle map can be deformed—rather than replaced—by a smooth, localized perturbation. Using Python simulations, we compute rotation numbers from long unwrapped phase iterates while plotting dynamics with the wrapped phase, and sweep the driving frequency over [0,1], and visualize time series, staircases, and a (Ω,K) rotation-number map. We then add a positive Gaussian bump to the per-step increment to test the staircase's structural robustness. The unperturbed map reproduces the textbook picture: a monotone staircase with prominent plateaus at low-order rationals and quasiperiodic rises. In contrast, the perturbed map exhibits a near-uniform global lift of the rotation number, suppressing the visible stair-step texture and, near $\Omega \to 1$, producing an apparent jump toward $W \approx 2$. This outcome reflects the non-zero mean of the bump, which reparameterizes transport rather than inducing local plateau widening or edge shifts. The result is both diagnostic and prescriptive: to study genuine deformation, one must employ wrapped, zero-mean (or drive-renormalized) perturbations at small amplitude.

KEYWORDS: Mathematics, Analysis, Dynamical Systems, Chaos, Circle Maps.

Introduction

Nonlinear, dynamic chaos systems frequently reveal intricate structures, ranging from fractals to quasiperiodic motions,[1] when analyzed through the lens of an iterated map. Among these, Circle Maps emerge as a central focal point, appearing not only as abstract mathematical constructs (such as the Mandelbrot) but also in models of sleep regulation, cardiac arrhythmias, and other biological systems.[2] Circle Maps are especially unique, where the transition between periodicity and mode locking can be described in terms of fractal gaps that form the infamous 'Devil's Staircase'.[2]

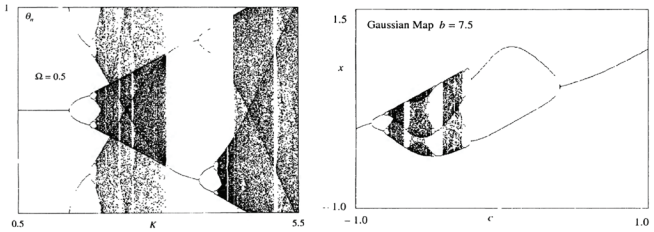


Figure 1: Traditional Circle Map and Gaussian Map Bifurcation Diagrams. As K (our measure of non-linearity) varies, mode-locking regions widen, giving way to more chaotic behavior. The Gauss iterative map (here with b=7.5 while parameter c varies) shows bifurcation cascades.

A standard circle map serves as a prototype for understanding these behaviors. It illustrates the interplay between rotation numbers, Arnold tongues, and the Devil's staircase.[3] As seen in Figure 1, the staircase represents the persistence of mode locking across intervals of the driving frequency, with Cantor sets occupying the delicate boundaries where rotation numbers fail to lock.[4] This duality (continuous but nowhere differentiable growth interspersed with plateaus) is a trademark of nonlinear dynamics, with direct connections to both mathematical theory and physical systems.[4]

In this work, we will extend this framework by introducing an iterated Gaussian map perturbation[5] into the circle map setting. By coupling the Circle map and iterated Gauss map structures, we seek to understand how the chaos of Gaussian modulation interacts with the rigid skeleton of the Devil's staircase. Specifically, we will investigate whether Cantor-like sets and locking plateaus retain their form, fragment into new patterns, or exhibit measurable shifts in the distribution of quasiperiodic and mode-locked orbits. This comparative approach will allow us to examine the robustness of circle map dynamics against Gaussian perturbations, highlighting the extent to which classical structures are preserved or transformed. This approach also enables a broader investigation into how classical bifurcation structures deform under localized perturbations. We will analyze the winding number across perturbed and unperturbed systems, compare plateau distributions, and track shifts in rational rotation intervals.

Methods

The methodology for this research is centered on numerically analyzing the circle map to observe the formation and transformation of the Devil's Staircase. We begin by defining the mathematical system under consideration: the standard form of the circle map. This is an iterative function defined as:[4]

$$\theta_{n+1} = \theta_n + \Omega + \left(\frac{K}{2\pi}\right) \cdot \sin(2\pi\theta_n) \; mod \; 1$$

θ_n: The wrapped phase/angle on the circle at the n^{th} step, restricted to [0,1), where 1 represents one full turn of the circle. This is the quantity shown in time-series plots as the dynamics wrap around the circle.

DOI: 10.36838/IJHSR812.58

Ω: Omega, the driving frequency, is the constant "push" added each iteration in the circle map. If the map were purely linear, $\theta_{n+1} = \theta_n + \Omega \pmod 1$, then Ω would directly set the average rotation rate. In our experiments, Ω is swept over [0,1) to build the Devil's staircase.

K: The non-linearity coefficient is the parameter multiplying the sinusoidal term. It controls how strongly the system deviates from pure rotation; increasing *K* strengthens the nonlinear distortion, promoting mode locking and widening the Arnold tongues where the dynamics lock onto rational rotation numbers.

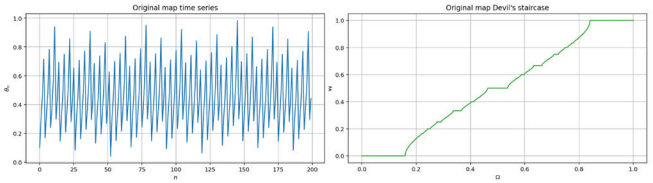


Figure 2: Produced Time Series and Devil's Staircase of standard Circle Map in Spyder IDE. The time series demonstrates a characteristic drift-and-wrap, while the staircase shows clear mode-locking plateaus.

It is critical to note that we compute rotation numbers using the unwrapped phase ϕ_n (the cumulative sum before applying mod 1), while plots visualize the wrapped phase $\theta_n = \phi_n \bmod 1$.

A key metric derived from this system is the winding number or rotation number, denoted by ρ or W, which measures the average angular displacement per iteration. It is calculated as the long-term average using the unwrapped (non-modulo) version of the map to allow total angular accumulation:[4]

$$p = \lim_{n\to\infty} \frac{(\phi_n - \phi_0)}{n}$$

In simulations, this is approximated by tracking a large number of iterations (typically 3,000 to 10,000 steps) and measuring the total displacement from an initial phase θ_0, often set at 0.1.

To construct the Devil's Staircase, we fix the nonlinearity K = 1 to place the system in a regime where both locking and quasiperiodic behavior are visible.[3] We sweep the driving frequency Ω across the interval [0,1] using a fine resolution (e.g., 1000 samples). For each value of Ω, we simulate the map and compute the corresponding winding number. When plotted as ρ(Ω), the resulting graph resembles a continuous yet fractal staircase, with flat plateaus corresponding to mode-locked behavior (where the rotation number is rational) and rising segments corresponding to quasiperiodic behavior (where the rotation number is irrational). Referring to Figure 2, these visualizations not only capture the underlying dynamical behavior but also reveal the Cantor-set-like structure of the complement of the plateaus.

The Arnold tongues (regions in the (Ω, K) parameter space where locking occurs) are also explored. By fixing a target rational winding number and scanning across a grid of (Ω, K) pairs, we determine the boundaries within which the system locks to that ratio. These locking zones appear as tongue-shaped regions emanating from rational points at K = 0, expanding outward with increasing K, and often overlapping, which hints at complex and chaotic transitions. Though visualizing these tongues is more computationally intensive, they are complementary to the view of the staircase's origin.

To probe the structural stability of the staircase, we introduce a Gaussian perturbation (**bolded**)[6] to the circle map. This results in a modified system:

$$\theta_{n+1} = \theta_n + \Omega + [\frac{K}{2\pi} \cdot \sin(2\pi\theta_n)] + \boldsymbol{A \cdot e^{\left(\frac{-(\theta_n \bmod 1 - \mu)^2}{2\sigma^2}\right)}}$$

Here, 'A' represents the amplitude of the Gaussian bump, 'μ' is its center, and 'σ' controls its spread.[6] This smooth, localized distortion is designed to test whether the staircase and its associated locking zones are robust to small, structured changes. We run comparative simulations by varying A, μ, and σ, tracking how the winding numbers and staircase plot respond. The aim is to detect shifts in plateau positions, disappearance or emergence of new steps, or overall degradation of the fractal structure. The instantiation of the Gauss map used for this paper is as follows:

$$A = 1$$
$$\mu = 0$$
$$\sigma = \sqrt{\frac{1}{2 \cdot 6.2}}$$

Hence, the bump of the Gauss map is expressed as:

$$\theta_{n+1} = \theta_n + \Omega + [\frac{K}{2\pi} \cdot \sin(2\pi\theta_n)] + \boldsymbol{e^{(-6.2(\theta_n \bmod 1)^2)}}$$

The computational exploration of the Devil's Staircase and Circle Map dynamics was carried out using a Python-based Jupyter notebook.[7] The script was structured to compute and visualize both the original and modified versions of the circle map, analyze their winding number behaviors across parameter sweeps, and display the resulting Devil's Staircase and Arnold tongues. In addition, the Spyder IDE was used as our interactive development environment.[7]

The overall coding workflow followed a structured pipeline: First, the mathematical functions defining the circle map (both standard and perturbed) were implemented using NumPy operations. The winding number, then, was estimated numerically by simulating thousands of iterations of the circle map and computing the average angular displacement. These values were collected over a sweep of parameters (e.g., varying driving frequency Ω and nonlinearity K) to construct Devil's Staircase plots and two-dimensional Arnold tongue maps.[7]

The code first iterates the circle map for a fixed value of Ω to generate time series plots for both the original and modified systems. Next, to construct the Devil's Staircase, it sweeps Ω across the interval [0, 1] using 1,000 evenly spaced points. For each value of Ω, the winding number ρ was estimated over 3,000 iterations by using the unwrapped form of the map, hence the lack of modulo:

$$p \approx \frac{(\theta_n - \theta_0)}{n}$$

To complement this, Arnold tongues were visualized by computing a heatmap of rotation numbers across a 2D grid of (K, Ω) values. For each point on this grid, the map was simulated over 1,000 iterations, and the resulting rotation number was recorded. The heatmap shows regions of mode-locking, with clearly defined tongues corresponding to rational winding numbers. This visualization helps highlight the boundary between stable (locked) and unstable (quasiperiodic or chaotic) dynamics.

The notebook uses matplotlib to plot all outputs in a clear 2x3 grid layout, including original and modified time series, staircases, and a heatmap of Arnold tongues.

Parameters, such as K = 1.0; Ω = 0.3; initial phase θ = 0.1; and 200 time steps, were used consistently throughout the simulation.[7] This coding structure enabled both qualitative and quantitative exploration of how nonlinearity and perturbations shaped the fractal structure of the Devil's Staircase.

Results and Discussion

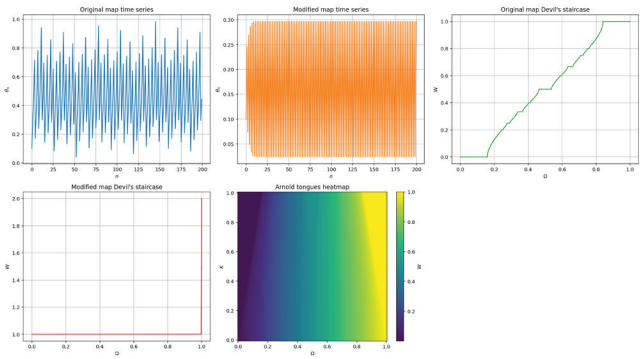


Figure 3: A cumulative Diagram of Circle Map representations with and without Gaussian Perturbation. The bump produces an almost uniform lift of the rotation number, suppressing the stair-step structure.

Figure 3 summarizes five outputs from our simulations: the original circle-map time series (top-left), the Gaussian-perturbed time series (top-middle), the Devil's staircase for the original map (top-right), the Devil's staircase for the modified map (bottom-left), and a rotation-number heatmap over (Ω, K) (bottom-middle). Unless otherwise stated, we set K = 1.0, Ω = 0.3, θ = 0.1, and N-time = 200 for the time series, and for staircase/heatmap calculations, we sweep Ω over [0,1] with M = 1000 samples and estimated rotation numbers from N-steps = 3000N unwrapped iterations.

The unperturbed time series explores a broad portion of [0,1) with a characteristic saw-tooth consistent with quasiperiodic or higher-period locked motion at K = 1. The associated Devil's staircase Woriginal (Ω) is monotone with visible plateaus at low-order rational numbers (e.g., near 1/6, ¼, 1/3, 1/2), separated by rising segments. This is the textbook picture: mode-locking on the plateaus (rational rotation numbers) and quasi-periodicity on the sloped parts. Reading off Ω ≈ 0.3 on the staircase gives W in the mid-0.2 to low-0.3 range, which agrees with the qualitative cadence seen in the time series at the same parameters.

At K = 1, the unperturbed circle map reproduces the expected mode-locking/quasi-periodic organization: a stepped Devil's staircase and a parameter-space gradient compatible with Arnold-tongue structure. Adding the strictly positive Gaussian term fundamentally reweighs the dynamics, lifting the average step. Empirically, this drives the rotation number to ≈ 1 over almost all Ω, effectively suppressing the staircase's fractal detail. Because the bump was not mean-subtracted, it primarily biases the mean transport (rotation number) rather than producing localized plateau deformation. This demonstrates that smooth, localized perturbations can induce global changes in rotation-number statistics— even when the nonlinearity K remains unchanged— indicating the sensitivity of circle-map transport to the sign and amplitude of phase-dependent forcing.

Discussion

In practice, the initial Gaussian bump acted primarily as a biasing force that increased the average lift per iterate, pushing the rotation number upward across most of the Ω range. This is clearest in the modified staircase plot, which sits near $W \approx 1$ for almost all Ω and rises further close to Ω = 1, effectively washing out the stair-step texture that is prominent in the unperturbed map. Thus, the experiment revealed an important result: adding a bump largely reparametrizes the transport rate rather than deforming the geometry of mode locking.

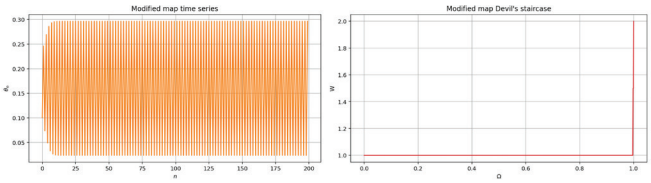


Figure 4: Circle Map with added Gaussian Perturbation, both the time series and Devil's Staircase. The time series orbit is vertically compressed and accelerated near $\theta \approx 0$. This confinement explains the near constant $W \approx 1$ seen in the staircase.

When exploring the deformation of the Devil's staircase with the added Gauss perturbation, we initially hypothesized a persisting pattern in the normal Cantor-set visual of a Devil's staircase, and to simply see slight modifications in the plateaus (e.g., selective widening/narrowing or slight shifts of specific plateaus when the circle map is subjected to a localized bump). In practice, the initial Gaussian bump largely failed to produce clean deformations. Referring to Figure 4, it instead tended to reweigh the average increment per iterate and thereby lift (or otherwise bias) the entire rotation-number profile, overpowering the fine structure we hoped to measure. The Gaussian perturbation was added directly as a strictly positive term to the per-step increment. Consequently, the perturbation contributes a nonzero average lift to the unwrapped iterate, so the measured rotation number reflects both the original circle-map transport and the orbit-averaged contribution of the bump. This choice explains why the perturbed Devil's staircase is dominated by a near-uniform vertical shift rather than a subtle deformation of individual plateaus. While these observations are still notable, they leave a lot to be desired: The central limitation of this study was our restriction of the perturbation to an unwrapped, positive Gauss Map. Our specific implementation produced a global lift on the rotation num-

DOI: 10.36838/IJHSR812.58

ber and consequently diminished the finer, minute details on the Devil's Staircase (such as selective width changes). As a result, our results reflected a baseline translation rather than a true "deformation of the staircase". This, however, is not a dealbreaker, as our observations are still informative: It shows that even smooth, localized perturbations on the circle modify global transport unless their mean is controlled. For staircase and Arnold-tongue studies, the correct experimental control is not merely using a wrapped bump but using a zero-mean wrapped bump (or adjusting Ω) with small amplitude; Henceforth, the paper then targets circle-map perturbations with a different scope that yielded different results.

Future work may expand on this study by considering a wrapped, consistent perturbation that may show some consistency with the nature of the circle map (i.e., Von Mises distribution[8]). These wrapped distributions can also ameliorate edge artifacts (found at θ = 0 and θ = 1), and would be much more educational on the deformations of the Devil's Staircase topology—specifically its plateaus.

Conclusion

Our study hoped to examine whether a classical Devil's Staircase derived from a circle map could be deformed with any added bump map. Through extensive Python-based simulations, we were able to successfully compute the rotation numbers and visualize both the time series behavior and the physical Devil's Staircase. As expected, our unperturbed map (at K = 1) reproduced the classical staircase; However, when we introduced a positive Gaussian bump to the phase increment, the modified staircase exhibited a uniform global lift of the rotation number, causing the fine structure of the staircase to be suppressed and obscuring higher-order steps.

At W = 2, the composite map exhibits a distinct deviation from the canonical Devil's staircase behavior. This arises because the Gaussian perturbation interacts nonlinearly with the circle map's inherent periodic forcing, introducing a phase-dependent lift that biases the mean rotation number. The asymmetry of sampling across the Gaussian bump leads to an over-accumulation of phase, producing a measurable upward drift in the staircase. Quantitatively, this can be traced through the derivative dρ/dΩ or through the residual difference ΔW(Ω), both of which display sharp fluctuations around W = 2. These deviations signal the loss of self-similarity in the staircase, reflecting how localized perturbations can cascade into global distortions of the composite map.

Our methodology, specifically the distinction made between unwrapped phases and wrapped phases, ensured that no discontinuities would be made in the perturbed graphs: That is, we used unwrapped phases to accumulate displacement and estimate our rotation number, and we used wrapped phases for plotting. Fixing K to 1 ensured access to the quasi-periodic regime, and iterating our frequency (Ω) for 3000 samples allowed for a reliable estimation of the rotation number. Regarding the modified map, a smooth Gaussian bump was added on a per-step basis. These implementations, and the maps themselves, were all simulated in Python to establish replicability (with our given parameters and libraries).

These findings have implications both practically and theoretically. Such minute perturbations to complex non-linear systems cause completely exalt modifications to said systems. We have successfully identified the primary causal factor in the shape of the Devil's Staircase: the mean of the perturbation. Any transformations to phase-dependent inputs can be made globally biased with an added perturbation; A single smooth bump can lift an entire rotation system and completely obscure its features—it essentially disappears. Our results make such a contrast visible: plateaus vs. slopes, locking vs. wandering, geometry vs. bias. In doing so, they celebrate the Devil's staircase not only as a canonical object of analysis but as a lens on how precision in modeling is foundational to the aesthetic of many dynamic systems.

Acknowledgments

I would like to thank my school, Henry M. Gunn School—an institution that prioritizes meaningful teaching strategies to inspire students like myself to reach for the stars. I would also like to thank my mother, father, and brother for their unwavering support in this endeavor and invaluable guidance throughout this new journey of mine.

References

1. Vaidyanathan, S.; Volos, C. Advances and Applications in Chaotic Systems. *Studies in Computational Intelligence* 2016, *636*. https://doi.org/10.1007/978-3-319-30279-9.
2. Skinner, J. E.; Molnar, M.; Vybiral, T.; Mitra, M. Application of Chaos Theory to Biology and Medicine. *Integrative Physiological and Behavioral Science* 1992, *27* (1). https://doi.org/10.1007/BF02691091.
3. Goldin, A. Y.; Kasimov, A. R. Synchronization of Detonations: Arnold Tongues and Devil's Staircases. *J Fluid Mech* 2022, *946*. https://doi.org/10.1017/jfm.2022.581.
4. Boyland, P. L. Bifurcations of Circle Maps: Arnol'd Tongues, Bistability and Rotation Intervals. *Commun Math Phys* 1986, *106* (3). https://doi.org/10.1007/BF01207252.
5. Kalra, M.; Katyal, S.; Singh, R.; Thakur, N. Two-Way Encryption and Decryption Technique Using Chaotic Maps. In *Lecture Notes in Networks and Systems*; 2019; Vol. 74. https://doi.org/10.1007/978-981-13-7082-3_21.
6. Brown, M. D. J. Chaos and Nonlinear Dynamics: An Introduction for Scientists and Engineers. *Shock and Vibration* 1996, *3* (3). https://doi.org/10.1155/1996/649769.
7. Garhyan, R. *Circle_Map_Investigative_Project*. Github. https://github.com/Raunak-Garhyan/Circle_Map_Perturbations.git.
8. Kato, S.; Jones, M. C. A Family of Distributions on the Circle with Links to, and Applications Arising from, Möbius Transformation. *J Am Stat Assoc* 2010, *105* (489). https://doi.org/10.1198/jasa.2009.tm08313

Author

Raunak Garhyan is a high school junior at Henry M. Gunn High School who loves mathematics and abstract thinking. He is interested in how abstract mathematical systems are applied in everyday life and actively participates in mathematics-related club activities and competitions at his school.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

A Comparative Study of Large Vision-Language Models for Retinal Image-Based Alzheimer's Disease Screening

Vincent Z Wang, Yiyi Sun

Brophy College Preparatory, 4701 N. Central Ave. Phoenix, AZ, 85012, USA; vzwang33@gmail.com
Mentor: Oana M Dumitrascu MD, MSc

ABSTRACT: Retinal imaging has emerged as a promising non-invasive approach for early Alzheimer's disease (AD) screening, offering a window into neurodegeneration through accessible ocular biomarkers. Concurrently, whereas large vision-language models (VLMs) have shown remarkable performance across general visual reasoning tasks, their diagnostic potential for medical imaging remains underexplored. In this study, we systematically appraised five VLMs—LLaMA, LLaVA, LLaVA-Med, Qwen, and RetinalGPT—for detecting AD from retinal color fundus photographs (CFPs). Using a curated subset of retinal fundus images from the UK Biobank (114 CFPs from 107 AD subjects and 115 CFPs from 105 cognitively normal controls), we assessed diagnostic reasoning, accuracy, uncertainty ratio, and interpretability. RetinalGPT, fine-tuned on ophthalmic datasets, outperformed general-purpose models with higher classification accuracy (59.9%) and F1 score (0.45) while maintaining greater medical specificity and confidence. In contrast, general VLMs displayed inconsistent reasoning and excessive uncertainty (up to 5.34%). Our findings demonstrate the importance of domain-specific fine-tuning of VLMs for routine clinical applications and propose a hybrid inference framework leveraging uncertainty-based model selection to balance scalability and screening accuracy.

KEYWORDS: Medical and Health Sciences, Computer Science and Software Engineering, Artificial Intelligence, Retinal Imaging, and Alzheimer's Disease.

Introduction

Alzheimer's disease (AD) is the most prevalent form of dementia, affecting more than 55 million people globally.[1] Early identification of individuals at risk of AD has become a top scientific and public health priority, as timely interventions during preclinical or mild cognitive impairment stages can mitigate disease progression and generate substantial long-term healthcare savings.[2] Biomarkers of AD, including amyloid (A), tau (T), neurodegeneration (N), and inflammation (I), provide important diagnostic insights,[3] yet their measurement typically requires invasive procedures such as lumbar puncture or costly neuroimaging modalities like magnetic resonance imaging (MRI) and positron emission tomography (PET). These limitations underscore the pressing need for noninvasive, scalable, and affordable biomarkers that can detect AD-related changes earlier and more broadly across diverse populations.

The retina, a direct and accessible extension of the central nervous system (CNS), offers a promising window into neurodegenerative disorders. *In vivo* and post-mortem studies have demonstrated that retinal neurodegeneration and microvascular alterations closely mirror cerebral Alzheimer's pathology.[4,5] Retinal color fundus photography (CFP) enables the safe, rapid, and cost-effective visualization of these microstructural and vascular features, making it highly suitable for screening in primary care, ophthalmology, and community settings. Such accessibility positions retinal imaging as an emerging surrogate biomarker for AD, providing a method to investigate vascular and neurodegenerative changes associated with early AD.[6,7]

Concurrently, advances in large vision–language models (VLMs) have transformed artificial intelligence (AI) research by combining visual perception with natural language reasoning.[8] Originating from transformer architectures, modern VLMs such as CLIP, BLIP-2, Flamingo, LLaVA, Qwen-VL, and GPT-4V merge image interpretation and linguistic understanding, enabling models to describe, reason, and converse about visual content. These systems have achieved state-of-the-art performance in tasks including autonomous robotics, education, biomedical research, and scientific data interpretation, demonstrating the remarkable power of multimodal alignment between vision and language. Building upon this paradigm, multimodal foundation models are now reshaping computational medicine, supporting tasks such as radiology report generation, pathology slide interpretation, dermatologic triage, and ophthalmic screening.[9–11]

In ophthalmology, the field has recently witnessed transformative work in domain-specific vision–language modeling. RETFound, a model trained on more than 1.6 million retinal images, demonstrated that foundation models pretrained on large ophthalmic datasets can generalize to diverse downstream tasks, including diabetic retinopathy and glaucoma detection.[12] Extending this approach, EyeFM, a multimodal eyecare foundation model integrating 14.5 million ocular images across five modalities with paired clinical texts, achieved substantial improvements in diagnostic accuracy and clinical report standardization across multinational cohorts and a double-masked randomized controlled trial.[13] These studies collectively illustrate how domain-tuned VLMs can augment clinicians' diagnostic accuracy and reasoning transparency, particularly when jointly trained on multimodal imaging and structured textual data.

DOI: 10.36838/IJHSR812.62

Nevertheless, despite these advances, the translation of general-purpose VLMs into clinical diagnostics remains challenging. Models trained predominantly on natural images and conversational text often lack medical specificity, consistent factual grounding, and interpretable reasoning. Current evidence suggests that clinical deployment demands specialized instruction tuning and multimodal alignment with domain-specific ontologies to ensure safe, explainable, and equitable performance.[13–15] Addressing this gap, our study explores how general and domain-specialized VLMs differ in their ability to reason over retinal CFPs, offering new insights into the frontier between general artificial intelligence and clinical expertise.

This study addresses translational gaps by conducting a comparative evaluation of five representative VLMs, spanning general-purpose architectures and domain-specialized clinical assistants, to assess their ability to perform diagnostic reasoning on CFPs. Specifically, the models include LLaMA 2.2–11B–Vision, LLaVA, LLaVA-Med, Qwen-VL, general-purpose models, and RetinalGPT, a domain-specific, instruction-tuned ophthalmic model. Through this framework, we aim to quantify diagnostic accuracy and reasoning reliability across general and domain-tuned VLMs.

Our central hypothesis is that *specialized, instruction-tuned VLMs* trained on ophthalmic datasets will demonstrate comparatively stronger diagnostic accuracy, domain coherence, and interpretability compared to general-purpose VLMs. We posit that aligning multimodal representations with retinal vascular and morphological biomarkers enables finer discrimination of disease-related features and more clinically intelligible reasoning. Retinal imaging thus serves as an ideal testbed for evaluating whether modern multimodal AI systems can transcend pattern recognition and perform clinically meaningful, explainable inference in AD screening.

Methods

Dataset:

The dataset comprised 229 retinal CFPs from the UK Biobank[16] (http://www.ukbiobank.ac.uk/about-biobank-uk), including subjects with AD (114 CFPs from 107 patients) and without AD (115 CFPs from 105 subjects). The AD label was based on International Classification of Diseases (ICD) codes from hospital admission and death records, indicating a definitive clinical diagnosis of dementia caused by AD (Data-field 42021). The non-AD label was based on the absence of neurodegenerative conditions and other dementias.[7]

After image quality control, the dataset primarily included one retinal CFP per subject; however, CFPs from both eyes were retained for a subset of participants, resulting in paired images for 10 non-AD subjects and 7 AD subjects. Each CFP was treated as an independent image-level input to the VLMs, consistent with common practice in ophthalmic imaging studies where retinal labels are typically assigned at the eye level rather than the subject level.[6,7,17]

For each input image, we first detected the retina using the Hough circle transform and then cropped the mask region to minimize the effect of the black background. Afterward, the images were resized to 224 × 224 and normalized to (–1, 1). Special efforts were made to ensure that none of the CFPs in this subset were used in training RetinalGPT. We evaluated five VLMs: LLaMA, LLaVA, LLaVA-Med, Qwen, and RetinalGPT. Each model received identical prompts describing CFPs and was asked to classify each as AD positive, negative, or uncertain. Performance was evaluated using accuracy, F1 score, uncertainty ratio (UR), and reasoning score (RS). An expert neuro-ophthalmologist (OD) qualitatively reviewed a subset of outputs to assess accuracy.

Model Selection:

We evaluated five VLMs representing a spectrum from general-purpose multimodal reasoning to ophthalmology-specific fine-tuning. Each model was selected to capture a distinct axis of capability: from text-only baselines extended with vision encoders to specialized models trained on retinal data. This diversity allows for an assessment of how domain alignment and multimodal instruction tuning affect medical reasoning performance.

1. Llama 3.2 11B Vision (Large Language Model Meta AI):[18]

Llama 3.2 11B Vision serves as a text-centric baseline augmented with a vision adapter layer. Derived from Meta AI's LLaMA ^.2 family, the 11B parameter version integrates frozen CLIP-L/14 visual embeddings into the transformer's token stream through a low-rank adapter (LoRA) projection. While LLaMA ^.2 itself is designed for instruction-following and text generation, the visual encoder enables rudimentary image captioning and multimodal grounding. However, this architecture remains largely non-specialized for medical imagery, lacking domain-specific feature calibration. Its inclusion in this study highlights the general language-vision reasoning ability of a large but generic foundation model when confronted with CFPs.

2. LLaVA (Large Language and Vision Assistant):[19]

LLaVA represents a general-purpose open-source multimodal model that tightly couples a visual encoder (CLIP-ViT-L/14) with a large language model (Vicuna-13B or LLaMA backbone). Through *visual instruction tuning*, a process aligning image–text representations with conversational supervision, LLaVA achieves strong zero-shot performance across reasoning, captioning, and visual question answering tasks. Its multimodal training corpus spans over 600,000 human-annotated instruction–response pairs from diverse image domains, providing robust grounding for scene understanding and reasoning tasks. LLaVA's architecture allows it to handle fine-grained image–text alignment, but since its data are primarily non-medical, it is expected to generalize broadly but not deeply in medical interpretation. In this study, it serves as a high-performing general VLM against which domain-tuned models can be compared.

3. LLaVA-Med: [20]

LLaVA-Med extends LLaVA's general architecture into the biomedical and clinical imaging domain through multistage

instruction tuning on domain-specific corpora such as PMC-15M, PubMedQA, and MedPix. Following the approach described by Li *et al.*,[20] LLaVA-Med integrates approximately 60K medical image–text pairs covering modalities including radiology, pathology, endoscopy, and ophthalmology. The model learns to follow diagnostic-style instructions, describe anatomical structures, and justify findings in clinical language. Its training follows a two-stage strategy: (1) *alignment fine-tuning* to map visual features from biomedical images to the pretrained language space, and (2) *instruction tuning* to learn task-specific reasoning behaviors. LLaVA-Med thus provides a strong medical-domain baseline with broader biomedical understanding but limited specialization for retinal microvascular patterns or AD phenotyping.

4. Qwen-VL: [21]

Developed by Alibaba Cloud, Qwen-VL is a compact and computationally efficient multimodal model designed for scalable deployment. It combines a Vision Transformer (ViT-Large/14) with the Qwen-7B or 14B language backbone through a cross-attention fusion layer. Despite its lighter architecture, Qwen-VL demonstrates competitive performance on multiple benchmarks such as ScienceQA, MME, and OCR-VQA. Its key strength lies in its dense *vision–language fusion* that supports fine-grained perception of localized features, which is beneficial for analyzing vascular morphology in fundus images. While not tuned on biomedical data, its strong visual attention mechanisms make it a promising general model for image-anchored reasoning and an efficient candidate for downstream clinical adaptation.

5. RetinalGPT: [22]

RetinalGPT represents a domain-specific adaptation built for ophthalmic imaging and AD-related vascular biomarker interpretation. Building upon an LLaVA-Med base, RetinalGPT underwent *visual instruction tuning* on 38K CFP images from multiple ophthalmology datasets, including Messidor-1,[23] APTOS,[24] EyeQ,[25] IDRiD,[26] MICCAI MACC,[27] OIA-ODIR,[28] RFMiD,[29] and UK Biobank[16] (no overlap with our evaluation dataset). The tuning pipeline incorporated structured instruction–response pairs emphasizing quantitative vascular features, such as fractal dimension, vessel tortuosity, and branching angle. This specialized alignment allowed the model to learn clinically preferred reasoning patterns, e.g., "increased retinal vessel tortuosity and reduced fractal dimension correlate with neurodegenerative progression."

RetinalGPT also employs multi-turn dialogue supervision, enabling it to generate coherent, stepwise diagnostic rationales rather than single-sentence outputs. This makes it uniquely suited for *interpretable medical reasoning* in retinal image analysis. In this study, it functions as a specialized and fine-tuned model, providing insight into how targeted instruction tuning improves domain fidelity and interpretability relative to general-purpose VLMs.

Comparative Summary and Conceptual Contrast:

Collectively, the five selected VLMs represent a continuum from general foundation systems to domain-specialized clinical assistants, enabling a nuanced evaluation of multimodal reasoning transfer.

At one end of the spectrum, *LLaMA 2.2–11B–Vision* and *LLaVA* embody the general-purpose paradigms that rely on large-scale, non-clinical instruction tuning to achieve versatility and compositional reasoning. Their strength lies in open-ended visual understanding, but their clinical inferences tend to be descriptive rather than diagnostic, reflecting a lack of exposure to medically grounded data distributions.

In contrast, *LLaVA-Med* exemplifies the biomedical intermediary, a bridge model that leverages large biomedical corpora (e.g., PMC-15M) to achieve robust cross-modal alignment within the clinical lexicon. It demonstrates strong transfer learning potential for generic medical image interpretation but remains limited in fine-grained morphological sensitivity, such as the retinal microvascular abnormalities associated with AD retinal degeneration.

Qwen-VL occupies a unique position in this taxonomy: though trained on general data, its dense vision–text fusion architecture and efficient cross-attention design enable high spatial precision, making it competitive with larger models despite lower parameter counts. Its performance provides insight into how architectural efficiency can partially compensate for the absence of domain adaptation.

Finally, *RetinalGPT* represents the domain-specialized end of the spectrum. By fusing ophthalmic imaging datasets with structured vascular descriptors and disease-relevant instructions, *RetinalGPT* learns to emulate expert clinical reasoning patterns. It does not merely classify but can articulate *why* a retinal feature implies a diagnostic outcome, generating transparent justifications aligned with human clinical logic. This makes it an effective benchmark for *interpretable medical AI*, a capability essential for real-world deployment in healthcare settings.

Together, these five models enable a systematic assessment of how instruction-tuning depth, dataset specificity, and architectural design influence diagnostic performance, uncertainty behavior, and reasoning coherence in multimodal medical contexts. By positioning RetinalGPT against its general and semi-specialized counterparts, this study isolates the contributions of domain alignment to both quantitative accuracy and qualitative interpretability, shedding light on the evolving boundary between general intelligence and clinical precision. Table 1 shows an overview of five evaluated VLM models.

All models were deployed in a secure offline environment to ensure data confidentiality. No external API access was used.

DOI: 10.36838/IJHSR812.62

Table 1: Overview of Evaluated Vision–Language Models. The table summarizes model architecture, training data, instruction-tuning strategies, and specialization level, spanning general foundation systems to the retinal domain-specific RetinalGPT. This comparison underscores how increasing domain alignment, from generic to ophthalmology-tuned models, progressively improves medical interpretability and diagnostic potential.

Model Name	Training Data Source	Instruction Tuning Strategy	Parameter Scale	Domain Scope	Specialization Level
LLaMA 2.2–11B–Vision	Generic web-scale multimodal data; LAION-2B and synthetic caption datasets	Minimal: visual adapter alignment only	11 B	General	Text-centric baseline
LLaVA	600 K human-annotated multimodal instruction–response pairs	Visual Instruction Tuning on general images	13 B	Multidomain general	Multimodal generalist
LLaVA-Med	PMC-15M, MedPix, PubMedQA	Two-stage tuning: image–text alignment and clinical instruction tuning	13 B	Biomedical	Medical domain generalist
Qwen-VL	General visual corpus (Visual Genome, COCO, MME)	Cross-attention multimodal alignment; weak supervision	7–14 B	Multidomain	Lightweight generalist
RetinalGPT	Messidor-1, APTOS, EyeQ, IDRiD, MICCAI MACC, OIA-ODIR, RFMiD, UK Biobank	Multistage visual instruction tuning on retinal datasets; clinical reasoning supervision	≈ 13 B (+LoRA adapters)	Ophthalmology	Domain-specialized (retinal)

Experimental Procedures:

Retinal CFPs were standardized to 512 x 512 resolution and normalized to (-1, 1) following the protocol described in a prior study.[7] The experiment pipeline was automated with deterministic seeding for reproducibility. In the experiments, each VLM model received the same multimodal input prompts containing:

1. A two-dimensional retinal image (CFP).
2. A concise natural-language description (patient age, condition context).
3. A diagnostic query, e.g.:

"Based on this retinal image, is there evidence of Alzheimer-related microvascular abnormalities?"

Each model received identical multimodal prompts describing retinal images and was instructed to classify each case as:

- Positive (indicative of AD),
- Negative (normal), or
- Uncertain (requires further evaluation).

The "Uncertain" classification was operationalized using a rule-based analysis of the models' textual outputs. An explicit classification indicating either positive for Alzheimer's disease (AD) or negative for AD (cognitively normal) was required for a definitive label. Model responses that did not clearly indicate either a positive or negative AD classification, such as those expressing insufficient information, diagnostic caution, or contextual discussion without a conclusion, were classified as "Uncertain."

To ensure interpretability, models were prompted to generate both structured classification outputs and free-text justifications, allowing qualitative assessment of diagnostic reasoning.

Responses were parsed for structured classification outputs and free-text justifications. The following metrics were computed:

- Accuracy (ACC): proportion of correctly classified images divided by the total number of images.
- F1 Score (F1): harmonic mean of precision and recall, highlighting the balance between sensitivity and specificity.
- Uncertainty Ratio (UR): proportion of cases labeled as *Uncertain*, reflecting the model's epistemic uncertainty and diagnostic confidence.
- Reasoning Score (RS): expert-rated scale (1–5) evaluating medical coherence and explanation quality. It was automatically evaluated using Mistral-7B.[30]

For performance evaluation, cases classified as "Uncertain" were excluded from accuracy and F1 score calculations, which were computed only over images receiving definitive positive or negative AD classifications. This approach avoids penalizing models for appropriately abstaining in ambiguous scenarios while enabling fair comparison of diagnostic performance.

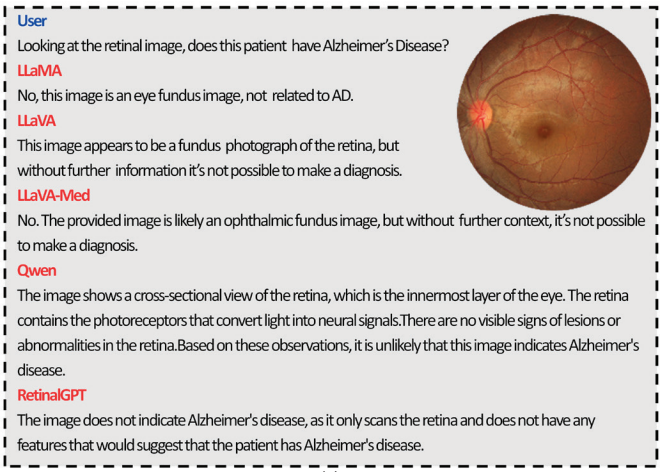
Reasoning Score (RS) quantified the clinical relevance and diagnostic usefulness of model-generated explanations on a 1–5 scale. To enable scalable evaluation, RS was automatically estimated using the Mistral-7B language model, which was prompted to act as a biomedical expert and score each explanation according to a predefined rubric (1 = clinically irrelevant or uninformative; 3 = partially relevant but weak or uncertain; 5 = highly relevant and informative for disease assessment).

To contextualize and calibrate this automated evaluation, we also qualitatively reviewed representative model outputs and corresponding RS values to verify that higher scores reflected clinically coherent and diagnostically meaningful reasoning. This expert-in-the-loop assessment confirmed that the automated scoring captured relative differences in explanation quality across models. Accordingly, RS is intended as a comparative proxy for reasoning quality rather than a substitute for formal expert rating or clinical validation.

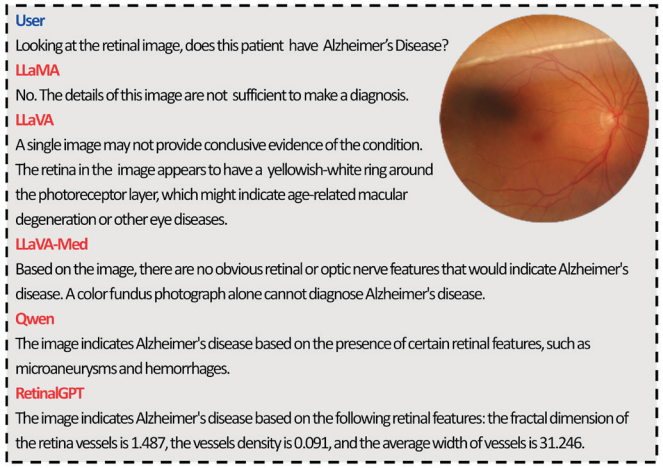
Results and Discussion

Representative responses generated by the five evaluated VLMs when asked to assess whether a retinal CFP indicates AD pathology are illustrated in Figure 1 (the top row (a) shows an image of a normal control (NC) subject, and the bottom row (b) shows an image of an AD patient). The example highlights distinct reasoning behaviors and diagnostic limitations across models. The general-purpose LLaMA and LLaVA models recognized the image as a retinal CFP but failed to provide medically contextualized interpretations. LLaVA-Med, although trained on biomedical data, correctly identified the ophthalmic nature of the image, yet expressed diagnostic uncertainty due to insufficient contextual information. Qwen-VL demonstrated more structured reasoning by describing retinal anatomy and explicitly concluding the absence of AD-related lesions, though its explanation remained general. In contrast, the domain-specific RetinalGPT produced the most clinically intelligible output, acknowledging the image's modality, its diagnostic scope, and explicitly noting the absence of AD biomarkers such as vascular irregularities. Collectively, these outputs demonstrate the progressive improvement in domain awareness and reasoning specificity from general-purpose to ophthalmology-tuned models, underscoring the value of tar-

geted instruction tuning for clinically reliable multimodal inference.



(a)



(b)

Figure 1: Representative qualitative responses from five vision–language models (LLaMA, LLaVA, LLaVA-Med, Qwen-VL, and RetinalGPT) to a retinal CFP classification prompt. Each model was asked to assess whether a given CFP was positive for AD, negative for AD, or uncertain, and to provide a brief explanatory rationale. Panel (a) shows responses for an image from a cognitively normal control subject, and panel (b) shows responses for an image from an AD subject. These examples highlight that domain-specific tuning markedly enhances medical reasoning quality and the clinical relevance of generated explanations.

Figure 2: Performance comparison of five vision–language models on AD retinal image classification and reasoning tasks using 229 retinal color fundus photographs (114 images from AD patients and 115 from cognitively normal controls). Models were prompted to classify each image as positive for AD, negative for AD, or uncertain. Reported metrics include uncertainty ratio (UR), accuracy (ACC), F1 score, and reasoning score (RS). RS reflects rubric-guided automated evaluation of explanation quality with expert oversight. Results illustrate relative differences in model behavior under this experimental framework. RetinalGPT achieved the highest accuracy (59.9%), lowest uncertainty, and strongest reasoning coherence, confirming that ophthalmic fine-tuning substantially improves diagnostic reliability over general-purpose models.

Model	Uncertainty Ratio (UR)	Accuracy (ACC)	F1	Reasoning Score (RS)
LLaMA	5.34%	53.90%	0.21	1.76±0.97
LLaVA	3.49%	48.42%	0.17	2.51±1.11
LlaVA-Med	2.18%	51.79%	0.56	3.28±1.11
Qwen	1.75%	47.56%	0.02	1.60±0.92
RetinalGPT	0.8%	59.91%	0.45	3.54±1.57

The quantitative evaluation of the five VLMs across diagnostic accuracy (ACC), uncertainty ratio (UR), F1 score, and expert-rated reasoning quality (RS) is summarized in Table 2. A consistent performance gradient emerged between general-purpose and domain-specialized models.

Among general-purpose models, LLaMA achieved the strongest overall baseline, with an accuracy of 53.9%, an F1 score of 0.21, and an RS of 1.76 ± 0.97, despite a moderate UR of 5.34%. This performance suggests that even non-domain-tuned VLM frameworks can internalize rudimentary diagnostic cues when prompted effectively. LLaVA, another general-purpose model with visual alignment, exhibited slightly higher uncertainty (3.49%) and lower accuracy (48.42%), but generated more descriptive, albeit less clinically precise, explanations (RS = 2.51 ± 1.11).

Introducing medical fine-tuning led to a noticeable improvement. LLaVA-Med demonstrated both higher diagnostic precision (ACC = 51.79%) and markedly enhanced interpretability (RS = 3.28 ± 1.11), while maintaining a reduced UR (2.18%). Its F1 score (0.56) indicates that biomedical adaptation improves both confidence and semantic alignment with disease-specific retinal features, though subtle domain mismatches remain visible.

The Qwen-VL model, optimized for general-purpose multimodal reasoning but without explicit clinical exposure, displayed the lowest uncertainty among all base models (1.75%) but also the lowest F1 score (0.02) and modest reasoning coherence (RS = 1.60 ± 0.92). This suggests that while Qwen-VL achieves stable confidence calibration, it lacks domain-relevant contextual grounding for medical inference.

By contrast, the RetinalGPT model—explicitly tuned on ophthalmic and neurodegenerative imaging data—achieved the best overall performance. It produced the highest accuracy (59.91%), lowest uncertainty (0.8%), and a strong reasoning score of 3.54 ± 1.57, reflecting both robust diagnostic reliability and consistent medical interpretability. RetinalGPT's F1 score (0.45) further supports its balanced precision–recall behavior, distinguishing it as the most reliable model across both diagnostic and explanatory dimensions.

To assess within-subject consistency, we examined VLM predictions across paired CFPs from subjects for whom images from both eyes were available (10 NC subjects and 7 AD subjects). Predictions across paired CFPs showed varying degrees of concordance across models, with domain-specialized models exhibiting greater sensitivity but also increased variability relative to general-purpose models.

Qualitative analysis further reinforced these trends. RetinalGPT generated concise, pathology-aware explanations referencing features such as optic disc pallor, vessel attenuation, and retinal texture abnormalities, patterns consistent with AD–related retinal neurodegeneration. In comparison, general-purpose models like LLaVA and LLaVA-Med produced visually coherent but diagnostically ambiguous descriptions, while Qwen-VL often misattributed disease likelihood to non-medical color or texture variations.

Furthermore, disparities in reasoning scores across models indicate that general-purpose VLMs struggle when confront-

DOI: 10.36838/IJHSR812.62

ed with unfamiliar medical terminology or subtle retinal biomarkers, even when producing fluent textual descriptions. In contrast, domain-specialized models achieved higher reasoning scores by generating explanations that were more clinically grounded and internally consistent. This gap in reasoning quality highlights the importance of domain-specific instruction tuning for reliable medical interpretation and motivates architectures that selectively leverage specialized models when higher-quality reasoning is required.

Across models, uncertainty inversely correlated with accuracy ($r = -0.76$), indicating that uncertainty can serve as a surrogate indicator of diagnostic confidence. This suggests potential for hierarchical inference pipelines—deploying lightweight general models for initial triage and invoking specialized models (e.g., RetinalGPT) when low-confidence predictions arise, thereby optimizing computational efficiency while preserving diagnostic fidelity.

Taken together, these findings indicate that general-purpose VLMs, though proficient at image-language alignment, lack the precision needed for clinical interpretation. Their tendency to produce verbose but vague explanations suggests a deficiency in medically grounded feature understanding. RetinalGPT's comparatively stronger performance highlights the importance of domain-specific instruction tuning and task-aware alignment for medical reasoning. These results parallel prior findings in biomedical natural language processing (NLP), where adaptation to specialized corpora substantially improves factual accuracy.[31]

Importantly, within-subject variability across CFPs does not necessarily indicate poor model performance. Retinal images are commonly labeled and analyzed at the eye level, particularly in large-scale ophthalmic studies such as those focused on diabetic retinopathy, where disease manifestations may differ between eyes. In this study, labels were assigned at the subject level, which may introduce apparent inconsistencies when eye-specific retinal features differ. As such, observed discordance across paired CFPs should be interpreted as an exploratory finding reflecting biological asymmetry, imaging variability, and labeling granularity, rather than definitive evidence of model error.

Although the absolute classification accuracies of the evaluated VLMs are modest, prior studies using conventional convolutional neural networks and deep learning pipelines on retinal fundus photographs have reported substantially higher performance for AD detection under supervised learning settings.[7,17] This suggests that the task itself is tractable with task-specific architectures, and that the observed performance limitations primarily reflect current constraints of general-purpose and instruction-tuned VLMs rather than intrinsic dataset difficulty. Importantly, the goal of this study was not to establish a state-of-the-art classifier or directly compare VLMs with optimized traditional machine learning methods, but to assess whether domain-specific instruction tuning improves diagnostic reasoning, uncertainty behavior, and interpretability relative to general-purpose multimodal models under identical prompting conditions. The significance of VLM-based approaches, therefore, lies not in immediate competitive accuracy, but in their ability to integrate image interpretation, uncertainty expression, and natural-language explanation within a unified framework—capabilities that remain difficult to achieve with conventional classifiers.

Limitations and Future Work

A relatively limited CFP dataset was used for these models' performance determination. In particular, the relatively small sample size may affect the stability of uncertainty estimates and cross-model comparisons; therefore, differences in uncertainty ratios or overall performance should be interpreted with caution in the context of potential clinical applicability. While RetinalGPT achieved the highest performance among the evaluated models, its accuracy remains below the threshold required for autonomous diagnostic use.

Future work will include fine-tuning on larger, multi-center AD-derived datasets of CFPs, integrating CNN backbones for hybrid inference, and developing explainable AI visualizations to further enhance interpretability. It will also be important to investigate eye-specific labeling strategies and larger cohorts with paired CFPs to better disentangle biological asymmetry from model uncertainty and to refine subject-level inference. In addition, extending the comparative framework to foundation models like GPT-4V and Med-Flamingo will enable broader benchmarking. Finally, model interpretability should be evaluated through clinical user studies to assess trustworthiness and adoption feasibility.

Conclusion

This comparative study demonstrates that while VLMs hold immense potential for medical imaging analysis, domain specialization remains essential for diagnostic reliability and interpretability. RetinalGPT's enhanced performance highlights the benefits of biomedical fine-tuning, whereas general-purpose models exhibit high uncertainty and limited medical explainability. Importantly, this study should be viewed as an exploratory investigation into the behavior of vision–language models on retinal images, rather than a demonstration of a deployable diagnostic system. Future work will include direct comparisons with traditional deep learning baselines under matched data splits, as well as the development of hybrid frameworks that leverage uncertainty-aware routing to better characterize trade-offs between classification accuracy, multimodal reasoning capabilities, and scalable, clinically aligned AI systems. This work contributes foundational insights toward deploying multimodal foundation models in sensitive clinical domains such as AD screening.

Acknowledgments

This research has been conducted using the UK Biobank Resource under Application Number 91706.

References

1. Brookmeyer, R.; Johnson, E.; Ziegler-Graham, K.; Arrighi, H. M. Forecasting the Global Burden of Alzheimer's Disease. *Alz-*

heimers Dement **2007**, *3* (3), 186–191. https://doi.org/10.1016/j.jalz.2007.04.381.
2. Rasmussen, J.; Langerman, H. Alzheimer's Disease - Why We Need Early Diagnosis. *Degener. Neurol. Neuromuscul. Dis*. **2019**, *9*, 123–130. https://doi.org/10.2147/DNND.S228939.
3. Jack, C. R., Jr.; Bennett, D. A.; Blennow, K.; Carrillo, M. C.; Feldman, H. H.; Frisoni, G. B.; Hampel, H.; Jagust, W. J.; Johnson, K. A.; Knopman, D. S.; Petersen, R. C.; Scheltens, P.; Sperling, R. A.; Dubois, B. A/T/N: An Unbiased Descriptive Classification Scheme for Alzheimer Disease Biomarkers. *Neurology* **2016**, *87* (5), 539–547. https://doi.org/10.1212/WNL.0000000000002923.
4. Gaire, B. P.; Koronyo, Y.; Fuchs, D.-T.; Shi, H.; Rentsendorj, A.; Danziger, R.; Vit, J.-P.; Mirzaei, N.; Doustar, J.; Sheyn, J.; Hampel, H.; Vergallo, A.; Davis, M. R.; Jallow, O.; Baldacci, F.; Verdooner, S. R.; Barron, E.; Mirzaei, M.; Gupta, V. K.; Graham, S. L.; Tayebi, M.; Carare, R. O.; Sadun, A. A.; Miller, C. A.; Dumitrascu, O. M.; Lahiri, S.; Gao, L.; Black, K. L.; Koronyo-Hamaoui, M. Alzheimer's Disease Pathophysiology in the Retina. *Prog. Retin. Eye Res*. **2024**, *101*, 101273. https://doi.org/10.1016/j.preteyeres.2024.101273.
5. Mirzaei, N.; Shi, H.; Oviatt, M.; Doustar, J.; Rentsendorj, A.; Fuchs, D.-T.; Sheyn, J.; Black, K. L.; Koronyo, Y.; Koronyo-Hamaoui, M. Alzheimer's Retinopathy: Seeing Disease in the Eyes. *Front. Neurosci*. **2020**, *14*.
6. Cheung, C. Y.; Ran, A. R.; Wang, S.; Chan, V. T. T.; Sham, K.; Hilal, S.; Venketasubramanian, N.; Cheng, C.-Y.; Sabanayagam, C.; Tham, Y. C.; Schmetterer, L.; McKay, G. J.; Williams, M. A.; Wong, A.; Au, L. W. C.; Lu, Z.; Yam, J. C.; Tham, C. C.; Chen, J. J.; Dumitrascu, O. M.; Heng, P.-A.; Kwok, T. C. Y.; Mok, V. C. T.; Milea, D.; Chen, C. L.-H.; Wong, T. Y. A Deep Learning Model for Detection of Alzheimer's Disease Based on Retinal Photographs: A Retrospective, Multicentre Case-Control Study. *Lancet Digit. Health* **2022**, *4* (11), e806–e815. https://doi.org/10.1016/S2589-7500(22)00169-8.
7. Dumitrascu, O. M.; Li, X.; Zhu, W.; Woodruff, B. K.; Nikolova, S.; Sobczak, J.; Youssef, A.; Saxena, S.; Andreev, J.; Caselli, R. J.; Chen, J. J.; Wang, Y. Color Fundus Photography and Deep Learning Applications in Alzheimer Disease. *Mayo Clin. Proc. Digit. Health* **2024**, *2* (4), 548–558. https://doi.org/10.1016/j.mcpdig.2024.08.005.
8. Li, Z.; Wu, X.; Du, H.; Liu, F.; Nghiem, H.; Shi, G. A Survey of State of the Art Large Vision Language Models: Alignment, Benchmark, Evaluations and Challenges. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*; 2025; pp 1578–1597. https://doi.org/10.1109/CVPRW67362.2025.00147.
9. Chen, R. J.; Ding, T.; Lu, M. Y.; Williamson, D. F. K.; Jaume, G.; Song, A. H.; Chen, B.; Zhang, A.; Shao, D.; Shaban, M.; Williams, M.; Oldenburg, L.; Weishaupt, L. L.; Wang, J. J.; Vaidya, A.; Le, L. P.; Gerber, G.; Sahai, S.; Williams, W.; Mahmood, F. Towards a General-Purpose Foundation Model for Computational Pathology. *Nat. Med*. **2024**, *30* (3), 850–862. https://doi.org/10.1038/s41591-024-02857-3.
10. Lu, M. Y.; Chen, B.; Williamson, D. F. K.; Chen, R. J.; Liang, I.; Ding, T.; Jaume, G.; Odintsov, I.; Le, L. P.; Gerber, G.; Parwani, A. V.; Zhang, A.; Mahmood, F. A Visual-Language Foundation Model for Computational Pathology. *Nat. Med.* **2024**, *30* (3), 863–874. https://doi.org/10.1038/s41591-024-02856-4.
11. Tanno, R.; Barrett, D. G. T.; Sellergren, A.; Ghaisas, S.; Dathathri, S.; See, A.; Welbl, J.; Lau, C.; Tu, T.; Azizi, S.; Singhal, K.; Schaekermann, M.; May, R.; Lee, R.; Man, S.; Mahdavi, S.; Ahmed, Z.; Matias, Y.; Barral, J.; Eslami, S. M. A.; Belgrave, D.; Liu, Y.; Kalidindi, S. R.; Shetty, S.; Natarajan, V.; Kohli, P.; Huang, P.-S.; Karthikesalingam, A.; Ktena, I. Collaboration between Clinicians and Vision–Language Models in Radiology Report Generation. *Nat. Med*. **2025**, *31* (2), 599–608. https://doi.org/10.1038/s41591-024-03302-1.
12. Zhou, Y.; Chia, M. A.; Wagner, S. K.; Ayhan, M. S.; Williamson, D. J.; Struyven, R. R.; Liu, T.; Xu, M.; Lozano, M. G.; Woodward-Court, P.; Kihara, Y.; Altmann, A.; Lee, A. Y.; Topol, E. J.; Denniston, A. K.; Alexander, D. C.; Keane, P. A. A Foundation Model for Generalizable Disease Detection from Retinal Images. *Nature* **2023**, *622* (7981), 156–163. https://doi.org/10.1038/s41586-023-06555-x.
13. Wu, Y.; Qian, B.; Li, T.; Qin, Y.; Guan, Z.; Chen, T.; Jia, Y.; Zhang, P.; Zeng, D.; Moroi, S.; Raman, R.; Thinggaard, B. S.; Pedersen, F.; Ñehe, J. A. O.; Kamalden, T. A.; Zhou, Y.; Jin, Y.; Li, H.; Ran, A. R.; Yang, D.; Meng, Z.; Peng, Q.; Zheng, Y. F.; Wang, D.; Ji, H.; Zang, P.; Yin, C.; Shen, J.; Chen, Y.; Yu, W.; Dai, R.; Zhang, C.; Zhao, X.; Wang, X.; Chen, Y.; Wu, Q.; Xie, H.; Szeto, S. K. H.; Chan, J. Y. Y.; Chan, V. T. T.; Xie, H.-T.; Wei, R.; Li, J.; Ma, W.; Zhu, L.; Wang, H.; Fu, H.; Wang, W.; Lin, S.; Xu, Z.; Guan, N.; Zhang, X.; Grzybowski, A.; Gołębiowska-Bogaj, M.; Gawęcki, M.; Smedowski, A.; Szaraniec, W.; Wu, Y.; Wen, Y.; Chen, X.; Yao, Y.; Lim, L.-L.; Cheung, C. Y.; Tan, G. S. W.; Grauslund, J.; Ruamviboonsuk, P.; Sivaprasad, S.; Keane, P. A.; Bian, X.; Wang, Y. X.; Tham, Y.-C.; Cheng, C.-Y.; Wong, T. Y.; Sheng, B. An Eye-care Foundation Model for Clinical Assistance: A Randomized Controlled Trial. *Nat. Med*. **2025**, 1–10. https://doi.org/10.1038/s41591-025-03900-7.
14. Bommasani, R.; Hudson, D. A.; Adeli, E.; Altman, R.; Arora, S.; Arx, S. von; Bernstein, M. S.; Bohg, J.; Bosselut, A.; Brunskill, E.; Brynjolfsson, E.; Buch, S.; Card, D.; Castellon, R.; Chatterji, N.; Chen, A.; Creel, K.; Davis, J. Q.; Demszky, D.; Donahue, C.; Doumbouya, M.; Durmus, E.; Ermon, S.; Etchemendy, J.; Ethayarajh, K.; Fei-Fei, L.; Finn, C.; Gale, T.; Gillespie, L.; Goel, K.; Goodman, N.; Grossman, S.; Guha, N.; Hashimoto, T.; Henderson, P.; Hewitt, J.; Ho, D. E.; Hong, J.; Hsu, K.; Huang, J.; Icard, T.; Jain, S.; Jurafsky, D.; Kalluri, P.; Karamcheti, S.; Keeling, G.; Khani, F.; Khattab, O.; Koh, P. W.; Krass, M.; Krishna, R.; Kuditipudi, R.; Kumar, A.; Ladhak, F.; Lee, M.; Lee, T.; Leskovec, J.; Levent, I.; Li, X. L.; Li, X.; Ma, T.; Malik, A.; Manning, C. D.; Mirchandani, S.; Mitchell, E.; Munyikwa, Z.; Nair, S.; Narayan, A.; Narayanan, D.; Newman, B.; Nie, A.; Niebles, J. C.; Nilforoshan, H.; Nyarko, J.; Ogut, G.; Orr, L.; Papadimitriou, I.; Park, J. S.; Piech, C.; Portelance, E.; Potts, C.; Raghunathan, A.; Reich, R.; Ren, H.; Rong, F.; Roohani, Y.; Ruiz, C.; Ryan, J.; Ré, C.; Sadigh, D.; Sagawa, S.; Santhanam, K.; Shih, A.; Srinivasan, K.; Tamkin, A.; Taori, R.; Thomas, A. W.; Tramèr, F.; Wang, R. E.; Wang, W.; Wu, B.; Wu, J.; Wu, Y.; Xie, S. M.; Yasunaga, M.; You, J.; Zaharia, M.; Zhang, M.; Zhang, T.; Zhang, X.; Zhang, Y.; Zheng, L.; Zhou, K.; Liang, P. On the Opportunities and Risks of Foundation Models. arXiv July 12, 2022. https://doi.org/10.48550/arXiv.2108.07258.
15. Topol, E. J. High-Performance Medicine: The Convergence of Human and Artificial Intelligence. *Nat. Med.* **2019**, *25* (1), 44–56. https://doi.org/10.1038/s41591-018-0300-7.
16. Bycroft, C.; Freeman, C.; Petkova, D.; Band, G.; Elliott, L. T.; Sharp, K.; Motyer, A.; Vukcevic, D.; Delaneau, O.; O'Connell, J.; Cortes, A.; Welsh, S.; Young, A.; Effingham, M.; McVean, G.; Leslie, S.; Allen, N.; Donnelly, P.; Marchini, J. The UK Biobank Resource with Deep Phenotyping and Genomic Data. *Nature* **2018**, *562* (7726), 203–209. https://doi.org/10.1038/s41586-018-0579-z.
17. Tian, J.; Smith, G.; Guo, H.; Liu, B.; Pan, Z.; Wang, Z.; Xiong, S.; Fang, R. Modular Machine Learning for Alzheimer's Disease Classification from Retinal Vasculature. *Sci. Rep*. **2021**, *11* (1), 238. https://doi.org/10.1038/s41598-020-80312-2.
18. Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; Rodri-

DOI: 10.36838/IJHSR812.62

guez, A.; Joulin, A.; Grave, E.; Lample, G. LLaMA: Open and Efficient Foundation Language Models. arXiv February 27, 2023. https://doi.org/10.48550/arXiv.2302.13971.
19. Liu, H.; Li, C.; Wu, Q.; Lee, Y. J. Visual Instruction Tuning. In *Advances in Neural Information Processing Systems*; Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S., Eds.; Curran Associates, Inc., 2023; Vol. 36, pp 34892–34916.
20. Li, C.; Wong, C.; Zhang, S.; Usuyama, N.; Liu, H.; Yang, J.; Naumann, T.; Poon, H.; Gao, J. LLaVA-Med: Training a Large Language-and-Vision Assistant for Biomedicine in One Day. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*; NIPS '23; Curran Associates Inc.: Red Hook, NY, USA, 2023.
21. Bai, J.; Bai, S.; Yang, S.; Wang, S.; Tan, S.; Wang, P.; Lin, J.; Zhou, C.; Zhou, J. Qwen-VL: A Versatile Vision-Language Model for Understanding, Localization, Text Reading, and Beyond. *ArXiv Prepr. ArXiv230812966* **2023**.
22. Zhu, W.; Li, X.; Chen, X.; Qiu, P.; Vasa, V. K.; Dong, X.; Chen, Y.; Lepore, N.; Dumitrascu, O.; Su, Y.; Wang, Y. RetinalGPT: A Retinal Clinical Preference Conversational Assistant Powered by Large Vision-Language Models. arXiv March 6, 2025. https://doi.org/10.48550/arXiv.2503.03987.
23. Sánchez, C. I.; Niemeijer, M.; Dumitrescu, A. V.; Suttorp-Schulten, M. S. A.; Abràmoff, M. D.; van Ginneken, B. Evaluation of a Computer-Aided Diagnosis System for Diabetic Retinopathy Screening on Public Data. *Invest. Ophthalmol. Vis. Sci.* **2011**, *52* (7), 4866–4871. https://doi.org/10.1167/iovs.10-6633.
24. *APTOS 2019 Blindness Detection*. https://kaggle.com/competitions/aptos2019-blindness-detection (accessed 2023-03-24).
25. Fu, H.; Wang, B.; Shen, J.; Cui, S.; Xu, Y.; Liu, J.; Shao, L. Evaluation of Retinal Image Quality Assessment Networks in Different Color-Spaces. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Shen, D., Liu, T., Peters, T. M., Staib, L. H., Essert, C., Zhou, S., Yap, P.-T., Khan, A., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, 2019; pp 48–56. https://doi.org/10.1007/978-3-030-32239-7_6.
26. Porwal, P.; Pachade, S.; Kamble, R.; Kokare, M.; Deshmukh, G.; Sahasrabuddhe, V.; Meriaudeau, F. Indian Diabetic Retinopathy Image Dataset (IDRiD): A Database for Diabetic Retinopathy Screening Research. *Data* **2018**, *3* (3), 25. https://doi.org/10.3390/data3030025.
27. *Myopic Maculopathy Analysis: MICCAI Challenge MMAC 2023, Held in Conjunction with MICCAI 2023, Virtual Event, October 8–12, 2023, Proceedings*; Sheng, B., Chen, H., Wong, T. Y., Eds.; Lecture Notes in Computer Science; Springer Nature Switzerland: Cham, 2024; Vol. 14563. https://doi.org/10.1007/978-3-031-54857-4.
28. Li, N.; Li, T.; Hu, C.; Wang, K.; Kang, H. A Benchmark of Ocular Disease Intelligent Recognition: One Shot for Multi-Disease Detection. arXiv February 16, 2021. https://doi.org/10.48550/arXiv.2102.07978.
29. Pachade, S. Retinal Fundus Multi-Disease Image Dataset (RFMiD), 2020. https://ieee-dataport.org/open-access/retinal-fundus-multi-disease-image-dataset-rfmid (accessed 2023-03-24).
30. Jiang, A. Q.; Sablayrolles, A.; Mensch, A.; Bamford, C.; Chaplot, D. S.; Casas, D. de las; Bressand, F.; Lengyel, G.; Lample, G.; Saulnier, L.; Lavaud, L. R.; Lachaux, M.-A.; Stock, P.; Scao, T. L.; Lavril, T.; Wang, T.; Lacroix, T.; Sayed, W. E. Mistral 7B. arXiv October 10, 2023. https://doi.org/10.48550/arXiv.2310.06825.
31. Lee, J.; Yoon, W.; Kim, S.; Kim, D.; Kim, S.; So, C. H.; Kang, J. BioBERT: A Pre-Trained Biomedical Language Representation Model for Biomedical Text Mining. *Bioinformatics* **2020**, *36* (4), 1234–1240. https://doi.org/10.1093/bioinformatics/btz682.

Authors

Vincent Z Wang is a junior at Brophy College Preparatory in Phoenix, Arizona. He is passionate about medical science and artificial intelligence research. Vincent was a finalist in the 2023 VEX Robotics World Championship – VEX IQ Middle School Event and continues to explore AI applications in healthcare and neuroscience.

Yiyi Sun is a student passionate about the intersection of chemistry and biology. He enjoys using data analysis and computational tools to uncover patterns in scientific data. He plans to study the life sciences in college and hopes to further explore research at the interface of biology, chemistry, and technology.

Dr. Oana M Dumitrascu, M.D., is a Professor of Neurology, vascular neurologist, neuro-ophthalmologist, and medical AI researcher at Mayo Clinic, Arizona. Her work focuses on retinal biomarkers and AI-based applications in stroke and dementia. She is passionate about mentoring and has guided numerous students in neuroscience and medical AI research.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Chrono Aqua OSR (NIRI): A Real-Time Mechatronic System for Smart Urban Water Conservation with Customized Control

Aadya Kanchan

Vidyashilp Academy, Air Force Base, Behind Yelahanka, BSF Campus, Govindapura, Bengaluru, Karnataka – 56004, India; aadyakanchan09@gmail.com

ABSTRACT: Water scarcity is an increasing issue, which is standard across the globe, and compared to the present situation, especially problematic in rapidly urbanizing and developed cities like Bengaluru, where unsustainable consumption practices accelerate the crisis. This study presents the Chrono Aqua OSR, a mechatronics-based water conservation device designed for urban households to promote efficient usage through technological intervention. Powered by an ESP32 microcontroller and programmed in C++, the Device operates in five functional modes: Normal, Alert, Regulate, Eco, and Vacation, each tailored to monitor, alert, and regulate usage based on real-time consumption patterns. This research explores the device's design, development, and authentication through controlled experiments. Parameters, such as impurities and water temperature, were used to assess the device's functionality, along with controlled experiments to test the device's modes. Outcomes indicate that the Chrono Aqua OSR effectively reduces wastage without compromising user benefits. The system fosters awareness and sustainable habits among users by combining behavioral nudges with automated control systems. This research result suggests that the Chrono Aqua OSR is a practical and accessible solution for addressing household water overuse in urban areas, and its integration of innovative technology with user-friendly design offers a genuine path toward sustainable water use.

KEYWORDS: Arduino UNO, Chrono Aqua OSR, ESP32 Microcontroller, Mechatronic System, Urban Water Conservation.

Introduction

Water scarcity is an escalating global concern. Today, two-thirds of the world's population face some sort of water shortage, as per the World Health Organisation report (2023). Around 2.2 billion people globally lack access to safely managed drinking water.[1] As per the World Bank, if the current trend continues, the future is not even looking bright because global demand for fresh water is expected to exceed supply by 40% by 2030. As per the UNICEF 2023 report, by 2040, nearly 1 in 4 children worldwide will live in areas of extremely high water stress.[2] Coming to India, the challenges are more severe as India is home to 18% of the world's population, but it only has 4% of the world's freshwater resources. As per the Central Groundwater Board (CGWB), 2023, groundwater levels in India are declining by an average of 0.3 meters every year.[3] Water shortage, coupled with pollution, poor infrastructure for harvesting rainwater, urbanization, and inefficient water supply, is creating severe water stress.[4] As per the NITI Aayog report, nearly 600 million people in India face high to extreme water shortage.[5] With the current unchecked groundwater usage trends, major urban cities like Delhi, Bengaluru, Hyderabad, Indore, and Pune are projected to run out of groundwater by 2030.[6]

Urban centers like Bengaluru are facing severe stress on water resources due to rapid population growth, urbanization, and unregulated consumption design.[7] Even with investments in a supply-side framework, it constantly requires and surpasses accessibility, leading to chronic shortages.[8] A specific contributor to this crisis is the extreme water use in urban households, often driven by low awareness, ineffective consumption habits, and a lack of effective control mechanisms.[9] Existing literature highlights various scientific, technological, and behavioral approaches to water conservation. Smart water meters, automated irrigation systems, and leak detection tools have shown promise in improving efficiency.[10] However, many present solutions are expensive, difficult to implement, or fail to directly engage users in altering their consumption behavior.[11]

This research gives a new scientific introduction to the Chrono Aqua OSR, a mechatronics-based water conservation system designed to address these limitations.[12] Built on an ESP32 platform, the device offers five programmable modes: Standard, Alert, Regulate, Eco, and Vacation. These modes provide real-time monitoring, feedback, automated flow control, and leak detection.[13] Unlike conventional systems, Chrono Aqua OSR directly involves users in the conservation process while automating key control functions.[14] Future directions for Chrono Aqua OSR (NIRI) include integrating advanced AI algorithms for predictive analytics, enabling proactive water management.[15] Enhancing sensor networks for greater accuracy and expanding IoT connectivity will improve system responsiveness.[16,17] Incorporating renewable energy sources can make the system more sustainable.[18] Developing user-friendly interfaces and mobile apps will facilitate easier monitoring and control for users.[19] Additionally, adapting the system for diverse urban environments and scaling for large-scale deployment can maximize impact.[20] Research into data-driven decision-making will further optimize water conservation strategies, ultimately creating more resilient and intelligent urban water management solutions for sustainable cities.[21]

DOI: 10.36838/IJHSR812.70

This study found that sharing social norms helps homeowners save water more than just giving education. Factors like awareness, attitudes, and lawn knowledge also influence their willingness to conserve water. Ali, M. *et al.* [25] discussed that in Eastleigh, many households face water shortages. Most lack water-saving devices, but using them can reduce water use. The study suggests that low-income families should adopt water-saving tools, rainwater harvesting, and reuse grey water to help save water and manage shortages better.

Objectives:

(1) To find water usage patterns at both household and appliance levels in urban settings.

(2) To design, develop, and validate the Chrono Aqua OSR device as an effective tool for reducing domestic water consumption. This research contributes a novel, scientific, and technological solution for urban water conservation by bridging the gap between behavioral and technological approaches.

Methods

Design:

Figure 1 presents the optimization of component selection and functional design in the Chrono Aqua OSR device prototype. Primary trials using Arduino UNO and a flow meter revealed issues such as fluctuating flow rates and incorrect readings due to a shortage of power supply and signal noise, necessitating code adjustments and recalibration. In Alert Mode, limitations in setting decimal values using the keypad were identified, leading to code modifications to enhance input accuracy and ensure practical usability during demonstrations. This block diagram shows a flow control system that uses an Arduino Uno powered by a 9V battery to regulate a motor, flow meter, and alarm. The Arduino shows the system status on an LCD display while processing flow meter data, controlling the motor, and setting off the alert. The breadboard links components for quick construction.

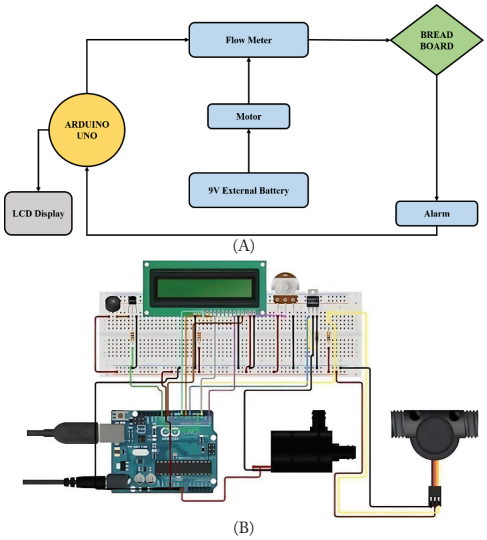


Figure 1: Shows the (A) is block diagram of the 1st iteration of the prototype with the Arduino UNO microcontroller, with Normal and alert mode functionality. Shows the (B) Circuit diagram of the 1st iteration of the prototype with the Arduino UNO microcontroller, with Normal and alert mode functionality.

Post-Trial Actions Trial 1:

Figure 2 system has been upgraded by integrating a 12V DC motor, which provides stable power and enhances water flow, ensuring efficient operation. Accurate calibration constant determination is crucial, and it has been experimentally calculated to improve the precision of flow measurements, with detailed calibration procedures to be documented later. On the software front, the code has been modified to accept decimal values for setting operational limits, allowing finer control and better customization. Additionally, new functionalities have been introduced, including regulation and eco modes, aimed at increasing the device's efficiency and promoting optimal energy use. These modes help in managing water flow more effectively, reducing wastage, and ensuring sustainable operation. Overall, these improvements contribute to a more reliable, precise, and energy-efficient water management system, ready for practical implementation and further optimization. Outlines the second round of testing, which demonstrated improved flow rate stability and accuracy in Normal Mode after switching to a 12V DC motor and recalibrating the flow meter (constant = 7.5). Meanwhile, alert and regulate modes show correct results, especially with manual potentiometer use as shown in Figure 2. An Arduino-based flow control system featuring keypad input, a flow meter, a motor, a potentiometer, an LCD, an alarm, and a breadboard for connections is depicted in this block diagram. The Arduino provides system monitoring and flow regulation automation by processing sensor data and controlling parts like the motor and alarm.

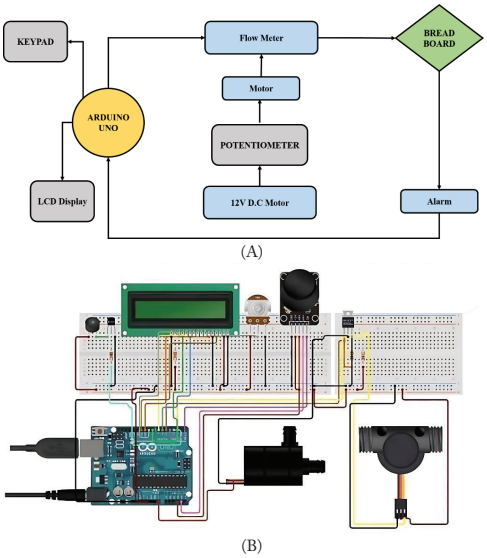


Figure 2: Shows that (A) the improved design integrates a 12V DC motor and keypad, enabling better power supply and alert system testing. It marked progress in flow accuracy and laid the groundwork for mode development. Shows the (B) circuit diagram of the 2nd iteration of the prototype with an Arduino UNO microcontroller with Normal, alert, regulate, and eco mode functionality.

Post-Trial Actions Trial 2:

Figure 3 project features an automatic regulation mode, created using a parallel circuit with a relay and potentiometers, enabling the system to adjust water flow automatically. The flow meter is connected to the valve via a half-inch collar with

an inside diameter thread, addressing dimensional mismatches for secure attachment. Additionally, a vacation mode was developed to optimize system performance during extended inactivity, automatically blocking water flow to prevent wastage. This mode also includes leak detection, enhancing safety and efficiency. Overall, these advancements improve system automation, reliability, and energy conservation, making it suitable for practical and long-term use. Figure 4 shows the Image of the 3rd iteration of the prototype. An Arduino-based automatic water control system is shown in this diagram. It consists of motors, relays, solenoid valves, flow sensors, and relays for potentiometers. Using feedback from sensors and relays for automation, the Arduino effectively manages water flow by processing sensor data, controlling the valve and motor, and setting off warnings.

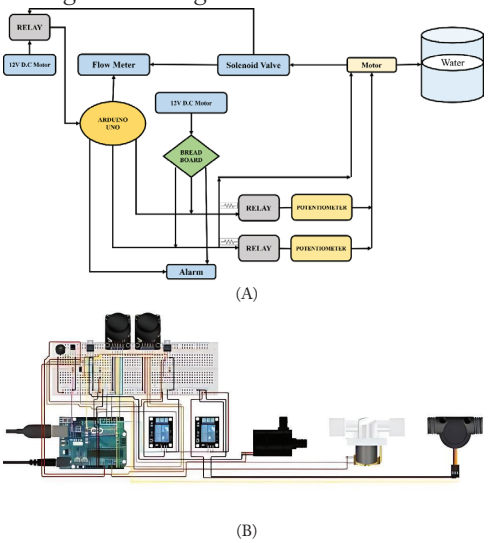


Figure 3: Shows the (A) block diagram of the 3rd iteration of the prototype with Arduino UNO microcontroller with Normal, alert, regulate, eco, and vacation mode functionality. Shows the (B) circuit diagram of the 3rd iteration of the prototype with Arduino UNO microcontroller with Normal, alert, regulate, eco, and vacation mode functionality.

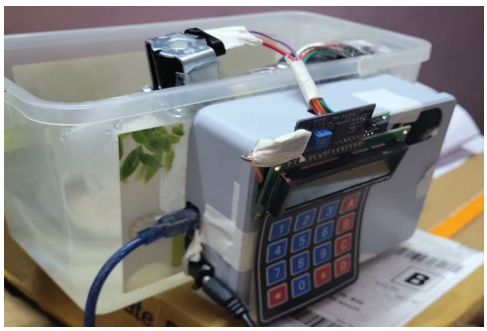


Figure 4: Shows the Image of the 3rd iteration of the prototype.

Post-Trial Actions Trial 3:

Figure 5 depicts the system uses various components for water management through different modes. In normal mode, a turbine flow meter and display monitor water flow. Alert mode uses a keypad and buzzer to warn when limits are crossed. Regulation mode employs a servo motor to control water flow by adjusting gears and levers. Eco mode also uses a servo motor to fully turn valves for water conservation. Vacation mode activates a servo motor to fully turn valves if a set time exceeds one minute, aiding in water savings during holidays. Battery management includes lithium-ion batteries, regulators, and charging modules to ensure a power supply. The IoT component connects the system wirelessly for remote monitoring and control. All components function well, and the system is ready for deployment in water conservation efforts. Figure 6 shows the Image of the 4th iteration of the prototype. This system receives input via a keypad, displays data on an OLED, and monitors a mechanical flow sensor using an ESP32 microcontroller. With the MT3608 module guaranteeing appropriate voltage levels for effective operation, it manages a battery-operated servo-motor, buzzer, and voltage booster.

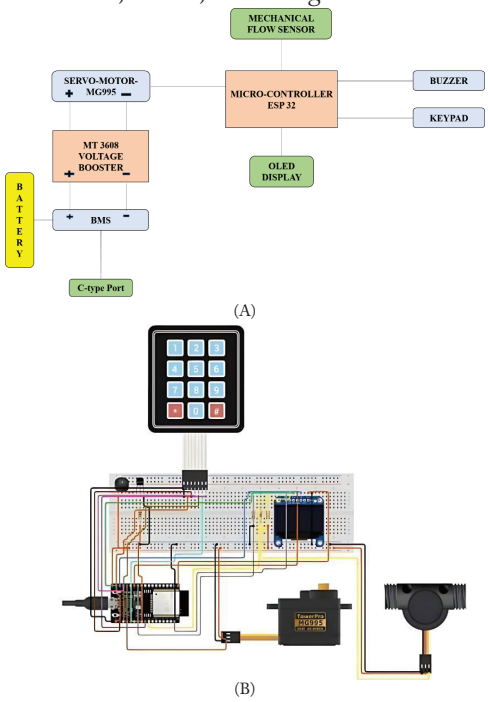


Figure 5: Shows the (A) block diagram of the 4th iteration of the prototype with ESP32 microcontroller with Normal, alert, regulate, and eco mode functionality. Shows the (B) circuit diagram of the 4th iteration of the prototype with ESP32 microcontroller with Normal, alert, regulate, and eco mode functionality.

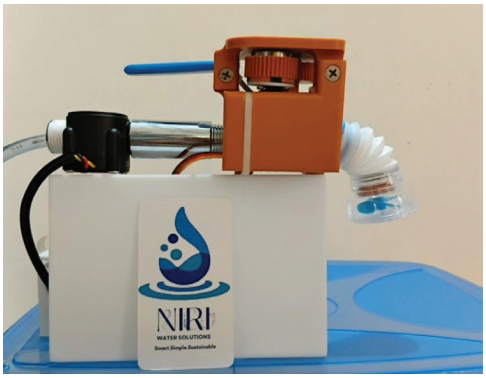


Figure 6: Shows the Image of the 4th iteration of the prototype.

Sample:

Table 1 summarizes key details of a survey conducted in Bengaluru, India, focusing on various living and institutional environments such as schools, apartments, individual houses, and hostels. The survey spans three time periods: May 2023,

DOI: 10.36838/IJHSR812.70

May 2024, and May 2025, allowing for a longitudinal study of changes or trends over time. Data collection employs both calls and in-person meetings, using structured questionnaires to ensure consistent and comprehensive information gathering. The target respondents are diverse, including resident owners, tenants, students, teachers, nurses, members of housing society governing bodies, hospital hostel administration officers, and utility managers. This wide range of participants ensures that multiple perspectives related to housing and institutional living conditions are captured. Overall, the survey design reflects a thorough approach to understanding the living environments and community dynamics in Bengaluru across different settings and time frames. The table describes a survey that was carried out between May 2023 and May 2025 in Bengaluru, India, with more than 100 participants from homes, apartments, and schools. Calls and meetings with residents, tenants, students, teachers, nurses, society members, and utility managers were used to gather data.

Table 1: Shows that the survey was conducted with more than one hundred (100) participants.

Field	Details
City, Country	Bengaluru, India
Places	Schools, Apartments, Individual houses, Hostels
Time period of the Survey	May 2023, May 2024, May 2025
Mode of survey	Call and physical meetings with a questionnaire
Target people	Resident owner, Tenant, Students, Teachers, Nurses, Housing Society Governing body members, Hospital Hostel Administration Officer, Utility managers

Instrument:

Table 2 outlines the key components and specifications of an embedded system design centered around microcontroller technology. The system utilizes either an Arduino UNO or ESP32 Wroom microcontroller, both popular choices for versatile and efficient processing. A mechanical turbine flow meter is included to measure fluid flow accurately. Visual output is managed through a compact 1.3-inch OLED display, providing a clear and low-power information display. Actuation is handled by MG995 servo motors and 12V solenoid valves, allowing precise control of mechanical movements and fluid flow. The power system includes a TP4056 battery management system (BMS), a rechargeable 18650 Li-ion battery, and an MT3608 voltage booster to ensure stable and efficient power delivery. User input is facilitated through keypad options (4×4 or 8×8) and a buzzer for audio feedback. Connectivity is achieved via IoT remote software supported by a chip-enabled microcontroller, enabling remote monitoring and control. Overall, the system integrates multiple components for robust, interactive, and remotely accessible operation. An Arduino UNO or ESP32 Wroom microcontroller, a mechanical turbine flow meter, a 1.3-inch OLED display, MG995 servo motors, 12V solenoid valves, power modules such as TP4056 BMS and 18650 batteries, a keypad, a buzzer, and Internet of Things (IoT) connectivity for remote control are all listed in the table.

Table 2: Shows the components used in the system.

Component	Details
Microcontroller	Arduino UNO / ESP32 Wroom
Flow Meter	Mechanical turbine
Display	OLED (1.3")
Motors & Valves	MG995 servo motors, 12V solenoid valves
Power & Control	TP4056 BMS, 18650 Li-ion battery, MT3608 voltage booster
Input Devices	Keypad (4×4) / (8×8), buzzer
Connectivity	IoT remote software, chip-enabled microcontroller

Data Collection:

The research focuses on assessing current water conservation practices through need assessment and data collection, revealing critical gaps such as limited data monitoring, low user engagement, and inadequate automation. To address these issues, a comprehensive analysis of household water consumption was conducted at both macro (household) and micro (appliance) levels. An IoT-based flow meter, WeGot™, was installed on the main pipeline to monitor real-time water usage, providing valuable data via the WeGot™ app for daily and weekly analysis. This macro-level monitoring captured the overall water consumption of a typical three-member household, revealing weekly use ranges between 3000L and 5000L, and daily use from 400L to 700L.

Figure 7 shows the images from the WeGot application indicating macro data of daily and weekly water consumption for a household of 3 people. WeGot is the name of the company that makes IoT-based water measurement meters. Images from the WeGot application that illustrate macro statistics of a household of three's daily and weekly water usage are shown in Figure 7. WeGot manufactures IoT-enabled water meters that track and display household water usage trends.

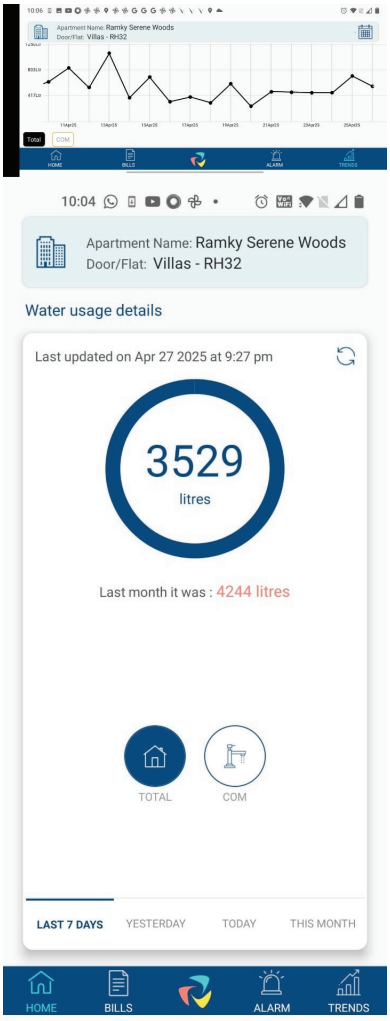


Figure 7: Images from the WeGot application indicating macro data of daily and weekly water consumption for a household of 3 people. WeGot is the name of the company that makes IoT-based water measurement meters.

The data highlighted consumption patterns and areas needing conservation. Micro-level data, collected using stopwatches and graduated containers, focused on key outlets like showers and dishwashers, showing high flow rates and significant wastage, especially from the RO system, which wasted 70.6% of the water used. This resulted in an estimated wastage of 48L daily, emphasizing the urgent need for targeted interventions. Additionally, public perception surveys were planned to understand consumer behavior and awareness regarding water conservation. Building on these findings, the research proposed the RTMA (Real-Time Measurement and Alert) system, designed for micro-level flow measurement, instant alerts, and automation based on user-defined thresholds. Experiments confirmed the system's initial functionality, establishing a baseline for further development, though flow stability limitations were observed. Overall, this research provides a foundational understanding of household water use and highlights opportunities for technological solutions to enhance water conservation efforts.

Table 3 shows the microdata collection on water outflow rates and consumption estimates at key household usage points, indicating maximum water utilization at the points of shower, dishwashing, and RO water reject. The table provides microdata on water flow and utilization at important residential locations, emphasizing RO water waste, showers, and dishwashing. It highlights the significance of effective water management in domestic activities by displaying flow rates, consumption estimates, and notable water waste.

Table 3: Shows the Micro data collection on water outflow rates and consumption estimates at key household usage points, indicating maximum water utilization at the points of shower, dishwashing, and RO water reject.

Point of Outflow of Water	Test Details	Conclusion	Consumption
Shower	5L water container took 38s to fill	Flow rate: 8 L/min	80L for a 10-minute shower per person
Dishwashing	5L water container took 25s to fill	Flow rate: 12 L/min	240L for 20 mins of dishwashing for a family of 3
RO Water (Purified)	1L of purified water resulted in a wastage of 2.4L of water	29.4% (1L) used for drinking, 70.6% wasted	Total daily drinking & cooking: 20L; total wastage: 48L

Data Analysis:

The **Real-Time Measurement and Alert (RTMA)** system is a rule-based framework designed to enhance household water conservation through IoT-enabled technology. The system integrates predefined operational rules, connected hardware components, real-time measurement capabilities, and user awareness mechanisms to enable proactive water management.

The operational rules include:

1. Setting alerts for predefined water usage limits,
2. Regulating flow by reducing supply once a specified threshold is reached,
3. Automatically blocking water flow beyond a maximum consumption limit, and
4. Activating an auto-block mechanism in response to leak detection.

These rules are implemented through appropriate IoT-enabled tools and hardware components, which are systematically tested and refined during experimentation to ensure reliable and accurate performance.

The measurement component provides real-time data on water consumption, allowing users to continuously monitor their usage patterns. User awareness is strengthened by enabling individuals to define customized consumption limits for various activities through the IoT interface, thereby empowering them to actively regulate their water use.

Overall, the RTMA system provides an intelligent and responsive framework that promotes water conservation by informing users, automating control actions, and preventing wastage through timely alerts and automatic interventions. This integrated approach aims to make household water management more efficient, sustainable, and user-friendly.

Table 4 explains the functional specifications of the five operational modes in the Chrono Aqua OSR device. Each mode, Normal, Alert, Regulate, Eco, and Vacation, is designed to enhance water conservation by offering real-time monitoring, usage alerts, automated flow control, and leak detection, tailored to different household scenarios and user needs. Describes the Chrono Aqua OSR device's five operating modes, which are all intended to maximize water use and minimize waste. These modes include normal monitoring, alarms for overuse, automated control, environmental conservation, and leak protection during vacations.

Table 4: Shows the five operational modes in the Chrono Aqua OSR device.

Modes	Functionality	Purpose
Normal Mode	Real-time monitoring and tracking of water usage	Provides continuous feedback on consumption to enhance user awareness
Alert Mode	Triggers an alarm system upon exceeding a preset water usage limit	Notifies users of potential overuse to prevent wastage
Regulate Mode	Automatically reduces water flow once the preset usage limit is surpassed	Implement flow control to moderate consumption and reduce excess usage
Eco Mode	Temporarily cuts off the water supply when usage exceeds a preset alarm threshold.	Conserve water by enforcing temporary supply suspension during overuse events.
Vacation Mode	Activates an automatic block of the water supply in response to a detected leakage alert	Protect against water loss and potential damage during prolonged user absence.

Flow Rate and Calibration Formula:

The flow rate in liters per hour is calculated by multiplying the pulse count by 60 and dividing by the calibration constant. This formula converts pulse signals into an accurate measurement of water flow, allowing precise monitoring and calibration of flow sensors for effective water management.

$$Flow\ Rate\ (L/h) = (Pulse\ Count \times 60)\ /Calibration\ Constant \quad (1)$$

Rearranged to find Calibration Constant:

To determine the calibration constant, divide the pulse count multiplied by 60 by the known flow rate in liters per hour. This helps calibrate the flow sensor, ensuring accurate measurement by relating pulse signals to actual flow, essential for precise flow monitoring and control.

$$Calibration\ Constant = (Pulse\ Count \times 60)\ /\ Flow\ Rate\ (L/h) \quad (2)$$

Results and Discussion

The Chrono Aqua OSR (NIRI) is an innovative mechatronic system designed for smart urban water conservation. It integrates real-time data monitoring and control to optimize water usage efficiently. Utilizing sensors, IoT technology, and customized algorithms, it dynamically adjusts water flow based on demand, reducing wastage. This system enhances sustain-

DOI: 10.36838/IJHSR812.70

ability by providing precise control and promoting responsible water management in urban environments. Its adaptability allows for integration into existing infrastructure, making it suitable for diverse applications. Chrono Aqua OSR (NIRI) exemplifies advanced technological solutions for addressing water scarcity challenges, contributing to smarter, eco-friendly urban water systems.

The experimental methods evaluate the Device's performance under different variable conditions. The key variables studied are:

- Variable 1: Impurities in water
- Variable 2: Temperature of water
- Variable 3: Calibration of the Device

Experiment 1: Effect of Impurities in Water:

Experiment 1 investigates how impurities like salt and dirt affect water flow rate. The variable is the presence of impurities. The aim is to determine whether impurities impact the flow rate, helping to understand how water purity influences the performance of flow measurement instruments. Two vessels were prepared, each containing 2 liters of water. One vessel held clean water, while the other contained water mixed with impurities, including a tablespoon of salt and some dirt. The prototype device was first tested with clean water, where the time required for 0.1 liters of water to flow through was measured and recorded. This measurement was repeated three times to minimize fluctuations and ensure accuracy. The same procedure was then conducted using the impure water, with the time and speed for 0.1 liters of flow carefully noted. Finally, the results from both setups were compared to analyze the effect of impurities on water flow speed and timing. Table 5 shows the flow measurement values for normal water. Over the course of three tests, the flow measurement data reveal normal water flow with varying speeds between 3.5 and 3.9 liters per hour. The duration varied between 1 minute 30 and 1 minute 40 seconds, suggesting steady flow rates under typical circumstances. Table 6 shows the flow measurement values for impure water. The impure water flow measurement shows somewhat lower flow rates, between 3.38 and 3.4 liters per hour. The timings, which varied from 1 minute 40 to 1 minute 47 seconds, demonstrated a steady but slightly lower flow than typical water, most likely as a result of contaminants influencing flow efficiency.

Table 5: Shows the flow measurement values for normal water.

Test Number	Time Taken	Speed
Test 1	1 min 30 s	3.5 L/h to 3.9 L/h (fluctuating)
Test 2	1 min 40 s	3.5 L/h
Test 3	1 min 40 s	3.5 L/h

Table 6: Shows the flow measurement values for impure water.

Test Number	Time Taken	Speed
Test 1	1 min 40 s	3.4 L/h
Test 2	1 min 45 s	3.4 L/h
Test 3	1 min 47 s	3.38 L/h

Initial fluctuations in the device's performance were caused by the inconsistent and variable power supply from a 9V battery. The issue was resolved by replacing it with a 12V DC motor. The overall speed of the prototype was relatively low because the 9V battery's limited power was divided among multiple components, including the flow meter, relay, and buzzer. In the experiment with normal water, the water flow remained steady, with consistent speed and timing across repeated tests. During the experiment with impure water, the results of Test 1 closely matched those observed with normal water. From Test 2 onwards, a slight buildup of impurities was noticed inside the motor and pipe, causing a slight decrease in flow speed and a slow increase in the time taken for water to pass. This buildup continued to increase in Test 3, further affecting the flow rate and timing.

Impurities in Water:

The effect of impurities in water on the Chrono Aqua OSR device was studied by comparing clean and impure water (containing salt and dirt). The results indicated that impurities significantly impacted the flow rate and speed through the device. A reduction in flow speed of approximately 0.1 L/h was observed when 0.1 liters of water were passed through the prototype, with a time variation of 5 to 7 seconds. The buildup of impurities in the flow meter and pipes caused a slight decrease in the water flow.

Experiment 2: Effect of Water Temperature on Flow and Speed:

To find the influence of water temperature on the flow rate and speed, three trials were conducted using hot (45°C), lukewarm (35°C), and cold (20°C) water. The results showed no significant flow rate or water speed change across the three temperatures. Small variations of 2–3 seconds in flow time and a 0.1 L/h change in speed were recorded in the Device, indicating that the Chrono Aqua OSR device operates efficiently across a range of typical household water temperatures. The slight changes could be attributed to the expansion of the flexible pipe under higher temperatures, affecting the flow characteristics.

Table 7 shows that hot water at 45°C recorded the fastest flow speed at 3.6 L/h, while cold water at 20°C had the slowest at 3.4 L/h, suggesting that higher water temperatures slightly improve flow efficiency due to reduced viscosity. The complete, functional Chrono Aqua OSR device. It demonstrates compact integration of all electronics and mechanical components for household installation. The data shows that flow speed drops marginally from 3.6 to 3.4 liters per hour as water temperature drops from 45°C to 20°C. Lower temperatures result in a modest decrease in flow rate and efficiency, as indicated by the corresponding slight increase in flow time.

Table 7: Shows the effect of water temperature on flow time and speed, showing a slight decrease in flow speed as water temperature decreases.

Temperature	Time Taken	Flow Speed (L/h)
Hot (45°C)	1 min 39 s	3.6
Lukewarm (35°C)	1 min 40 s	3.5
Cold (20°C)	1 min 41 s	3.4

Experiment 3: Effect of Flow Meter Calibration on Accuracy:

The flow meter was connected to a microcontroller (Arduino UNO). The flow meter's 5V and GND (ground) pins were attached to the corresponding pins on the Arduino, while the signal pin was connected to a designated digital input pin as specified in the code. A program was then written and uploaded to the Arduino, designed to count the number of pulses generated by the flow meter over a fixed time interval. A grad-

uated container was used to pre-measure exactly 0.1 liters of water to ensure consistency across trials.

The water flow was initiated and stabilized before pulse counting began. Once the flow was steady, pulse counting started simultaneously with a timer, allowing the flow to continue until exactly 0.1 liters of water had passed through the meter. Figure 8 summarizes the results of experiments assessing the impact of impurities, temperature, and calibration on device accuracy. It visually supports the reliability of performance under variable conditions. After the experiment, the number of pulses and the time taken for the flow of 0.1 liters were recorded. Flow rate is calculated as (Pulse Count × 60) divided by the calibration constant, converting pulses per minute to liters per hour. Rearranged, the calibration constant is (Pulse Count × 60) divided by flow rate. Given pulses per minute (0.43) and flow rate (3.5 L/h), the calibration constant is 7.5. Figure 8 (A) demonstrates that when impurity levels rise, water flow velocity falls. Impure water has the lowest flow rate, whereas normal water has the greatest. This suggests that contaminants obstruct water flow, resulting in a steady decrease in flow rate throughout tests, with the most notable decrease seen in the most contaminated sample. As the temperature lowers from hot to cold, Figure 8 (B) illustrates how the water flow speed falls. In every test, hot water has the fastest speed and cold water has the slowest, suggesting that lower temperatures cause water to flow more slowly. The pattern holds for several tests.

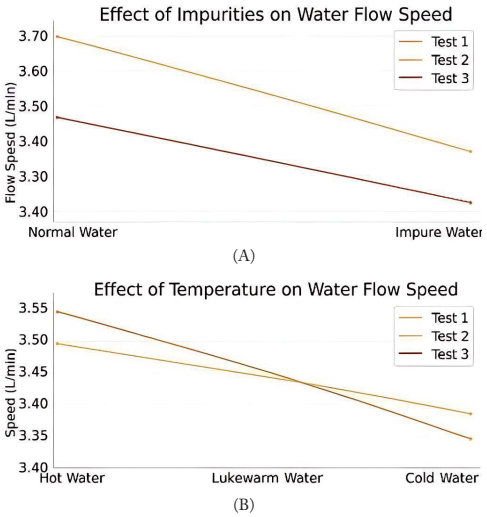


Figure 8: Shows the (A) and (B) are graphical representations of experiments with variables. (A) demonstrates that when impurity levels rise, water flow velocity falls. Impure water has the lowest flow rate, whereas normal water has the greatest. As the temperature lowers from hot to cold, (B) illustrates how the water flow speed falls. In every test, hot water has the fastest speed and cold water has the slowest, suggesting that lower temperatures cause water to flow more slowly.

Mode 1- NORMAL: Experiments to test flow rate and value:

Table 8 experiment with 9V external battery flow rate of pump shows inconsistent water flow rates ranging from 3 L/h to 7.5 L/h over 30-second intervals. These fluctuations indicate that a 9V external battery does not provide stable power to the pump, leading to unreliable and varied flow measurements. Inadequate power supply causes unstable and reduced water flow. While Arduino UNO provides 5V, it's not a reliable term. Using an external battery via breadboard or a 12V DC motor ensures a stable power source, improving flow consistency and device performance for better results. Water flow was measured in the experiment using an external 9V battery. Tests 1 and 2 recorded 3.6-4 L/h, 6-7.5 L/h, and 3-3.8 L/h, respectively. Various water flow behaviors with the battery configuration were shown by the various flow rates throughout testing.

Table 8: Shows the experiment with a 9V external battery.

Tests	Time Taken to Record the Flow	Flow of Water as Displayed on LCD
1	30 s	3.6 L/h - 4 L/h
2	30 s	6 L/h - 7.5 L/h
3	30 s	3 L/h - 3.8 L/h

Table 9 experiment with 12V external DC battery flow rate of pump demonstrates a significantly improved and more stable water flow, with flow rates ranging from 30 L/h to 35 L/h. A 12V external DC battery provided sufficient and consistent power, minimizing fluctuations and ensuring reliable pump performance across all tests. The flow speed had increased, but fluctuation was still observed. Modifications in the code were made to cancel false pulses from any external noise. Water flow was measured in the experiment using a 12V DC battery. While Test 2 fluctuated between 30 and 35 L/h, Tests 1 and 3 consistently displayed 30 L/h. The outcomes show steady flow rates with minor variations, indicating how well the battery powers the water flow system.

Table 9: Shows the experiment with a 12V External DC battery.

Tests	Time Taken to Record the Flow	Flow of Water as Displayed on LCD
1	30 s	30 L/h
2	30 s	30 L/h - 35 L/h
3	30 s	30 L/h

Table 10 experiment after code modifications using a 12V external battery confirms the success of the code optimization in achieving consistent flow measurements. All three tests showed a stable and uniform water flow rate of 30 L/h, indicating that software adjustments effectively eliminated false pulses and fluctuations. The flow rate is the same in different trials after changing the code and the power supply. Water flow was consistent in the modified code experiment using a 12V external battery. For 30 seconds, a constant flow rate of 30 L/h was measured in all three tests. Following code modifications with the 12V power source, these findings show increased stability and dependability in water measurement.

Table 10: Shows the experiment after code modifications using a 12V external battery.

Tests	Time Taken to Record the Flow	Flow of Water
1	30 s	30 L/h
2	30 s	30 L/h
3	30 s	30 L/h

Table 11 experiment to measure actual flow rate and water outflow reveals a discrepancy between the displayed flow rate and the actual volume of water measured using a graduated container. Although the flow rate consistently showed 30 L/h, the actual outflow was 0.5 L in each trial, highlighting inaccuracies likely due to incorrect calibration constants or sensor sensitivity, necessitating recalibration for improved accuracy. The experiment measured actual water flow over 60 seconds, with a consistent flow rate of 30 L/h. The total volume varied

DOI: 10.36838/IJHSR812.70

between 0.8 and 1.2 L, but the actual water flown was about 0.5 L in each test. This indicates some measurement discrepancies, but overall flow consistency.

Observation: Inaccuracy of the flow value displayed on the LCD, while the actual water outflow was the same

Solution: The calibration constant was 7 (as per data sheets). After experimentally calculating the calibration constant. New calibration constant- 7.5.

Table 11: Shows the experiment to measure the actual flow rate and water outflow.

Test	Time	Flow Rate of Water	Total Volume of Water	Actual Amount of Water Flown*
1	60 s	30 L/h	1.2 L	0.5 L
2	60 s	30 L/h	0.8 L	0.5 L
3	60 s	30 L/h	0.9 L	0.5 L

Table 12 presents an experiment to verify the water outflow display value using a new calibration constant (7.5), which demonstrates improved accuracy in flow readings. After recalibrating the flow meter using a constant of 7.5, the displayed water flow closely matched the actual measured volume in all tests, confirming that the calibration adjustment successfully corrected previous discrepancies. The experiment verified water outflow readings with a new calibration constant over 60 seconds. Displayed flow was around 30 L/h, matching the actual 0.5 L displaced in tests. Minor discrepancies suggest calibration improved accuracy, ensuring reliable water flow measurement.

Result: The flow rate is constant; the Water flow value displayed is correct and consistent.

Table 12: Shows the experiment to verify the water outflow display value using the new calibration constant.

Tests	Time	Flow of Water Displayed	Volume of Water Displaced	Actual Amount of Water
1	60 s	30 L/h	0.4 L	0.5 L
2	60 s	30 L/h	0.5 L	0.5 L
3	60 s	30 L/h	0.5 L	0.5 L

Mode 2 – ALERT MODE:

Table 13 experiments to test alarm activation after preset water limit confirm that the alarm system reliably activates after reaching the preset threshold. In both tests, the alarm rang exactly after 1 liter of water was used, with consistent flow rates of 30 L/h, validating the mode's accuracy and responsiveness in real-time monitoring. A minor fluctuation was observed in the value of water outflow. The test checked alarm activation after reaching a 1 L water limit. The alarm sounded after 120 seconds at a flow rate of 30 L/h, with 1 L displayed and flowed. The system successfully triggered the alarm upon reaching the preset water volume.

Table 13: Shows the test alarm activation after the preset water limit.

Tests	Limit Set	Time Taken for Alarm to Ring	Flow of Water Displayed	Total Volume Displayed When Alarm Activates	Amount of Water Flowed
1	1 L	120 s	30 L/h	1 L	1 L
2	1 L	120 s	30 L/h	1 L	1 L

Table 14 shows an accuracy check of the alert mode for a 0.1-liter limit, demonstrating that this device's alert system performs with high precision at small volumes. In both tests, the alarm triggered close to the preset 0.1 L threshold, with the actual flow ranging from 0.1 L to 0.12 L, confirming reliable detection and minimal deviation in real-time usage conditions. The accuracy check for alert mode tested activation at a 0.1 L limit. The alarm triggered within about 12-13 seconds at 30 L/h flow, with displayed and flowed water close to the set limit. Results indicate the alert mode reliably detects near-limit water volumes.

Table 14: Shows the accuracy check of the alert mode.

Tests	Limit Set	Time Taken for Alarm to Ring	Flow of Water Displayed	Total Volume Displayed When Alarm Activates	Amount of Water Flowed
1	0.1 L	13 s	30 L/h	0.1 L	0.12 L
2	0.1 L	12 s	30 L/h	0.1 L	0.1 L

Mode 3- REGULATION MODE: Experiment with the regulation system:

The setup was done to manually regulate the flow after a limit was reached using a potentiometer. A parallel circuit with relays and potentiometers was utilized to make the system automatic. Table 15 shows the effectiveness of automatic water flow regulation in the regulate mode, demonstrating that the system successfully reduces the flow rate once the preset limit of 0.5 liters is reached. In all three tests, the flow dropped from 30 L/h to approximately 15–16 L/h after 60 seconds, leading to a moderate total water volume (around 0.75–0.78 L), thus validating the effectiveness of the automatic regulation mechanism in conserving water beyond the set threshold. The result is regulation mode is functioning well. The effectiveness test assessed automatic water flow regulation at a 0.5 L limit. The alarm activated after 60 seconds at 30 L/h flow, with displayed and actual water volumes matching exactly. This confirms that the regulation mode accurately maintains and detects the set water volume.

Table 15: Shows the effectiveness of automatic water flow regulation in the regulation mode.

Tests	Limit Set	Time Taken for Alarm to Activate	Flow of Water Displayed	Total Volume Displayed When Alarm Activates	Actual Amount of Water That Flowed
1	0.5 L	60 s	30 L/h	0.5 L	0.5 L
2	0.5 L	60 s	30 L/h	0.5 L	0.5 L

Mode 4- Eco mode (1) - Set up a limit of 0.1 L using a solenoid valve and relay:

Table 16 shows the performance of the eco mode auto water cut-off after the preset limit, illustrating the Device's ability to stop water flow once the 0.1-liter limit is reached automatically. In all three tests, the flow rate of 30 L/h was consistently cut off within 13–14 seconds, confirming the eco mode's effectiveness in enforcing timely water conservation. The result is eco mode is functioning well, as water flow stops after 0.1 L in 14 sec. The eco mode auto cut-off test showed water flow reduction after 60 seconds at a 0.5 L limit. Flow decreased from 30 L/h to around 15-16 L/h, with total water usage around 0.75-0.78 L. This demonstrates effective automatic regulation to conserve water after reaching the preset limit.

Table 16: Shows the performance of the eco mode auto water cut-off after the preset limit.

Tests	Limit Set	Time After Which Water Flow Is Reduced	Flow of Water Before Regulation	Flow of Water After Regulation (for 30 s)	Total Volume of Water
1	0.5 L	60 s	30 L/h	16 L/h	0.78 L
2	0.5 L	60 s	30 L/h	15 L/h	0.75 L
3	0.5 L	60 s	30 L/h	15 L/h	0.75 L

Mode 5- Vacation mode preset threshold leakage amount of 1L in the code:

Table 17 explained detected continuous water flow (simulating a leak) and automatically blocked it after approximately 120–121 seconds at a flow rate of 30 L/h. This confirms the vacation mode's reliability in preventing water loss during prolonged user absence. Result: water leakage stopped after 2 minutes (120 seconds); the quantity of water leaked is 1 liter. The test detected continuous water flow at 30 liters/hour. Water flow persisted for approximately 120-121 seconds before stopping, demonstrating the system's ability to monitor and detect continuous water flow effectively. This indicates reliable performance in identifying sustained water movement.

Table 17: Shows the detected continuous water flow.

Test	Flow of Water	Time After Which the Flow Stops
1	30 Litre/hour	120 seconds
2	30 Litre/hour	121 seconds
3	30 Litre/hour	121 seconds

Pilot study:

Table 18 explains reduced water usage and billing after installing the NIRI instrument as a part of a pilot study in April and May of 2025. Water usage dropped by 4,537 liters from the month of March to May, saving approximately ₹504.21, with additional daily savings from leak control, indicating NIRI's effectiveness in promoting conservation and lowering utility costs. Figure 9 shows the comparative analysis of water usage and Bill (with and without NIRI). Installing the NIRI instrument reduced water usage and bills. In April and May, water consumption dropped significantly compared to March, saving around 3,749-4,537 liters and ₹265-₹504 monthly. Daily savings of 125 liters were achieved, with NIRI active, indicating effective water conservation. The figure explains that water usage and bills significantly decreased after installing the NIRI device. From March to May, water consumption dropped from 19,429 liters to 14,892 liters, and water bills reduced from ₹2,142 to ₹1,638, demonstrating effective water conservation and cost savings with the NIRI system.

Table 18: Shows the reduced water usage and billing after installing the NIRI instrument.

Months	Water Used (litres)	Water Bill (INR/USD)	Savings (Litres versus March details)	Savings (INR)	Daily Savings	NIRI Status
March	19,429	₹2,142.33 (25 USD)	Not Applicable	Not Applicable	Not Applicable	Not used
April	15,680	₹1,877.32 (20 USD)	3,749 L	₹265.01 (4 USD)	125 L/day (shower)	NIRI Active
May	14,892	₹1,638.12 (18 USD)	4,537 L	₹504.21 (6 USD)	125 L/day (shower) + 788 L (Leak Control)	NIRI Active

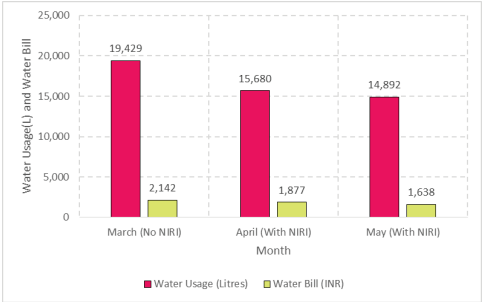


Figure 9: Shows the comparative analysis of water usage and Bill (with and without NIRI),showing a 20% reduction in water use.

Table 19 shows the inferred annual water and cost saving potential of the device, highlighting that by saving 200 liters per day through optimized shower use and leak control, a household can conserve up to 73,000 liters annually, resulting in an estimated ₹8,030/year (100 USD) in cost savings. This demonstrates the significant impact of the device in reducing both water consumption and utility expenses. The device's inferred annual savings indicate a potential of saving 73,000 liters of water and ₹8,030 (about $100 USD) in costs. By controlling leaks and optimizing water use in showers and main lines, significant water and money can be conserved annually.

Table 19: Shows the inferred annual water and cost saving potential of the device.

Water Saving Potential (inferred)	Location	Annual Water Saved	Annual Cost Saved
200 L/day (Shower faucet + Main line leak control)	Shower faucet + Main line (leak control)	73,000 L	₹8,030/year (100 USD)

Table 20 shows the projected large-scale impact of device deployment. This table illustrates the broader effect of implementing the device across 1,000 households (from 10 cities × 100 societies × 10 houses each). It projects an annual water saving of 73 million liters, translating to a cost saving of ₹8,000,000 (approximately USD 93,848). All images emphasize the substantial environmental and economic benefits of scaled adoption. Implementing the device in 1,000 households across 10 cities can save 73 million liters of water annually, reducing costs by ₹8 million ($93,848). This large-scale adoption significantly conserves water resources and offers substantial financial savings across communities.

Table 20: Shows the broader effect of implementing the device across 1,000 households.

Scope	Households	Annual Water Saved (Liters)	Annual Cost Saved (INR)	Annual Cost Saved (USD)
10 cities × 100 societies × 10 houses each	1,000	73,000,000 L	₹8,000,000	$93,848

Conclusion

NIRI Chrono Aqua OSR device demonstrates significant potential as a scalable, cost-effective intervention for urban domestic water conservation. A structured experimental protocol was implemented to evaluate system performance across its five operational modes: **Normal, Alert, Regulation, Eco, and Vacation**. Each mode was tested under controlled household-simulated flow conditions to assess accuracy, responsiveness, and functional reliability.

In **Normal Mode**, volumetric flow measurement accuracy was validated against a calibrated reference container using repeated trials ($n \geq 20$). Statistical analysis indicated consistent linearity between actual and displayed flow values following recalibration. Experimental refinement yielded an optimized calibration constant of 7.5, reducing measurement deviation and improving inter-trial consistency.

In **Alert Mode**, the device reliably triggered auditory signals upon reaching predefined consumption thresholds, including a 10 L pre-threshold warning buffer. Response latency remained within operational tolerance limits (<1 second delay).

In **Regulation Mode**, post-threshold flow reduction was achieved through controlled mechanical constriction, resulting

DOI: 10.36838/IJHSR812.70

in approximately 50% reduction in discharge rate relative to baseline conditions. Flow normalization required a manual reset, confirming deterministic control logic.

In **Eco Mode**, complete flow termination occurred immediately upon threshold breach, with restart governed by preset temporal conditions, ensuring enforced conservation compliance.

In **Vacation Mode**, leak detection sensitivity was tested under micro-leak simulations (low continuous discharge conditions), demonstrating reliable anomaly detection based on sustained non-zero flow without active usage input.

Environmental robustness tests evaluated the influence of temperature variation and particulate impurities on flow velocity and sensor accuracy. While effects were statistically minimal, measurable deviations highlighted the importance of integrated filtration or debris-resistant turbine design to maintain long-term precision. Electrical optimization of the power supply system, combined with firmware-level pulse filtering algorithms, significantly reduced signal noise and improved measurement stability.

Pilot household deployment studies demonstrated measurable reductions in monthly water consumption of up to **4,537 liters per household**, corresponding to both financial savings and reduced freshwater extraction. Extrapolated to large-scale urban implementation, projected annual savings could reach approximately **73 million liters of water and ₹8 million (≈ USD 93,848)** in utility cost reductions, assuming moderate adoption rates.

Collectively, the Chrono Aqua OSR addresses critical gaps in existing domestic water-saving technologies by integrating real-time volumetric monitoring, automated leak detection, adaptive consumption regulation, and affordability within a compact, tap-mounted architecture. Its multi-modal control framework positions it as both a behavioral intervention tool and an engineering solution for sustainable urban water management.

Recommendations

- Pilot Testing & Feedback: Broaden field testing in diverse urban contexts to guide iterative design enhancements.
- Filtration Integration: Include filters to prevent clogging from impurities.
- Expanded Customization: Allow user-defined settings based on usage and tariff patterns.
- Awareness Campaigns: Launch parallel educational initiatives to promote responsible water use.
- Municipal Collaboration: Work with local authorities to subsidize or promote the Device via innovative city programs.
- Scalability & Robustness: Invest in improving battery life, connectivity, and hardware durability to support widespread adoption.

Acknowledgments

I express heartfelt gratitude to Ms. Kalai Selvi, Head of Vidyashilp Academy, for her inspiring mentorship and the opportunity to work in the School ATL Lab, where this project began. I am especially thankful to Mr. Aravindhan at the ATL Lab for his expert guidance and support in developing the NIRI - Chrono Aqua OSR device.

Above all, I am deeply grateful to the Almighty for His blessings and guidance in making this work possible.

References

1. Kamyab, H.; Khademi, T.; Chelliapan, S.; SaberiKamarposhti, M.; Rezania, S.; Yusuf, M.; Farajnezhad, M.; Abbas, M.; Hun Jeon, B.; Ahn, Y. The Latest Innovative Avenues for the Utilization of Artificial Intelligence and Big Data Analytics in Water Resource Management. *Results Eng*. **2023**. https://doi.org/10.1016/j.rineng.2023.101566.
2. Pravallika, D.; Prathyusha, D.; Srinivasa Kumar Professor, D. IoT Based Water Level Monitoring System with an Android Application. *Ijritcc* **2018**.
3. Chen, H.; He, H.; You, J.; Xie, X.; Fang, G.; Xiao, P. A Study on Urban Household Water Consumption Behavior under Drought Conditions. *J. Environ. Manage*. **2023**. https://doi.org/10.1016/j.jenvman.2023.118963.
4. Michael Ayorinde Dada; Michael Tega Majemite; Alexander Obaigbena; Onyeka Henry Daraojimba; Johnson Sunday Oliha; Zamathula Queen Sikhakhane Nwokediegwu. Review of Smart Water Management: IoT and AI in Water and Wastewater Treatment. *World J. Adv. Res. Rev*. **2024**. https://doi.org/10.30574/wjarr.2024.21.1.0171.
5. Rodríguez Montoya, C. A Taxonomy of Demand Management Strategies for Sustainable Water Consumption in Urban Households. *Urban Water J.* **2024**. https://doi.org/10.1080/1573062X.2024.2314652.
6. Aydamo, A. A.; Robele Gari, S.; Mereta, S. T. Seasonal Variations in Household Water Use, Microbiological Water Quality, and Challenges to the Provision of Adequate Drinking Water: A Case of Peri-Urban and Informal Settlements of Hosanna Town, Southern Ethiopia. *Environ. Health Insights* **2024**. https://doi.org/10.1177/11786302241238940.
7. Santos, J. C.; Allison, A. L.; Jankovic-Nisic, B.; Campos, L. C. Impact of Behavioural Factors on the Household Water Consumption in Urban Areas. *Proc. Inst. Civ. Eng. Munic. Eng*. **2022**. https://doi.org/10.1680/jmuen.21.00032.
8. Laha, S. R.; Pattanayak, B. K.; Pattnaik, S. Advancement of Environmental Monitoring System Using IoT and Sensor: A Comprehensive Analysis. *AIMS Environ. Sci.* **2022**. https://doi.org/10.3934/environsci.2022044.
9. Rahman, H. A. A.; Al-Farsi, H. A.; Ahmed, M.; Goosen, M. F. A. Evaluation of Some Water Saving Devices in Urban Areas: A Case Study from the Sultanate of Oman. *J. Agric. Mar. Sci. [JAMS]* **2018**. https://doi.org/10.24200/jams.vol22iss1pp18-26.
10. Casazza, M.; Xue, J.; Du, S.; Liu, G.; Ulgiati, S. Simulations of Scenarios for Urban Household Water and Energy Consumption. *PLoS One* **2021**. https://doi.org/10.1371/journal.pone.0249781.
11. Danquah, L.; Awuah, E.; Agyemang, S.; Mensah, C. M. Investigating the Predictors of Domestic Water Consumption in Urban Households with Children Under-Five Years: A Panel Study in the Atwima Nwabiagya District, Ghana. *J. Sustain. Dev*. **2015**. https://doi.org/10.5539/jsd.v8n8p1.

12. Adedotun, S. B.; Ogundahunsi, D. S.; Ibrahim, R. B.; Adedotun, D. O.; Yakubu, D. A. Analysis of Households' Water Access and Consumption in Differential Urban Neighbourhoods of Osogbo, Nigeria. *Ghana J. Geogr*. **2024**. https://doi.org/10.4314/gjg.v16i1.9.
13. Rebouças, R. da S. de O.; Soares, M. D. R.; Noguchi, H. S.; De Souza, M. S.; do Nascimento, F. R.; Alves, K. de V.; Pantoja, L. P.; De Souza, Z. M. Water Quality for Human Consumption in Semi Artesian Wells in the City of Lábrea/AM. *Contrib. A LAS CIENCIAS Soc*. **2024**. https://doi.org/10.55905/revconv.17n.1-316.
14. Palermo, S. A.; Maiolo, M.; Brusco, A. C.; Turco, M.; Pirouz, B.; Greco, E.; Spezzano, G.; Piro, P. Smart Technologies for Water Resource Management: An Overview. *Sensors*. **2022**. https://doi.org/10.3390/s22166225.
15. Carriazo-Regino, Y.; Baena-Navarro, R.; Torres-Hoyos, F.; Vergara-Villadiego, J.; Roa-Prada, S. IoT-Based Drinking Water Quality Measurement: Systematic Literature Review. *Indones. J. Electr. Eng. Comput. Sci.* **2022**. https://doi.org/10.11591/ijeecs.v28.i1.pp405-418.
16. Mirauda, D.; Erra, U.; Agatiello, R.; Cerverizzo, M. Applications of Mobile Augmented Reality to Water Resources Management. *Water (Switzerland)* **2017**. https://doi.org/10.3390/w9090699.
17. Chen, J. F.; Liao, Y. T.; Wang, P. C. Development and Deployment of a Virtual Water Gauge System Utilizing the ResNet-50 Convolutional Neural Network for Real-Time River Water Level Monitoring: A Case Study of the Keelung River in Taiwan. *Water (Switzerland)* **2024**. https://doi.org/10.3390/w16010158.
18. Al-Naemi, S.; Al-Otoom, A. Smart Sustainable Greenhouses Utilizing Microcontroller and IoT in the GCC Countries; Energy Requirements & Economical Analyses Study for a Concept Model in the State of Qatar. *Results Eng*. **2023**. https://doi.org/10.1016/j.rineng.2023.100889.
19. Jamaaluddin; Akbar, A.; Khoiri. Ultrasonic Flow Meters and Microcontrollers for Precise Water Management with 6.45% Error Margin. In *IOP Conference Series: Earth and Environmental Science*; **2023**. https://doi.org/10.1088/1755-1315/1242/1/012017.
20. Gautam, G.; Sharma, G.; Magar, B. T.; Shrestha, B.; Cho, S.; Seo, C. Usage of IoT Framework in Water Supply Management for Smart City in Nepal. *Appl. Sci*. **2021**. https://doi.org/10.3390/app11125662.
21. Sood, R.; Kaur, M.; Lenka, H. Design and Development of Automatic Water Flowmeter. *Int. J. Comput. Sci. Eng. Appl*. **2013**. https://doi.org/10.5121/ijcsea.2013.3306.
22. Lamprom, W.; Jotaworn, S.; Iamsomboon, N.; Bhumkittipich, P.; Siramaneerat, I.; Rukwong, A. Exploration of Wastewater Management Behavior for Enhancing Water Conservation in Urban Area, Thailand. *AIMS Environ. Sci.* **2022**. https://doi.org/10.3934/environsci.2022005.
23. Wang, B.; Niu, J.; Berndtsson, R.; Zhang, L.; Chen, X.; Li, X.; Zhu, Z. Efficient Organic Mulch Thickness for Soil and Water Conservation in Urban Areas. *Sci. Rep*. **2021**. https://doi.org/10.1038/s41598-021-85343-x.
24. Yue, C.; Cui, M.; Kong, X.; Watkins, E.; Barnes, M. Landscape Irrigation and Water Conservation in Urban Areas: An Analysis of Information-Based Strategies. *Horttechnology* **2022**. https://doi.org/10.21273/HORTTECH05001-21.
25. Ali, M.; Munala, G.; Muhoro, T.; Shikuku, J.; Nyakundi, V.; Gremley, A. Water Usage Patterns and Water Saving Devices in Households: A Case of Eastleigh, Nairobi. *J. Water Resour. Prot*. **2020**. https://doi.org/10.4236/jwarp.2020.124018.

Author

Aadya Kanchan, a Grade 11 student at Vidyashilp Academy, Bengaluru, India, is passionate about mechatronics, sustainable innovation, and real-world problem-solving. Driven by curiosity, creativity, and critical thinking, I develop practical solutions through hands-on experimentation. My background in debating and writing strengthens my ability to communicate complex scientific ideas clearly and persuasively. Science and technology are powerful tools to simplify complexity and drive meaningful societal change.

DOI: 10.36838/IJHSR812.70

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Understanding Financial Inclusion: Patterns and Determinants of Formal Borrowing in India

Aaran Nihalani

Eton College, Windsor, Berkshire, SL4 6DW, UK; aaran.nihalani@gmail.com

ABSTRACT: Financial inclusion plays a major role in household welfare and economic freedom in India, where borrowing is often needed to finance consumption, agriculture, housing, and emergencies like health shocks. The paper examines the household characteristics' association with formal credit access and documents the structural patterns that continue to shape the borrowing landscape across the country. The study exploits the nationally representative All India Debt and Investment Survey (AIDIS), 2019, employing OLS and Logit regression models with fixed effects at the state and district level. MPCE, a proxy for well-being, is positively associated with an increase in formal borrowing by ~1% per thousand rupees. Household size has a small negative effect on formal borrowing. The SCs and STs are less likely to avail credit from formal sources compared to OBCs, whereas the 'Others' caste is more likely to avail credit than OBCs. Muslim and Christian households are also less likely to have access to formal loans. The study highlights the role of loan purpose and relative urgency in shaping borrowing behavior. Addressing structural inequality in the credit market will require policies that explicitly recognize and counteract social disadvantage, rather than assuming that broader financial development will automatically translate into equal access.

KEYWORDS: Economics, Econometrics, Financial Inclusion, Household Credit Behavior, Formal and Informal Borrowing.

Introduction

Financial inclusion plays a major role in household welfare and economic freedom in India, where borrowing is often needed to finance consumption, agriculture, housing, and emergencies like health shocks. Despite significant policy changes over decades, Indian households, across all groups, continue to borrow heavily from informal sources, with the determinants of formal borrowing unclear and varied across regions and socio-economic groups.[1] Understanding why certain households decide to go with formal credit while others choose informal credit is thus essential for identifying long-standing problems in India's financial system.

Existing literature highlights the complexity of credit access in developing economies. Research on microfinance shows that while expanding credit may increase borrowing activity, it does not necessarily translate into sustained or significant improvements in income or consumption. Loan design can constrain productive investment due to short-term repayments and the small volume of these loans.[1-3] Studies specific to India document substantial diversity in borrowing behavior driven by demographic and economic characteristics, including caste, landholding, income, and education.[4] Work on repayment structures further shows that rigid or frequent repayment requirements can discourage entrepreneurship among low-income households, who may not have enough time to deploy the credit before payment deadlines arrive meaningfully.[3,5] Criticisms of the current microfinance model in India include the much higher interest rates that impact the viability of borrowing, continued limited reach to the poorest, adverse selection, and weak infrastructure for expanding businesses.[1-3] A lack of collateral also drives people towards informal borrowing, and despite land being theoretically ideal as collateral, land records are often poorly maintained, impacting whether or not people can use land as collateral. This leads to land being underutilized as collateral, forcing more households to borrow informally.[6] Demand-side analysis indicates small but unimpressive improvements in financial inclusion, with rural households and disadvantaged groups continuing to face persistent frictions.[7]

Earlier empirical evidence highlights the importance of institutional finance for rural development, especially during the mid-20th century, when India made many changes to reduce the share of informal credit in the lending market. Increases in access to formal rural credit have historically supported agricultural productivity and poverty reduction, with the percentage of loans coming from moneylenders declining from 83% in 1951 to 36% in 1971. The AIDIS data used in this study shows that this number has further reduced to 22% in rural areas.[8,9] However, more recent reassessments question the robustness or magnitude of these effects, suggesting that outcomes may depend heavily on data quality, regional variation, and model specification.[10] Complementary research shows that institutional failures, such as corruption in the allocation of formal credit and the relative inconvenience of choosing a formal lender, may push households toward informal lenders even when formal channels are present.[11] Taken together, the literature suggests that both household characteristics and institutional constraints shape borrowing decisions, yet contemporary evidence using detailed, nationally representative microdata remains limited.

Recent contributions in *Economic and Political Weekly* argue that headline gains in financial inclusion, such as expanded banking infrastructure and account ownership, have not necessarily translated into substantively equitable credit access.

DOI: 10.36838/IJHSR812.81

Tagade and Reza highlight persistent structural gaps in meaningful participation within the formal financial system, while Chavan emphasizes that aggregate inclusion metrics often obscure disparities in usage patterns and borrower vulnerability across social groups.[12,13] These insights reinforce the need to examine disaggregated patterns of formal borrowing by caste, religion, and region, which the present study undertakes using nationally representative microdata.

This study fills this gap by analyzing borrowing patterns using the 2019 All India Debt and Investment Survey (AIDIS), one of the most comprehensive and recent datasets available on household indebtedness in India. The analysis examines how demographic and economic characteristics (including consumption levels, household structure, land ownership, caste, religion, and location) influence the probability of taking a formal loan as opposed to an informal one. Five econometric models are estimated, using logit and ordinary least squares (OLS) regression with variation of state and district-level fixed effects to control for unobserved regional heterogeneity. Average Marginal Effects are used for interpretability instead of log-odds coefficients. The findings reveal clear differences in borrowing determinants across households and substantial variation in both loan purposes and sources between rural and urban sectors. We find that groups including SC, ST, and Christians are less likely to borrow formally, indicating lower availability of formal sources, showing that structural inequality in India persists through the credit system. Existing literature comparing the propensity of different social groups to borrow formally ranks access to formal credit in agriculture as Others, OBC, SC, and ST.[14] This study finds that Others and OBC are, similarly, the top two, but puts SC below ST in terms of access to formal credit in general, as opposed to just agriculture, supported by average marginal effects (AME) and OLS coefficients.

Structural barriers to formal credit access operate through multiple channels: collateral constraints linked to land titling, historical exclusion from institutional banking networks, informational asymmetries, social discrimination in screening processes, and geographic clustering of financial infrastructure. These mechanisms imply that disparities in formal borrowing are not solely income-driven but embedded within broader institutional arrangements.

By providing updated national evidence, isolating within-region variation, and comparing formal and informal borrowing patterns at a granular level, this study contributes to the literature by offering a contemporary assessment of demand-side financial inclusion in India. It clarifies which household characteristics most strongly predict formal credit access and documents the structural patterns that continue to shape the borrowing landscape across the country.

The paper is structured with Section 2 describing the dataset being used in the study, the All India Debt and Investment Survey, including the details of the survey and the contents of the data. It then looks at the details of the two regression models used, OLS and Logit regression. Next, Section 3 contains the results and analysis of the paper's findings, bringing insights from the data with figures and tables. Finally, there is a conclusion to summarize the points made and the findings of the paper.

Methods and Data

Data:

AIDIS, or the All India Debt and Investment Survey, is a nationally representative survey conducted by India's National Sample Survey Office (NSSO).[15] AIDIS data was collected for the sixth time by the NSSO in 2019, during the National Sample Survey's 77th Round. It was first conducted in 1951, and it has been conducted roughly every 10 years since. AIDIS offers data on asset stock, indebtedness, capital formation, and indicators of rural/urban economies.[16] The survey covered 5,940 villages and 69,455 households in rural sectors and 3,995 blocks and 47,006 households in urban sectors.

The study was sampled through a stratified two-part sample, using simple random sampling without replacement for the villages/blocks in the rural or urban sectors. After this, simple random sampling was used once more for each of the strata based on Monthly Per Capita Consumer Expenditure (MPCE), used as a proxy for well-being in the survey. The survey covers the entirety of the Indian Union, excluding only inaccessible villages in the Andaman and Nicobar Islands. The data was verified post-collection to address errors, and outliers, based on MPCE, were removed from this study, reducing the number of observations by 0.2%.

The data comes with 18 datasets, or 'Levels', all focusing on different areas of the survey. For this study, we use Level 2, which addresses demographic and particulars of household members, Levels 3 and 4, which address household characteristics, and Level 14, which contains the data for loans taken by each household from both formal and informal sources. In this study, we load the datasets into individual R data frames before merging them into one large data frame, with Household ID as the variable being merged upon.

Other datasets, such as the India Human Development Survey (IHDS) from 2012 and the National Bank for Agriculture and Rural Development (NABARD) All-India Rural Financial Inclusion Survey (NAFIS), were considered.17,18 IHDS was not used because its loan data is limited and less detailed for household-level analysis. NAFIS focuses on rural data, so it was not suitable for a study comparing rural and urban borrowing patterns.

Methods:

The study employs OLS regression and Logit regression models with fixed effects at the state and district level, using nationally representative data on Loans in India. OLS regression calculates the response variable using Equation 1, with a closed-form solution for the coefficients given by Equation 2.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \epsilon \quad (1)$$

Equation 1

$$\beta = (X^T X)^{-1} X^T y \quad (2)$$

Equation 2

The first equation shows how response variables are produced during regression, with the β terms being the coefficients that the explanatory variables are multiplied by to arrive at the final response variable, with the ϵ representing the error term, which OLS aims to minimize, taking the sum of the squares of the residuals as its loss function.[19] The second equation offers an analytical solution to find the coefficients. However, as X increases in size, it becomes computationally intensive to invert the matrix, if it even is invertible, with a computational complexity of $O(n^3)$. This study also uses Logit Regression, as OLS has a limitation in that it is purely linear in X, and is not bounded between 0 and 1 like Logit is, which is necessary for a binary response variable. Therefore, Logit Regression is used to output values that are to be interpreted as probabilities. OLS is used for continuous outcomes, while Logit Regression uses Equation 3 to model the log-odds of the event's probability:

$$P(y = 1 \mid x) = \frac{exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ . \ . \ . \ + \beta_p x_p)}{1 + exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ . \ . \ . \ + \beta_p x_p)} \quad (3)$$

Equation 3

This study uses 5 models, shown in Table 1, to analyze different aspects of the AIDIS loan data. Logit Regression is used here to show the probability of borrowers taking a loan based on a certain characteristic, like their expenditure or household size. This allows for comparing how different groups of people borrow. Fixed effects are used at both state and district levels to control for unobserved regionally varying factors.[20] State-level effects control for differences in economic conditions or policies state by state, and district-level effects absorb bias from localized differences, like access to banks. This isolates the within-region variation for more consistent results in the regression. This ensures that the estimated relationships between variables and outcomes reflect variation within regions instead of between them.

Table 1: List of models being used in the study.

Model 1	Logit Regression on MPCE (Monthly Per-Capita Consumer Expenditure)
Model 2	Logit Regression on all characteristics, e.g., household size, land owned, caste
Model 3	Logit Regression on all characteristics with state-level fixed effects
Model 4	OLS Regression on all characteristics with district-level fixed effects
Model 5	Logit Regression on all characteristics with district-level fixed effects

While log odds offer a comparative measure of likelihood for borrowing characteristics, they are not intuitive and are difficult to interpret. A more understandable alternative is using Average Marginal Effects.[21] AME tells us what the expected impact on the outcome is when an independent variable is increased by one unit, averaged over all observations. This gives us an idea of how much of an impact each variable has on the final probability of borrowing, to see which factors are the most significant for identifying borrowing trends.

Replication/Reproducibility:

Replication code and data for this paper can be found at https://github.com/AaranNihalani/AIDIS.

Results and Discussion

Sample Characteristics:

Table 2: Summary of sample characteristics in AIDIS data. We see an even gender distribution and a majority of Hindus. Most respondents are located in rural areas, and loans are evenly distributed between formal and informal.

Summary Characteristics (AIDIS 2019)	
Variable	Value
Gender	
Male (%)	51.1
Female (%)	48.9
Religion	
Hinduism (%)	78.5
Islam (%)	12.6
Christianity (%)	4.9
Other religion (%)	4.0
Caste	
OBC (%)	44.2
Other caste (%)	27.8
SC (%)	16.7
ST (%)	11.4
Sector	
Rural (%)	62.9
Urban (%)	37.1
Loan Type	
Formal Loan (%)	49.8
Informal Loan (%)	50.2
Household Attributes	
Avg age (years)	30.6
Avg HH size (people)	5.5
Avg land owned (acres)	0.2
Avg MPCE (₹)	13,571.0
No. of Observations	480,687.0
Source: Author's Calculations from AIDIS, 2019	

Table 2 provides a summary of the observations made in the AIDIS data, where each observation is a loan. There is a very even distribution between male and female borrowers, ensuring we get representative data. The religion of borrowers is mostly Hindu, as is to be expected, with the next major groups being Islam and Christianity, with proportions consistent with official census numbers. With regards to caste, the largest portion is OBC (Other Backward Class), and the proportions of each caste are representative of the entire population. There are more rural respondents than urban respondents, but this is due to 63% of India's population being rural, consistent with AIDIS data.

Loan Characteristics:

Table 3: Loan data characteristics, by lender type. Formal loans are larger on average than informal loans. Interest for informal loans is almost twice as much, on average, as the interest rate of formal loans.

Loan Characteristics by Lender Type (AIDIS 2019)

Weighted using survey multiplier (MLT.x/100)

Lender Type	Number of Loans	% of Total Loans	Mean Amount (₹)	Median Amount (₹)	Mean Interest (%)	Median Interest (%)	Mean Outstanding (₹)	Median Outstanding (₹)
Formal	237,757	49.9%	225,774.39	70,000.00	10.52	10.00	203,430	55,120
Informal	238,483	50.1%	73,755.89	32,000.00	18.65	16.00	71,824	29,000

Source: Author's Calculations from AIDIS, 2019

DOI: 10.36838/IJHSR812.81

Table 3 offers a comparison of formal and informal loans. We have an even number of formal and informal loans available to us in the dataset, and the mean amount of each loan is much higher for formal loans, as would be expected. With the average formal loan over 250% greater than the average informal loan, borrower behavior for larger loans is very clear, with farm loans typically going through formal sources. Informal loans also have much higher interest rates, which is another reason why borrowers prefer formal sources of lending when larger amounts or less urgent matters are at hand.

Seeing the differences between the two loan types raises questions about other differences there may be between borrowers, in particular, their characteristics. To compare borrowers with and without a given characteristic, this study uses two-sample t-tests to evaluate whether the mean value of the characteristic differs across the two groups. The null hypothesis is that the true group means are equal. Rejecting this hypothesis indicates that the characteristic is statistically associated with formal borrowing and that the difference in the means isn't due to pure chance within the sampling methods. Regression is another way to investigate the characteristics that pertain to formal borrowing, using coefficients to see expected impacts on chosen borrowing source based on changes in relevant characteristics, which this study uses to understand the precise effect that each characteristic has on the choice of formal or informal borrowing sources.

Loan Distributions:

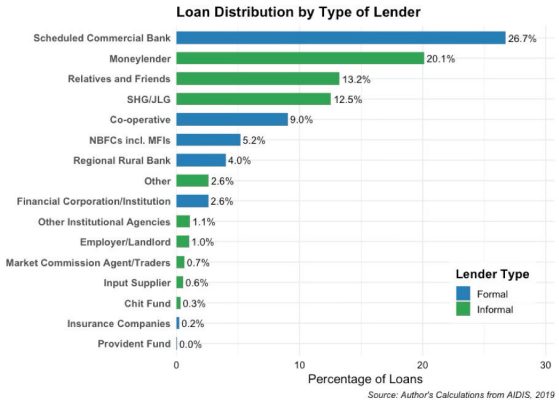


Figure 1: Distribution of loans by type of lender - both formal and informal sources. Scheduled commercial banks are the most common formal source, and moneylenders are the most common informal source.

Loans are distributed relatively evenly between formal and informal sources, as shown by Figure 1, where the formal lenders (shown in blue) make up 44.7% of loans taken (the sum of the blue formal bars), dominated by scheduled commercial banks, which have more capital on hand than many of the other sources, therefore attracting borrowers looking for larger loans, for businesses or farms. Moneylenders are the most common informal source of loans, as they are often the most accessible, through personal relations or just by being based in the same village as many borrowers. As expected, relatives and friends are common informal lenders, while SHGs (self-help groups) and JLGs (joint liability groups) facilitate small-scale credit through group responsibility. It is clear from Figure 1 that many Indians choose to borrow from community-based sources for credit access.

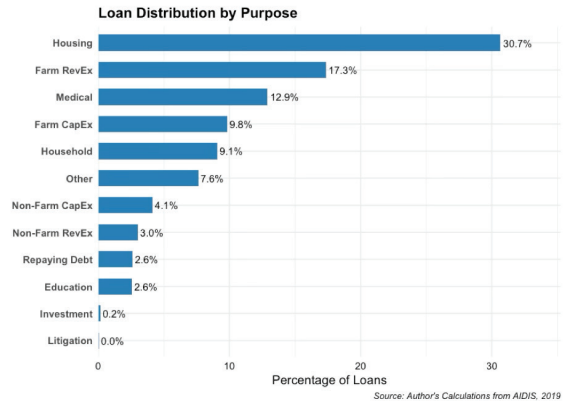


Figure 2: Distribution of loans by purpose. Housing is the largest cited purpose of loans, and farm-related loans form over a quarter of all loans.

Understanding the reason for loans being taken out allows us to draw conclusions about the behavior of borrowers in India and identify trends. To that end, Figure 2 gives us information about the purpose for each loan, weighted with the multipliers provided by NSS. Housing makes up the largest share of loan purposes, with Farm Revenue Expenditure and Medical expenses being the 2nd and 3rd most common loan purposes. The high proportion of Farm-related expenditure shows the scale of the rural population taking loans, as Farm expenditure forms over a quarter of all loans. This highlights the Indian economy's dependency on agriculture and the scale of the enterprise that leads to so many loans being taken.

T-Tests:

Table 4: T-test results table. Land ownership, household size, and MPCE exhibit statistically significant differences in mean values between households that borrow formally and those that borrow informally.

Independent-Samples *t*-Test Results by Formal Lending Status

Outcome Variable	*M* (No formal lending)	*M* (Formal lending)	*t*	*df*	*p*	95% CI
Land ownership	0.217	0.225	−5.916	437,546	3.304e-09	[-0.00954, -0.00479]
Household size	5.448	5.687	−30.671	421,707	< 2.2e-16	[-0.25452, -0.22394]
Monthly per-capita expenditure (MPCE)	11,897.740	15,831.670	−134.550	329,388	< 2.2e-16	[-3991.24, -3876.62]

Note. Group 0 = households without formal lending; Group 1 = households with formal lending.

Table 4 shows the results of statistical t-tests, used to evaluate the null hypothesis, which states that, in this case, the characteristics we evaluate don't have any association with formal borrowing. All three of the t-tests have a p-value less than 0.05, by many orders of magnitude, which allows us to reject the null hypothesis and deduce that the means of the groups of formal and informal borrowers differ.

Formal loan borrowers own more land on average than informal loan borrowers, with 0.217 acres for informal as compared to 0.225 acres for formal, t(427536) = -5.9, p = 3.3e-9. Formal borrowers also tend to have larger household size, with an average of 5.69 people, while informal borrowers have 5.45, t(421707) = 30.6, p < 2.2e16. Formal borrowers also have higher Monthly Per Capita Expenditure, a proxy for wellbeing, at 15832 rupees, while informal borrowers have 11898, 33% higher among formal borrowers, t(329388) = -134.6, p < 2.2e-16.

Land ownership has p < 0.01, indicating that there is a statistically significant difference between land ownership of

borrowers who borrow formally compared with informal borrowers. Since the mean in the informal borrowing group is less than the mean of the formal borrowing group, we can deduce that formal borrowers have higher land ownership than informal borrowers, although the magnitude of the difference is small. This t-test tells us that formal loan borrowers have more land holdings than informal loan borrowers. Landholding also facilitates loans, as it can be used as a form of collateral.

Household size, again $p < 0.01$, has a statistically significant difference in household size between formal and informal borrowers. However, with a difference of 0.2 people per household on average between each group, while statistically significant, it may not be economically significant. While household size is slightly larger on average for formal borrowers, the difference is minimal and likely doesn't have a major effect on borrowing decisions.

MPCE has another sufficiently small p-value of $p < 0.01$, meaning that there is a statistically significant difference in Monthly Per Capita Consumer Expenditure (MPCE) between formal and informal borrowers. Here we see the largest difference in the groups, with the difference between the two groups' means at around 4000 rupees. This tells us that formal borrowers tend to spend a significant 33% more than informal borrowers. We can conclude that households accessing formal credit have higher consumption levels, suggesting higher income or wealth.

While the statistical significance is clear, the economic magnitude differs across variables. The difference in land ownership, although statistically significant, is small in absolute terms (0.008 acres). This suggests that land ownership alone may not explain formal borrowing access, but could operate through collateral constraints. In contrast, the 33% higher MPCE among formal borrowers represents a substantively large gap, indicating that income-related screening mechanisms may play a central role in formal credit allocation. These unconditional differences motivate the multivariate regression analysis that follows, where confounding factors are jointly controlled.

Loan Distributions for Urban and Rural:

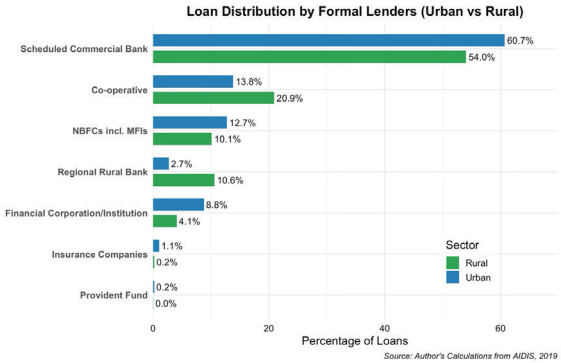


Figure 3: Distribution of loans from formal sources, comparing rural and urban borrowers. Scheduled Commercial Banks provide over half of all loans. Co-operative lending is more present in rural areas, which is reflected in borrower behavior.

Figure 3 shows us how rural and urban communities borrow when they choose to borrow from a formal source. Looking at the population as a whole, Scheduled Commercial Banks are clearly the most popular option, providing over 50% of the loans for both rural and urban populations. Co-operative lending is more popular among rural borrowers than urban borrowers, due to its stronger rural presence. Microfinance institutions are one of the more popular sources among both populations, with over 10% of formal loans going through Non-Bank Financial Corporations (NBFCs).

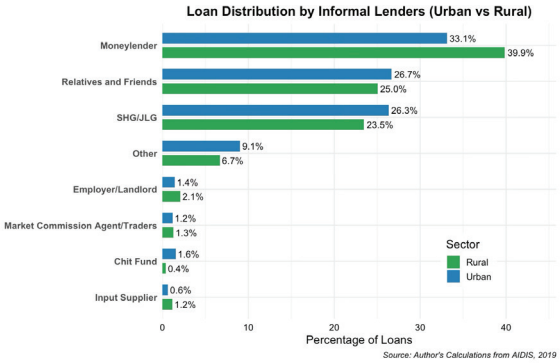


Figure 4: Distribution of loans from informal sources, comparing rural and urban borrowers. The distribution between informal lending sources is consistent between rural and urban borrowers, and moneylenders are the most popular choice.

Figure 4 shows us the differences between the sources of rural and urban borrowers when choosing an informal loan source. We see that the differences between rural and urban borrowing for each of the sources are much smaller than for formal borrowing, with all sources having less than a 3% difference between rural and urban borrowers, except for moneylenders, the most common choice. Rural borrowers are more likely to choose a moneylender by about 7 percentage points, attributed to moneylenders, who usually operate in villages or rural areas, accounting for the larger difference between the two populations for moneylenders.

Loan Purpose for Urban and Rural:

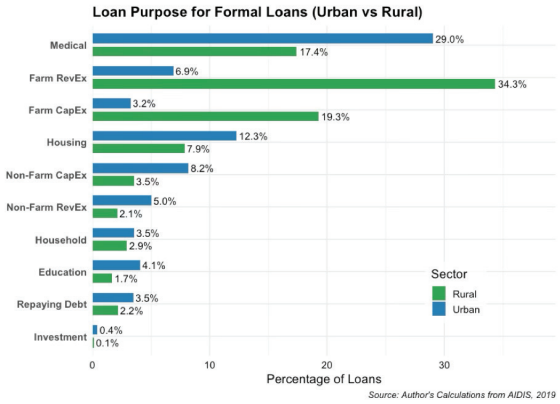


Figure 5: Distribution of loan purposes from formal sources, comparing rural and urban borrowers. The most common urban formal loan purpose is medical, while the most common rural formal loan purpose is farm revenue expenditure.

DOI: 10.36838/IJHSR812.81

Medical reasons are the most cited reason for borrowing, according to Figure 5, mainly from urban sources. One possible explanation for this is that when there is an urgent medical emergency and a loan has to be taken, urban communities have an easier time getting a formal loan than rural communities. Farm-related expenditure forms the next two largest purposes named for formal borrowing, showing the agricultural reliance in modern Indian society. As would be expected, the vast majority of borrowers for farm expenditure are rural, with over half of all rural formal loans being for farm revenue expenditure (RevEx) or capital expenditure (CapEx).

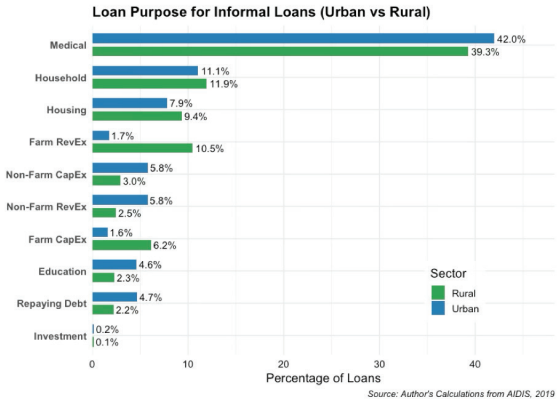


Figure 6: Distribution of loan purposes from informal sources, comparing rural and urban borrowers. For informal loans, both rural and urban borrowers cite medical purposes as the dominant purpose.

Just as with formal loans, Figure 6 tells us that medical reasons are the most common reason for borrowing from informal sources, with much more even numbers than with formal loans and a much higher proportion. This can be explained as borrowers needing medical loans urgently, in the event of a sudden emergency that requires money as soon as possible. Since informal loans are easier and generally quicker to obtain, more borrowers go to informal sources. Farm-related expenditure is much lower here, as farm loans are usually much larger than normal informal loans, so going to a formal source offers more competitive interest rates and the potential for much larger loans.

Loan Purpose for Formal and Informal:

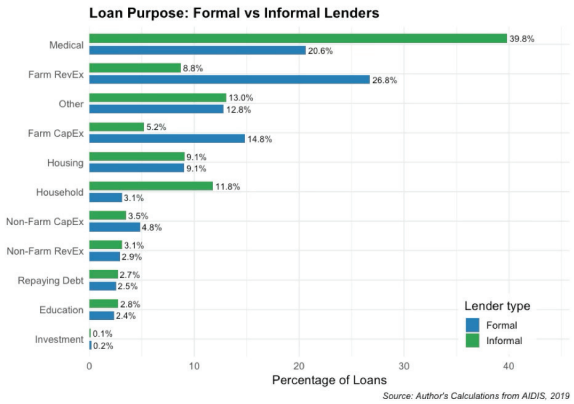


Figure 7: Distribution of loan purposes for formal and informal lenders. The largest share of informal loans is medical, and the largest share of formal loans is farm revenue expenditure. Loan purposes that are typically more urgent, like medical, are more likely to be informal.

Figure 7 displays the distribution of each loan purpose for formal and informal sources. Most medical loans are informal, due to the relative urgency of the loan and the increased accessibility of informal lending, particularly for rural communities. Farm-related expenditure is dominated by formal loans, due to the higher interest rates for informal loans and the higher formal loan size. We notice that urgent loan purposes, like medical and household, have a higher proportion of informal borrowing, detailing the behavior of borrowers when faced with an urgent situation, leaning away from formal loan sources.

The dominance of informal credit for medical purposes suggests that time-to-disbursement and procedural requirements act as binding constraints in formal markets. Formal lenders typically require documentation, collateral assessment, and processing delays, which are incompatible with emergency expenditure. This complements the t-test and regression findings, where income and land ownership increase formal access, suggesting that formal borrowing is structured around verifiable and planned financial needs rather than liquidity shocks.

Loan Distribution by Sector:

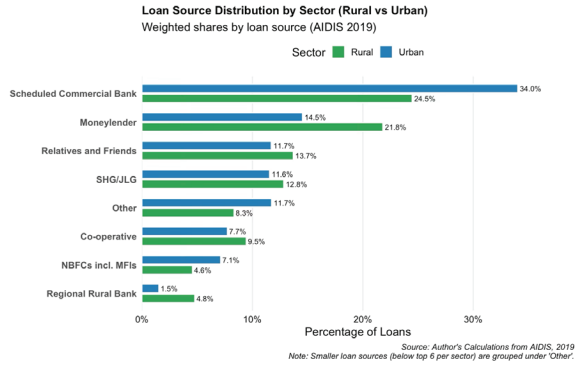


Figure 8: Distribution of loan sources by sector. Both rural and urban borrowers use SCBs the most. As would be expected, moneylenders are more common in rural communities, with other values relatively consistent between the two groups.

To compare where borrowers in rural and urban communities get their loans from, Figure 8 illustrates the top 6 sources, with the rest falling under the category of 'Other'. Scheduled Commercial Banks are the most common, once again, with both sectors having SCBs as the preferred source. Moneylenders, which are often considered much more common in rural sectors, are also the second major choice for urban borrowers, although there is a higher proportion of rural loans than urban loans from moneylenders. Relatives and Friends, as well as SHGs and JLGs, persist with similar rates across both sectors, as borrowing from people you already know makes sense for many borrowers who seek smaller-scale credit, and is much more convenient.

Regression Results:

Table 5: Logit and OLS regression table, with AME for interpretability. MPCE increases the likelihood of borrowing formally. OBCs and 'Others' are the most likely castes for formal borrowing, while Muslim and Christian households are less likely to borrow formally as compared to Hindu households.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Model description	Logit: MPCE only	Logit: Full controls	Logit + State FE	OLS + District FE	Logit + District FE
Monthly Per Capita Expenditure (MPCE per 1000 units)	1.36	1.25	1.08	0.72	1.13
Household Size	–	-0.48	-0.46	-0.06	-0.59
Land Owned (acres)	–	2.62	2.55	2.58	2.52
Urban (1 = Urban, 0 = Rural)	–	3.36	2.58	4.08	2.26
Caste: SC (ref = OBC)	–	-4.71	-5.72	-6.80	-6.24
Caste: ST (ref = OBC)	–	2.09	-2.32	-4.20	-3.69
Caste: Others (ref = OBC)	–	10.78	7.53	6.68	6.16
Religion: Christianity (ref = Hinduism)	–	-7.92	-3.90	-2.57	-3.23
Religion: Islam (ref = Hinduism)	–	-8.53	-9.49	-11.42	-10.67
Religion: Others (ref = Hinduism)	–	-0.16	1.98	3.10	2.67

AMEs expressed as percentage-point changes. Computed on a 10% subsample for efficiency. Models with fixed effects (3 & 5) computed without standard errors due to memory and package limitations. Source: Author's calculations from AIDIS, 2019.

Table 5 displays the results of the 5 regression models used in this study, with descriptions of each model in the top row. The inclusion of caste and religion indicators allows for comparison between different groups and shows how differences in characteristics impact accessibility of formal borrowing. Instead of log-odds, Average Marginal Effects have been chosen, as they provide changes in predicted probability, which are easier to interpret than log-odds coefficients. It shows how much a one-unit change in the explanatory variable will affect the outcome variable, in percentage points. It is worth noting that Model 4 uses OLS regression instead of logit regression, but the coefficients offer similar interpretations to the AME values.

MPCE, as would be expected, is positively associated with an increase in formal borrowing, by about 1% per thousand rupees, as borrowers with higher income can take larger loans out from banks. Household size has a small negative effect on formal borrowing, possibly explained by larger households implying lower income, needing to fit more people under a single roof. Increased land ownership is associated with an increase in the chance of formal borrowing by about 2.5 percentage points per acre, and the large proportion of farm-related formal loans offers qualification for this trend. Urban residence also increases the likelihood of formal borrowing, reflecting better access to formal financial institutions.

Caste is compared to the reference category of OBC, and relative to this, AME coefficients show a reduction in the likelihood of formal borrowing for SC households of between -4.71 and -6.80 percentage points, showing how inequality in socioeconomic groups persists through the credit market, where fewer SC households can borrow formally. We see a mixed range for ST households, from a positive coefficient in the baseline model to negative coefficients in the fixed effects models, highlighting the importance of adjusting for micro-variations that are unseen in the data. The large increase in the probability of formal borrowing for the 'Others' caste category, between 6.16 and 10.78 percentage points, reflects how socially advantaged groups have much more access to formal borrowing.

We see similar results with religion, where Muslims are consistently less likely to borrow formally relative to Hindus, showing persistent inequality in access to credit. Christians also show a lower likelihood of formal borrowing, while the 'Others' religion group displays smaller and mixed effects. This further demonstrates that disparities in formal borrowing are not only driven by economic characteristics but also by social and religious group identity.

The OLS coefficients indicate how a one-unit change in each explanatory variable affects the predicted value of the outcome, holding all other variables constant. Negative coefficients for SC, ST, as well as Christian and Muslim households, imply a lower predicted likelihood of formal borrowing relative to their respective reference groups (OBC for caste and Hindu for religion). Positive coefficients for land ownership, urban residence, and the 'Others' caste and religion categories indicate higher predicted likelihoods of formal borrowing.

These disparities persist even after controlling for income, land ownership, and fixed regional effects, suggesting that differences are not fully explained by observable economic characteristics. This raises the possibility of structural exclusion mechanisms operating within credit screening processes or through differential access to documentation and financial networks.

■ Conclusion

This study provides updated, nationally representative evidence on the determinants of formal borrowing in India, using the 2019 All India Debt and Investment Survey (AIDIS). The results show that, while formal and informal borrowing coexist at comparable levels, access to formal credit is highly uneven across households, often associated with particular characteristics of the borrowers. Higher consumption, land ownership, and urban location significantly increase the likelihood of formal borrowing, reflecting the importance of income and proximity to financial institutions when it comes to credit accessibility. In contrast, socially disadvantaged groups - particularly Scheduled Caste households, Muslims, and Christians - remain systematically less likely to access formal credit, even after controlling for state and district-level variations through fixed effects. These findings indicate that inequalities in India's credit market closely mirror broader social and structural inequalities, rather than being driven solely by economic characteristics or rural-urban differences.

The study also highlights the role of loan purpose and relative urgency in shaping borrowing behavior. Formal credit dominates larger, planned expenditure, such as agricultural investment, while informal credit is the primary source for urgent needs, especially medical and household expenses. This suggests that barriers to formal borrowing are institutional as well as situation-dependent, with speed and accessibility playing an important role in household decision-making when it comes to borrowing.

From a policy perspective, the findings imply that expanding financial inclusion requires more than a narrow focus on rural outreach. Existing policies mandate that a fixed share of lending must be supplied to priority sectors (like the 40% sector requirement for rural credit through the Priority Sector Lending scheme), but these regulations are insufficient on their own. Disparities by caste and religion hint towards the need for more targeted interventions within both rural and urban credit markets. These interventions could include improved access to

DOI: 10.36838/IJHSR812.81

collateral through better land record systems, simplified loan procedures for urgent credit needs, and continued and explicit monitoring of credit allocation across social groups to mitigate exclusion.

Overall, the evidence suggests that formal financial expansion in India has been uneven in terms of its inclusivity. Addressing structural inequality in the credit market will require policies that explicitly recognize and counteract social disadvantage, rather than assuming that broader financial development will automatically translate into equal access.

Acknowledgments

I would like to extend my sincere appreciation to Atif Anwar for his guidance and mentorship throughout the study.

References

1. Banerjee, A. V.; Duflo, E. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*; PublicAffairs: New York, 2011; Chapter 7.
2. Banerjee, A. V.; Duflo, E. Giving Credit Where It Is Due. *J. Econ. Perspect*. 2010, *24* (3), 61–80. https://doi.org/10.1257/jep.24.3.61
3. Bateman, M.; Chang, H.-J. Microfinance and the Illusion of Development. *World Econ. Rev*. 2012, 1, 13–36.
4. Prakash, A.; Kumar, A.; Singh, D. Determinants of Household Borrowing in India: Evidence from National Microdata. *Indian J. Econ. Dev*. 2019, *15* (3), 512–522.
5. Field, E.; Pande, R.; Papp, J.; Rigol, N. Does the Classic Microfinance Model Discourage Entrepreneurship among the Poor? *Am. Econ. Rev*. 2013, *103* (6), 2196–2226. https://doi.org/10.1257/aer.103.6.2196
6. Narayanan, S.; Chakraborty, J. *Land as collateral in India*; IGIDR Working Paper 2019-006, Indira Gandhi Institute of Development Research: Mumbai, 2019.
7. Venumuddala, V. R. Patterns in Demand Side Financial Inclusion in India — An Inquiry Using IHDS Panel Data. arXiv. 2020, arXiv:2005.08961. https://arxiv.org/abs/2005.08961
8. Binswanger, H. P.; Khandker, S. R. The Impact of Formal Finance on the Rural Economy of India. *J. Dev. Stud*. 1995, *32* (2), 234–262. https://doi.org/10.1080/00220389508422413
9. Burgess, R.; Pande, R. Do Rural Banks Matter? Evidence from the Indian Social Banking Experiment. *Am. Econ. Rev*. 2005, *95* (3), 780–795. https://doi.org/10.1257/0002828054201242 aeaweb.org+1
10. Buliskeria, N.; Baxa, J. Do Rural Banks Matter That Much? Burgess and Pande (2005) Reconsidered. *J. Appl. Econ*. 2022, DOI: 10.1002/jae.2922 library.utia.cas.cz+1
11. Ghosh, P. Formal Credit, Corruption and Informal Credit. *J. Dev. Econ*. 2013, *104*, 47–58. https://doi.org/10.1016/j.jdeveco.2013.05.001
12. Tagade, N.; Reza, M. S. *Financial Inclusion in India*; Economic and Political Weekly 2025, *60* (22).
13. Chavan, P. *How Inclusive Is Financial Inclusion in India?*; Economic and Political Weekly 2025, *60* (40).
14. Karthick, V., and Madheswaran, S. (2018). Access to formal credit in the Indian agriculture: Does caste matter? *Journal of Social Inclusion Studies*, 4(2). https://doi.org/10.1177/2394481118814064
15. National Sample Survey Office. *All India Debt and Investment Survey (AIDIS), NSS 77th Round*; Government of India: New Delhi, 2019.
16. National Data Archive (NADA). *AIDIS 2019 Microdata Catalogue Entry*. https://microdata.gov.in/NADA/index.php/catalog/156.
17. Desai, S.; Vanneman, R.; NCAER India Human Development Survey (IHDS) Team. *India Human Development Survey-II (IHDS-II), 2011–12*; University of Maryland & NCAER: New Delhi, 2015.
18. NABARD. *NABARD All India Rural Financial Inclusion Survey 2016–17 (NAFIS)*; National Bank for Agriculture and Rural Development: Mumbai, 2018.
19. Wooldridge, J. M. *Introductory Econometrics: A Modern Approach*, 5th ed.; South-Western Cengage Learning: Mason, OH, 2013.
20. Allison, P. D. *Fixed Effects Regression Models*; SAGE Publications: Thousand Oaks, CA, 2009.
21. UVA Library. *A Beginner's Guide to Marginal Effects*. UVA Library StatLab. https://library.virginia.edu/data/articles/a-beginners-guide-to-marginal-effects

Author

Aaran Nihalani is a Year 12 Student at Eton College with interests in Computer Science, Machine Learning, Maths, Statistics, and Economics. He finds applied Machine Learning particularly interesting in the context of Finance and Econometrics, and plans to study Computer Science with a focus on Finance and Machine Learning at university.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

The Rise of Digital Trading: Accessibility Versus Risk of Financial Ruin

Adidev Panday

Woodstock School, Landour Road, Mussoorie, Uttarakhand, 248179, India; panday.adidev@gmail.com

ABSTRACT: This paper looks into the double-edged impact of digital trading and investing. Currently, apps and websites like Robinhood, Groww, and Zerodha have taken over the market by removing things such as fees and account minimums. In addition, they also make access to the market merely a few clicks away from any user, whether properly educated or not. The research in the paper looks into the causes and effects of the rapid shift into digitized trading. Additionally, terms such as "finfluencer" and "gamify" are reviewed. Reports and studies in the paper inform us about the large negative effects of ill-advised trading and investing. The paper argues that though digitization of the markets is helpful and convenient for numerous investors, it carries with it a darker side. A side in which users with little to no financial knowledge end up losing more than they bargained for; a side that carries numerous hidden dangers.

KEYWORDS: Economics, Finance, Digital Trading Platforms, Retail Investors, Financial Risk.

Introduction

Warren Buffett is the first name that comes up in a household when talking about investing; however, most people do not understand what investing, trading, and the stock market really are. Electronic trading and new investment sites have transformed stock market involvement at an unprecedented rate. Websites and apps like RobinHood, Groww, and Zerodha now offer easy-to-use interfaces, real-time statistics, and low barriers to entry – capabilities that were previously present only in the territory of institutional traders – thereby allowing trading to reach new heights. These websites and apps use automated algorithms and algorithm-driven analysis, which make transactions easier and hassle-free for the modern and wider population.[1] Consequently, large numbers of new retail investors are inclined to come into the market with unprecedented ease. While democratizing investing by enabling anyone with a smartphone to invest, it could also instill impulse-based trades and ill-informed choices.

The simplicity of online market entry on these sites usually hides the high learning curve that is generally required to successfully trade and invest. For example, current FINRA research finds that most investors younger than 35 use social media sites like YouTube, Reddit, and TikTok for investment advice instead of traditional financial institutions or certified professionals.[1] This trend is particularly disturbing because information on social media is not always accurate. There are "finfluencers" or financial influencers, who have brought a whole new generation of market-driven, money-hungry, and emotionally charged people into the scene of investing. According to a report by Barron's, multiple companies have been or are under investigation for the use of finfluencers to manipulate users by conducting activities against advertising and disclosure laws.[2] All of this makes it abundantly clear to us that financial literacy is an important issue that needs to be addressed. There are millions of people, both young and old, who are falling victim to scams targeted towards the ill-advised; meanwhile, large hedge funds and investment banks rake in billions of dollars every year. Along with this, many companies are attempting to gamify their apps and websites to attract modern audiences. This is not just a marketing gimmick, but could also be a dangerous feature targeted towards financially illiterate users.

The main purpose of this paper is to look into the two-faced effect of online trading platforms: on one side, they expand access and participation in financial markets, while on the other side, they can mislead and put those in danger who do not possess the expertise to invest wisely. By delving into these aspects, the aim is to find out whether the advantages of online trading outweigh the potentially dangerous effects it could introduce into modern-day investing.

The history of the stock market goes back centuries before the introduction of electronic trading. The first official stock exchange happened in 1602 in Amsterdam. This occurred in order for merchants to trade shares in the Dutch East India Company, a very successful business at that time. Since then, stock trading has only boomed further and has now become the core of finance.[3] Stock exchanges such as the New York Stock Exchange (NYSE), the London Stock Exchange (LSE), and the Bombay Stock Exchange (BSE) have popped up around the globe. However, until the end of the 20th century, stocks were traded physically in these exchanges.[4] The process was much harder, wherein brokers had to cry, shout, and scream just to be heard, only sometimes. Due to this, trades were more expensive, time-consuming, and inefficient. During this time period, the stock market was run by wealthy individuals with elite knowledge about the market. All of this began to change at the turn of the century; the internet was booming, and so was online trading. The first stock exchange to shift to an electronic means of trading was NASDAQ in 1971; in the coming decades, other popular exchanges digitized as well.[5]

DOI: 10.36838/IJHSR812.89

Due to this, today's stock markets exist in very different contexts. Trading is no longer limited to Wall Street professionals or wealthy, highly intelligent people. Trades can be made faster than ever on a simple smartphone on apps and sites such as the ones mentioned before. These sites allow the user to access and manage their portfolio from the comfort of their home. In return, they rarely ask for anything and are generally free to download and use. Nevertheless, paid portals such as the Bloomberg portal still exist for those who are much more serious about trading and investing.[6] Meanwhile, the social media explosion has added a new and risky factor to investing. Sites such as Reddit, YouTube, and TikTok have become casual spaces of financial conversation and advice, with groups like WallStreetBets on Reddit gaining worldwide recognition for their ability to shape stock prices through group action. A prime example of this is the recent 2021 GameStop short squeeze; an incident so incredible that it barely falls under miraculous.[7] In summary, it brought together millions of young people to buy stocks of a failing business while running Wall Street firms out of business. Through incidents such as this one, the community of investors grows larger and welcomes new investors every day. Nevertheless, this is seen as extremely risky by some, as the advice is often based on hunches or speculation.

As shown in Figure 1, the history of the stock market has been long and vast. For hundreds of years, different means of trade have been used throughout the globe. Figure 1 is key in understanding how, through 1602 - 2021, trading has gone from the simple exchange of goods via in-person merchant trade to online investing, which is greatly impacted by the influence of social media figures. From a time in the 1970s when stock manipulation was common to the highly monitored markets of the modern day.

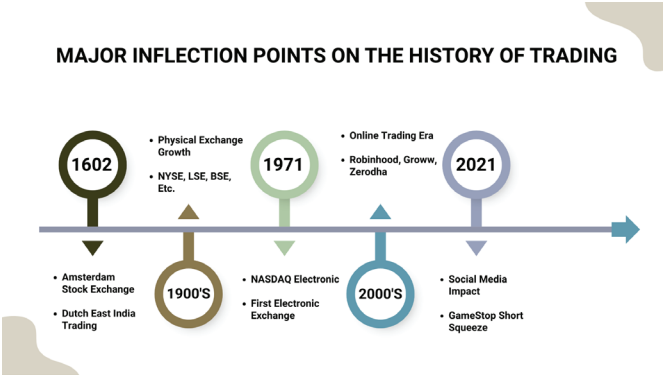


Figure 1: Major inflection points in trading history, across 5 key periods between 1602 and 2021.

During this period, the history has been relatively well kept with records dating back centuries. Since then, an increasing number of studies have explored how digitization has changed investor participation and results. Nathani and Dharmani's 2024 study finds that fintech platforms popularize market access by lowering fees, offering novel interfaces, and delivering real-time data, which importantly increases retail investor participation. However, their mixed methods analysis reveals that trading apps also enhance speculation-based trading and information overload for most users.[8] These results are similar to the findings of a 2022 research study of Chinese families, which revealed that digital financial literacy is important for effective and knowledgeable participation; less financially literate users are more likely to abuse and use digital trading sites, raising their risk exposure.[9] Another research conducted in 2025, among Italian robo-advisor users, confirms this. They state that more financially and digitally literate investors use platforms responsibly, and others tend to excessively depend on automated advice at the expense of personal judgment.[10] Once again, in a 2023 OECD global report, it was mentioned that while usage of fintech has increased a lot, investor readiness has not caught up, particularly among users under 30 who tend to be overconfident in their financial decision-making capabilities.[11] Overall, the literature presents a balanced picture: digital access enhances market inclusion but does not automatically ensure safe or well-informed decision-making.

Furthermore, historical evidence reinforces the balanced nature of digital access to markets. Before digitization, stock market participation by retailers was disrupted by high fees and commissions. However, due to digitization, zero-commission trading was initiated by Robinhood in 2013 and was soon picked up by existing players.[12] In accordance with this time period, retail trading activity picked up exponentially: Robinhood now has close to 26 million subscribers and $255 billion in assets under management in mid-2025.[13] In India, the growth of demat accounts has gone from 40 million in 2020 to more than 150 million in 2024. This is representative of the global phenomenon, but SEBI indicates that 90% of retail traders in derivatives are losing money, with 93% of F&O trading newcomers posting net losses between 2021–24 (view Figure 2).[14]

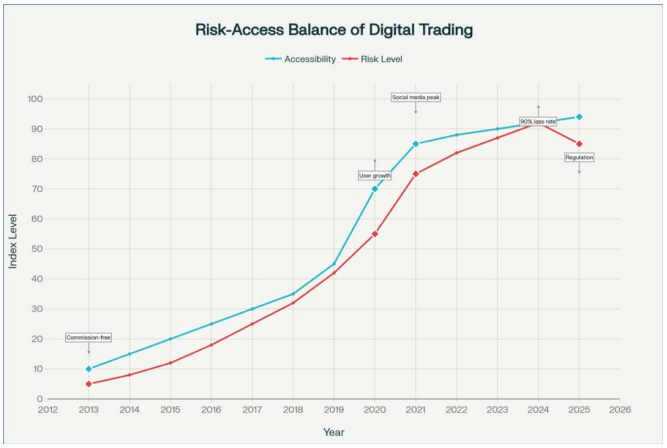


Figure 2: Risk–access balance of digital trading, showing that increased accessibility is associated with higher risk levels.

Looking at Figure 2 offers a key insight into how much progress has been made over 2013 - 2025. Looking at 2013 – 2019, the rise of accessibility vs risk remained relatively similar, however the year COVID-19 hit, everything began booming. During this period, risk levels were extremely high, and so was the accessibility. Thankfully, as seen in the graph, risk levels have decreased over the past year.

Other markets like the UK and South Korea follow the same pattern, in which post-pandemic trading app downloads boomed but failed to become long-term investment winners.

The research also examines the role of social platforms on trading behavior: platforms such as StockTwits, eToro, and Reddit tend to result in increased risk-taking through "mirror trading," a strategy in which an inexperienced user copies trades of successful users.[15] Lastly, a 2023 survey conducted by the CFA Institute also substantiates that even though social media makes individuals feel more informed about financial trends, it also strongly enhances emotional investing and short-term success.[16] These studies, surrounded by historical data, reinforce the overarching thesis that although digitization and social media have undoubtedly enhanced access, they simultaneously spread susceptibilities associated with a lack of financial literacy and emotional patterns of trading.

Methods

This research uses a qualitative research methodology involving largely secondary data analysis. The materials used include scholarly research papers, industry reports, government and regulatory documents, case studies, market data, and news reports on digital trading platforms, fintech innovation, and investor attitudes. Some of the most important sources are research studies by the OECD, FINRA, CFA Institute, Amundi Research Center, and financial news reporters like Barron's, Bloomberg, and MarketWatch. The approach entailed a systematic literature review to collect pertinent information concerning the impacts of mobile trading applications such as Robinhood, Groww, and Zerodha on retail investors. The review entailed examination of trends and evidence concerning investor access, behavioral biases, gamified interfaces, financial literacy, social media influence, and corporate practices. Besides, real-life case studies like the 2021 GameStop short squeeze and the Sky Capital scam were studied to emphasize how digital investing habits can cause both market disruptions as well as personal financial losses. Synthesis of data was conducted by grouping findings into dominant themes: accessibility vs. risk, behavioral effects, financial competence, and regulatory issues. This facilitated a holistic perspective of how online trading platforms influence market participation and decision-making without conducting original data collection or experimentation. This technique aims to provide an enlightened analysis given current knowledge and actual events, as well as gaps that must be addressed through further research.

Results and Discussion

The historical record and review of the literature highlight a two-faced story: electronic trading platforms reduce barriers to access and make access to markets available to all, but they also expose users to greater financial risk. On the one hand, Fintech tools like robo-advisors and zero-commission trades have widened retail access.[11] On the other hand, many young investors lack financial literacy, leading to impulsive and risky trading.

Digital platforms have eliminated things like brokerage commissions and minimum account balances; this enables investors to trade equities, derivatives, futures, options, and stocks without any hassle.[17] However, a lack of foundational financial literacy can lead some users to trade speculatively, without a full understanding of risk management, or even a little understanding of risk management, diversification, or the basics of the stock market. Although general awareness of financial issues is pretty high, young adults rarely compare investment products or consult with advisors; this pattern is especially troubling as more people invest. The outcome is not just individual ignorance but also the spreading of false information. Misinformed investors then contribute to market volatility in events such as meme stocks, echo chambers, and sentiment-driven results.[18] These events, fueled by group behavior enhanced through social media, can cause unstable market movements and inflict large amounts of financial damage on the users.

A key observation is that behavioral biases are heightened in fintech platforms that gamify trading. Impulsive trades and momentum chasing are created by website and app UI's (User Interfaces). For example, product recommendation features on investment sites really increase low/low-medium income investor participation, but such investors obtain bad returns, do little to no research, and hold stocks for too long or too short, depending on the bad advice. More commonly than ever, sites tend to implement "confirmation bias" designs by implementing and pushing forward user-chosen alerts, trending stocks, and cherry-picked social feeds, prompting further trades without gaining further understanding.[19] Such a design can be compared to the mannerisms of social network algorithms in which they push emotionally appealing material, subsequently driving "mirror trading" behavior as new users mirror peer activity during market frenzies.

One pattern that can be found throughout the course of history, literature, and evidence is that the role and position of financial literacy ends up being one of the most crucial factors in deciding whether access to digital investing mediums increases positive or negative results. Research on fintech development across more than 100 nations indicates that there is a strong correlation that exists, that being: increased literacy ensures sustainable digital finance use, while decreased literacy is generally and more often than not associated with suboptimal decision making and higher financial vulnerability. Studies have further established that when users do not possess basic financial education, fintech products usually tend to escalate the exact risk they attempt to minimize. Digitally well-educated investors, however, can more effectively analyze products, practice trades, and be careful about available information. These findings reiterate that financial literacy is not an afterthought; it is essential to trade and invest properly.[20]

Digital accessibility has allowed for actual innovation, but it has also invited destructive abuse. For instance, the concept of micro trading and derivative speculation among retail users, much driven by online brokers and finfluencers, has led to large amounts of losses. In some areas, the vast majority of retail traders have lost a lot of money, particularly newer users who joined markets during volatile post-pandemic circumstances.[14] A classic historical case is the Sky Capital scam, which involved deceptive electronic advertising and the distribution of unregistered securities and stocks. While occurring before the widespread use of trading apps, the case illustrates how fraud works in online spaces when regulations are weak

DOI: 10.36838/IJHSR812.89

and improper action is taken to prevent it.[21] A more recent example is the GameStop short of early 2021, coordinated by users on a Reddit discussion board. This incident, fueled by the ordinary crowd instead of conventional financial brokers, generated extreme amounts of market turbulence. Some users made money, but numerous others got in late and suffered losses when the bubble burst. It featured both the potential and danger of digital investing.[22]

Big financial and technology corporations have taken advantage of the increasing pool of new and inexperienced investors by employing convincing marketing strategies, game-like interfaces, and behavioral data gathering. These companies frequently bundle "recommended" products into platforms and subtly nudge users toward particular funds or stocks that better suit the companies' best interests or fee arrangements than investor requirements. Instead of replacing traditional finance, most fintech companies now help big financial institutions sell their products. They maintain product ownership, expand reach, and gather large amounts of behavioral insights. As investment products like ESG baskets or clean energy portfolios become more popular, most of which are not properly tested by companies or heavily sponsored finfluencers, retail investors get attracted by compelling stories but remain clueless about the underlying risks or expenses. While this goes on, international payment giants keep on adding fintech layers on their platforms, earning more from the systems even before registered users officially start investing. This business model, coupled with low patience among new users, can prioritize corporate profitability at the expense of optimal user outcomes.

When looking at all the research, there is one question that stands out: can digital platforms have the benefits of access while providing safeguards for economically disadvantaged users? Several possible answers can be found in different areas. Financial readiness and education programs should be introduced by schools and financial regulators, starting in high school and continuing through university. These programs should focus on helping students compare financial options, understand risks, and think carefully about financial decisions. Ethical app designs should also be nudged into fintech businesses; these designs include: risk analysis, loss simulation, and delayed execution of trades for large-value transactions. Platforms should be able to mitigate dangerous conduct by making sure users are competent in terms of financial knowledge before they can access advanced trading capabilities.

Transparency and accountability should also be implemented by regulators, wherein they should require companies to clearly explain how their recommendations work, reveal any conflicts of interest, and show a clear breakdown of all investment costs. Companies can also do with standard Know Your Client (KYC), which is a small background check, procedures for ensuring responsibility in the industry. Enhanced supervision of social media influence should also be taken seriously. Since the influence of online communities on investor choice is strong, regulators need to keep coordinated behavior under control and surveillance, ban manipulative campaigns, and make fintech platforms give warnings whenever social media trends are advertised.

There are some strong trends, yet there remain some limitations. The majority of research works examine developed markets only, which underrepresent emerging economies with high trading rates and volume. It is also hard to make a case against the bad effects of gaining financial literacy from social media trends. In addition, fintech change is fast and continuous. New products like BPNL, AI-driven portfolios, and crypto combinations move very fast in pace. This presents an ongoing challenge: keeping up to date in an environment where the environment shifts more quickly than quicksand. Future research should focus on long-term data, testing behavior-based solutions, and comparing results across different populations to see how education and design affect investor success.

The creation of digital trading platforms has revolutionized the way ordinary people access financial markets. Simple mobile apps have made it easier than ever to invest in markets that were formerly the domain of high-net-worth professionals. But this accessibility has been accompanied by very serious and generally overlooked threats. Numerous new investors, especially young ones, join the market with little capital and knowledge. In addition, nowadays markets are known to sway a lot, frequently depending on what is happening in the news or on social media sites. This setting encourages impulsive action, speculative trading, and emotional choice, all of which can cause substantial financial losses. Instances such as the GameStop short squeeze and mass losses in retail derivatives trading are a case in point where technology has gotten ahead of education. In addition, businesses operating these platforms have, on occasion, taken advantage of the innocence of the users. Gamified user interfaces, manipulative "recommended" investment instruments, and the increasing power of finfluencers make a market ecosystem that focuses more on interaction rather than learning. This imbalance exposes users, particularly those who lack a good background in financial literacy.

Conclusion

To conclude, the results of this study were not quantitative but rather suggestive. Through research and deduction, I found that companies and corporations take advantage of ill-advised investors using gamified investing platforms and modern technology. Nevertheless, the technology of investing is not necessarily bad when used wisely. When it is combined with education, regulation, and ethical design, it has the potential to be one of the largest accessible economic tools on Earth. It eliminates the old norms, such as location, money, and complexity, and allows everyone with a smartphone to enter the world of building wealth and contributing to the world economy. The goal is not to exclude or cancel digital finance but to evolve with it. With the right safeguards, the digital investing landscape can become not only a great tool but also a way for people to create wealth. It can do so, as long as people can make informed decisions, build sustainable portfolios, and contribute meaningfully to a healthier financial future for themselves.

Acknowledgments

I would like to acknowledge and graciously thank Ms.Punya Panday, my mother, and Ms. Shiny Lohani, a mentor, for their support during the development process of my paper.

References

1. FINRA. "Following the Crowd: Investing and Social Media." FINRA, https://www.finra.org/investors/insights/following-crowd-investing-and-social-media?. Accessed 19 June 2025.
2. "Advisor Fined for Using Finfluencers to Promote Trading App." *Barron's*, https://www.barrons.com/advisor/articles/finfluencers-finra-fines-tradezero-america-online-brokerage-af83d063. Accessed 19 June 2025.
3. "Investopedia: Who Issued the First Stock?" *Investopedia*, https://www.investopedia.com/ask/answers/08/first-company-issue-stock-dutch-east-india.asp. Accessed 19 June 2025.
4. "A Brief History of the Amsterdam Stock Exchange." *Beursgeschiedenis*, https://www.beursgeschiedenis.nl/en/the-story/. Accessed 19 June 2025.
5. IEEE Spectrum. "NASDAQ's Technology Floor: Its President Takes Stock." *IEEE Spectrum*, https://spectrum.ieee.org/nasdaqs-technology-floor-its-president-takes-stock. Accessed 19 June 2025.
6. "Bloomberg Terminal Overview." *Bloomberg Professional Services*, https://www.bloomberg.com/professional/terminal-overview/. Accessed 19 June 2025.
7. "Trading Behavior of Retail Investors in Online Communities." *ACM Digital Library*, 2024, https://dl.acm.org/doi/10.1145/3660760. Accessed 19 June 2025.
8. Nathani, Riya, and Meher Dharmani. "A Study of Impact of Financial Technology (Fintech) on Retail Investors in Trading Securities." *European Economic and Legal Education Times*, vol. 6, no. 1, 2024, pp. 22–33. https://www.eelet.org.uk/index.php/journal/article/view/1836. Accessed 19 June 2025.
9. Yue, Pengpeng, Aslihan Gizem Korkmaz, Zhichao Yin, and Haigang Zhou. "Digital Financial Literacy and Household Investment Behavior." arXiv, 2022, https://arxiv.org/abs/2201.09221. Accessed 19 June 2025.
10. Aristei, David, and Manuela Gallo. "Financial Literacy, Robo-Advising, and the Demand for Human Financial Advice: Evidence from Italy." *SSRN*, 2025, https://ssrn.com/abstract=5290783. Accessed 19 June 2025.
11. OECD. "Financial Markets." Organisation for Economic Co-operation and Development, https://www.oecd.org/en/topics/financial-markets.html. Accessed 19 June 2025.
12. "Robinhood's Vlad Tenev Has Led a Trading Revolution to Become a New Wall Street Power." *MarketWatch*, https://www.marketwatch.com/story/robinhoods-vlad-tenev-has-led-a-trading-revolution-to-become-a-new-wall-street-power-6a73bf9b. Accessed 19 June 2025.
13. Social Capital Markets. "Robinhood Statistics." *SoCap*, https://socialcapitalmarkets.net/crypto-trading/robinhood-statistics/. Accessed 19 June 2025.
14. "India's Retail Trading Boom: Why Millions of New Investors Are Losing Money." *LiveMint*, https://www.livemint.com/market/stock-market-news/indias-retail-trading-boom-why-millions-of-new-investors-are-losing-money-11688234567890.html. Accessed 19 June 2025.
15. Amundi Research Center. "Retail Investors' Behaviour in the Digital Age: How Digitalisation Is Impacting Investment Decisions." *Amundi*, https://research-center.amundi.com/article/retail-investors-behaviour-digital-age-how-digitalisation-impacting-investment-decisions. Accessed 19 June 2025.
16. CFA Institute. "Finfluencer Report and Social Media Impact on Retail Investors." *CFA Institute*, 2023.
17. "Robinhood Launches Desktop Platform, Adds Futures and Index Options Trading to App." *Investing.com*, https://www.investing.com/news/stock-market-news/robinhood-launches-desktop-platform-adds-futures-and-index-options-trading-to-app-3667464. Accessed 19 June 2025.
18. Zhu, Ruiqi Rich, Cheng He, and Yu Jeffrey Hu. "The Effect of Product Recommendations on Online Investor Behaviors." arXiv, 2023, https://arxiv.org/abs/2303.14263. Accessed 19 June 2025.
19. Zhu, Ruiqi Rich, Cheng He, and Yu Jeffrey Hu. "Algorithm-Driven Product Recommendations and Investor Behavior." arXiv, 2025, https://arxiv.org/pdf/2303.14263. Accessed 19 June 2025.
20. Lusardi, Annamaria, and Olivia S. Mitchell. "The Economic Importance of Financial Literacy: Theory and Evidence." *Journal of Economic Literature*, vol. 52, no. 1, 2014, pp. 5–44.
21. "Sky Capital Scam Case." *CFA Institute Finfluencer Report*, https://rpc.cfainstitute.org/sites/default/files/-/media/documents/article/industry-research/finfluencer-report.pdf. Accessed 19 June 2025.
22. "The GameStop Trading Frenzy Was Fueled by Social Media. It May Be the Start of a New Era." *Time*, https://time.com/5934285/gamestop-trading-wall-street/. Accessed 19 June 2025.

Author

Adi is currently studying in a boarding school in the foothills of the Himalayas. Adi's passion for finance and tech has helped Adi view data science as a potential future college major and encouraged Adi to write this paper.!

DOI: 10.36838/IJHSR812.89

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Qualitative Outcomes of MindBridge: Group CBT Curriculum for Transitional Aged Youth with Developmental Disabilities

Ariya Kaushek, Kinsey Nam

Menlo School, 50 Valparaiso Avenue, Atherton, California, 94027, USA; ariyakaushek@gmail.com, kinsey.nam@gmail.com
Mentor: Dylan Citrin-Cummins

ABSTRACT: Transitional-aged youth (TAY; 16–26) with developmental disabilities (DD) often experience greater mental health comorbidities than other young adults, partly due to gaps in mental health services during the transition from pediatric to adult care. While Group Cognitive Behavioral Therapy (GCBT) is effective for many psychiatric conditions, research on its use for TAY with diverse developmental disabilities is limited. This pilot study examined whether an 8-week brief GCBT intervention (MindBridge) could improve emotional regulation and social awareness among TAY with DD. We recruited 10 participants via opportunity sampling and implemented pre- and post-intervention surveys (18 items) as well as interviews with participants and educators. Sessions were 60 minutes weekly and incorporated CBT-based psychoeducation and art therapy. Only 5 of 10 participants had complete quantitative data; the resulting analyses yielded no statistically significant changes in survey scores. However, qualitative analysis identified two significant themes: (1) Increased Self-Regulation Skills and (2) Enhanced Social Confidence. The study highlights the value of qualitative assessment in research of this type and recommends integrating adapted measurement tools, expanded sample sizes, and longer interventions in future research. These findings contribute to research on adapting evidence-based mental health interventions for youth with DD and suggest that community-based GCBT can offer psychosocial benefits, even if not captured in quantitative data.

KEYWORDS: Social and Behavioral Sciences, Mental Health, Intellectual and Developmental Disabilities (IDD), Transitional-aged Youth (TAY), Group Cognitive Behavioral Therapy (GCBT).

Introduction

In our research, we explored the extent to which *Brief Cognitive Behavioral Therapy* (GCBT) curricula can be effectively implemented to address emotion regulation and social awareness challenges faced by Transitional Aged Youth (TAY) with developmental disabilities and the impacts of such curricula. Our study aims to expand on the limited research regarding the impact of GCBT on the mental health of transition-aged youth with a variety of developmental disabilities—addressing areas such as social awareness and emotion regulation. Unlike existing studies, our research uniquely examines how a brief, 8-week CBT-based program called MindBridge can improve mental health outcomes across TAY with a range of developmental disability types and symptom profiles in individuals aged 18 to 24.

Background on Group Cognitive Behavioral Therapy:

CBT centers on the idea that our thoughts, feelings, and behaviors are all interconnected. By identifying and challenging unhelpful thought patterns, CBT helps individuals change how they act and feel, thereby reducing psychological distress and improving emotional resilience. GCBT further supports the tenets of CBT, focusing on cognitive restructuring through behavioral tasks, normalization through identifying with others, cooperative therapeutic relationships, and positive reinforcement in a safe group environment. GCBT is often delivered in sessions with co-therapists, allowing for more flexible monitoring and content delivery. Although GCBT approaches date back to the late 1980s and 1990s, targeting substance use disorders and comorbid mental health conditions, over the past couple of decades, there has been a large body of evidence supporting the usage of GCBT for young adults with depression, generalized anxiety disorders, obsessive compulsive disorder, social phobias, and PTSD.[1]

For example, one of the earlier studies in 1993 by Heimberg *et al.* examined the long-term effectiveness of GCBT for social phobia, assessing outcomes 4.5 to 6.25 years after treatment.[2] The study involved 19 participants and compared GCBT to a credible alternative, Educational Supportive Group Psychotherapy (ES). Results showed that patients who received GCBT maintained greater improvements in social phobia than those who received ES, as measured by self-report questionnaires, behavioral tests, and structured interviews. Group settings can be more accessible, less stigmatizing, and more approachable compared to an individualized CBT setting.

Early randomized controlled trials demonstrated that GCBT, including models like Relapse Prevention and Guided Self-Change, could produce lasting reductions in substance use and associated psychiatric symptoms, particularly among individuals with mental health comorbidities. A leading 1990s GCBT study, using a 27-session relapse prevention model, found that participants with higher levels of comorbid psychopathology and sociopathy remained abstinent longer and gained greater long-term benefit from GCBT than from interactional group therapy. Additionally, survival analyses revealed better sustained outcomes for the CBT group two years

DOI: 10.36838/IJHSR812.94

post-treatment. These group-based approaches set the stage for modern interventions by emphasizing goal setting and skills training for complex, diverse clinical populations, with the additional benefit of being more cost-effective than treatments on the individual level.[3] However, group CBT presents limitations, including reduced individualization, higher dropout rates, a weaker evidence base compared to individual CBT, and potentially lower perceived acceptability among participants, particularly due to privacy concerns and discomfort with sharing personal issues in a group setting.[4]

TAY with Developmental Disabilities:

Interestingly, today, one of the specific age groups most impacted by high comorbidity rates of mental health disorders and developmental disabilities (DD) is transitional-aged youth (TAY). TAY are defined as individuals aged 16 - 25 navigating the transition from child-centered systems to increased adult services and responsibilities.[5] In a 2008 review published in *Current Opinion in Psychiatry* examining 85 studies on the co-occurrence of mental disorders in adolescents with developmental disabilities, among the four studies with a comparison group, the prevalence of co-occurring mental disorders in youth with Intellectual and Developmental Disabilities (ID/IDD) ranged from 30% to 50%, compared to 8% to 18% in typically developing peers.[6] In fact, 75% of serious mental illnesses (e.g., bipolar disorder or borderline personality disorder) show symptoms by the age of 25, overlapping with TAY developmental stages.[7] Furthermore, developmental disabilities often amplify stress responses to these transition-related demands. For example, Transitional-Aged Youth with DD are faced with demanding systemic and social challenges, often resulting in a concurrent mental illness due to service gaps with pediatric to adult care transitions, enhanced stigma and isolation as a result of their disability, and an overall loss of the structure that school-based support systems provide.[8] However, due to Diagnostic overshadowing—a cognitive bias where healthcare professionals mistakenly attribute symptoms of physical or mental health conditions to a person's pre-existing diagnosis—there exists an extreme amount of missed diagnoses and inadequate treatment for all adolescents and transitional-aged youth with developmental disabilities.[9] For example, a 2015 cross-sectional descriptive study on ADHD found that 38.8% of a cohort of 685 children with neurological disabilities met criteria for ADHD. However, only 28.2% of those were formally diagnosed—likely due at least in part to their syndromic or physical disabilities masking ADHD symptoms.[10] It is also important to note that TAY with developmental disabilities are disproportionately affected by adverse childhood experiences (ACEs) (defined as potentially traumatic events occurring before age 18) with a higher prevalence and adjusted odds of experiencing any ACE (72.1% vs 60.8%; adjusted odds ratio [AOR], 2.0) and four or more ACEs (26.8% vs 14.7%; AOR, 2.6).[11]

Limitations Within Current Treatment Options:

Currently, there are a range of treatment options for Transitional-Aged Youth with DD, including but not limited to occupation-based and advocacy interventions, behavioral and cognitive therapies, care coordination and community-based support, as well as assessment and transition planning.[12] However, there remain immense systematic barriers regarding transition issues with TAY who have developmental disabilities and financial hurdles such as limited insurance coverage and high out-of-pocket costs for many of these treatment options.[13] For example, while Applied Behavioral Analysis (ABA) or Cognitive Behavioral Therapy (CBT) are two commonly recommended forms of treatment, they are less universally accessed due to their high price points; One session of ABA can range anywhere from $120 - $150, one session of individual CBT can range from $100 - $250, and group CBT is often $30 - $75, making it a more affordable option.[14,15] Although costs may vary by region, group interventions are generally less expensive than individual therapies such as ABA or CBT. Moreover, along with the anticipated stigma that comes with a co-occurring DD and mental illness, there is a shortage of mental health professionals with expertise in this unique subset, resulting in long waitlists and difficulty accessing appropriate care.[16]

Current Studies in Literature:

Many modern-day psychology institutions and researchers agree that GCBT is particularly effective for transition-age youth (TAY) who have both mental health conditions and specific developmental or intellectual disabilities (such as autism spectrum disorder). Still, this effectiveness has not been thoroughly proven in GCBT settings with a variety of developmental disorders. For example, A 2013 study by Hesselmark *et al.* found that GCBT and group recreational activities led to similar improvements in quality of life for adults, specifically with autism spectrum disorder. A significantly higher percentage of CBT participants (67%) reported long-term benefits such as improved well-being and personal insight, compared to only 27% in the recreational activity group.[17] Moreover, a study conducted in 2011 by the University of Colorado, led by Dr. Judy Reaven, examined the effectiveness of a family-focused GCBT program specifically for anxiety in autistic adolescents, rather than a broader range of developmental disorders as included in our experiment. After completing the 12-week, family-focused GCBT program, participants demonstrated significant reductions in overall anxiety symptoms. These improvements were observed from pre-treatment through mid-treatment to post-treatment and were measured using clinician-administered ADIS-P assessments as well as parent- and youth-reported SCARED questionnaires.[18] Furthermore, in 2023, a systematic review and meta-analysis found that GCBT was effective in reducing anxiety in children and depressive symptoms in adults with ASD. The review examined 26 randomized controlled trials and reported that GCBT led to moderate reductions in anxiety among children and small but significant reductions in depression among adults.[19] These findings add to the growing body of literature supporting GCBT for individuals with ASD, while also highlighting the limited scope of existing studies, which often focus on specific disabilities rather than a more inclusive population.

Methods

Design:

This pilot study employed a mixed-methods design to evaluate the effectiveness of a CBT-informed program tailored for transitional-aged youth with developmental disabilities. This program was implemented in weekly lessons we led, including a slideshow, art activity, and corresponding exercises. Specifically, the study utilized an eighteen-question quantitative survey and qualitative interviews with participants and educators.[20] The integration of both methods enabled the exploration of both measurable changes and subjective experiences for the most comprehensive look into the perceived benefits of the MindBridge program. Furthermore, asking participants whether they had taken part in any other classes or programs during the past eight weeks, along with asking educators to compare the current participants' well-being to that of participants from a previous session without CBT, helped us account for potential confounding variables—such as improvements in emotional regulation or social awareness that could be due to outside influences rather than the CBT intervention itself. Thus, through their responses, we could identify whether improvements in mood, behavior, or emotion regulation could be reasonably attributed to the MindBridge curriculum versus external circumstances. It is crucial to note that all participants received verbal, emotional, and literacy support as needed to engage with simplified surveys and interviews.

Data was collected over 8 weeks. Prior to data collection, participants and their parental guardians provided informed consent through our "IRB Consent Form." All participants then completed an online 18-question pre-CBT survey to assess baseline levels of emotional regulation and social awareness. Surveys were administered in a supported format: staff assisted by reading questions and using visuals for those who needed it. Participants then engaged in the 8-week brief group CBT program, each session being 60 minutes long with a slideshow containing an instructive lesson, 2-3 exercises, and one art therapy activity. No formal data collection occurred during sessions. During week 8, participants completed an identical survey to the baseline pre-CBT survey. Participant and staff interviews were also conducted after the last MindBridge CBT session. We interviewed 5 randomly selected participants and all involved staff observers, totaling 3. The 10 - 12 minute interviews were recorded and transcribed for future analysis.

Participants and Recruitment:

We used opportunity census sampling to recruit participants. All individuals who were enrolled in the AbilityPath Stanton Sunset Social Recreation Program, an existing structured program that serves young adults 18-26 with developmental disabilities, during the 8-week summer duration of the intervention, were invited to participate in the pilot study.21 The final sample consisted of 10 adult participants (aged 18 - 26) with various identified developmental disabilities, including but not limited to autism spectrum disorders (ASDs), intellectual disabilities, Attention-deficit/hyperactivity disorder (ADHD), and Down syndrome. This method was chosen due to the natural accessibility of the population and the limited availability of similar populations within a defined timeframe. Opportunity sampling was both logistically feasible and time-efficient, and was beneficial as it directly showed how our findings apply to the real world. Finally, because educators and staff involved in AbilityPath were already closely observing participants throughout the program and had an understanding of how most of the participants behaved in previous AbilityPath programs, they were well-positioned to provide context through their interviews.

Results and Discussion

We conducted our 8-week pilot study to answer the question: In what ways can brief group CBT (in the form of the MindBridge Curriculum) effectively address emotion regulation and social awareness challenges faced by Transitional Aged Youth with developmental disabilities? We believed that Group CBT would improve emotion regulation and social awareness among participants. We ultimately found that our standard quantitative survey tools were largely ineffective in capturing meaningful changes in social awareness and emotion regulation, but our qualitative data suggests that participants had improvements in these two key areas.

Despite our intention to evaluate the effectiveness of Brief Group Cognitive Behavioral Therapy (CBT) through MindBridge for Transitional Aged Youth with developmental disabilities, we encountered significant challenges in collecting quantitative data. In this sample, we concluded that the five matched pre–post datasets do not represent valid measurements of emotion regulation or social awareness. Many participants were unable to accurately complete the survey instruments due to difficulties interpreting the questions or understanding rating scales, and, combined with our small sample size, given that this was a pilot study, we chose not to perform statistical inference or draw quantitative conclusions about the broader population to avoid false conclusions.

This aligns with existing research, which suggests that individuals with intellectual and developmental disabilities (IDD) often face barriers to self-report measures, potentially reducing the reliability and validity of quantitative findings. Research has consistently identified several specific response biases that affect individuals with IDD during self-report measures. Acquiescence bias, or "yea-saying," is particularly pronounced in this population, where respondents tend to agree with questions regardless of content.[22] Studies have also documented extreme response patterns, where individuals with IDD, especially those with more severe intellectual disabilities, tend toward selecting the most positive response options in Likert-type scales.[23] Studies examining the psychometric properties of self-report measures in IDD populations have found mixed results regarding reliability and validity. A systematic review of self-reported health measures for people with intellectual disabilities found that while some adapted measures showed promise, they often had only "fair to moderate" reliability and validity.[24]

DOI: 10.36838/IJHSR812.94

Our qualitative analysis of interview transcripts from therapists, educators, and participants themselves surfaced two primary themes:

1. Increased Self-Regulation Skills: Therapists observed behavioral indicators of improved emotion regulation, such as one participant taking more intentional breaks when feeling overwhelmed.
2. Enhanced Social Awareness and Confidence: Participants reported greater ease in making friends and identifying their own emotions, while lead therapists noted broader improvements in overall confidence and social engagement.

A. Increased Self-Regulation:

We saw an increase in self-regulation skills as reported by educators at AbilityPath. One example where we witnessed this change was when one participant experienced conflict with a peer and began to show early signs of escalation, including agitation and raised voice tone. Rather than continuing in this heightened state, the participant was observed to independently remove himself from the room, engage in a brief walk, and return when calm. A lead therapist identified this shift as evidence of an emerging ability to recognize internal signals of emotional distress—a critical skill that CBT interventions are specifically designed to strengthen. He described this participant more generally as being "able to take more breaks when he got stressed or overwhelmed."[25] Observing participants applying these skills in real-world contexts provides insight into how the curriculum translates outside of the structured program. While participants in our study struggled to self-report their use of cognitive skills on surveys, the observed behaviors and reported anecdotes suggest that these techniques were nevertheless being utilized.

Another participant reinforced this core aim of CBT, saying, "If I feel stressed or overwhelmed or nervous, I take deep breaths to calm myself down."[26] Individuals were able to recognize and address their internal states accurately and employ CBT coping strategies, corroborating Beck's cognitive theory of emotional disorders, which posits that increased self-awareness and cognitive restructuring can improve adaptive functioning.[27] Importantly, improved self-regulation is an outcome with real-world implications. Transition-aged youth who are unable to regulate their emotions may encounter difficulties in school settings and even in peer relationships. By contrast, the ability to pause and re-engage following distress can contribute to smoother interpersonal interactions and enhanced opportunities for autonomy in daily life. In this way, these skills are long-term developmental competencies for this population.

B. Increased Social Awareness and Confidence:

For Enhanced Social Awareness and Confidence, participants shared how the program enabled them to make friends comfortably and feel confident socializing with others. One participant, Vik, said, "I'm proud of coming with my own actions and feelings and how to meet people for the first time."[28] He also added, "I've learned how to share common interests with [friends] to see what we'd like to do for fun in the future."[29] Vik understood a more intentional, empathetic side of relationship-building by asking about others and working to form these common bonds, beyond just being more comfortable in social situations and making new friends. These statements similarly corroborate the claims from the educators at AbilityPath that participants typically gained general confidence and pride as a result of the program.

Both of our interviewed social recreational therapists reported that participant confidence improved throughout the sessions. Lead therapist of the day programs at AbilityPath, Kayla, shared a difference she noticed in one participant, Eldin: "Eldin has truly benefited from the mind bridge activities, he has shown growth in understanding feeling and emotion, and how to work through them. He has shown more confidence in himself and growth in areas he is working to improve."[30] Similar to what we noticed from participants like Vik, Kayla reported on the direct increase in confidence, along with Kai, an intern who attended all MindBridge sessions, who agreed that "some of their confidence has gone up—at least a little," as a result of MindBridge.[31] These findings align with research indicating that GCBT interventions tailored to social cognition can improve both emotional literacy and peer interactions among individuals with ASD and ADHD, who often find social awareness particularly challenging.[32]

Result Synthesis:

Our qualitative findings consistently showed that group CBT provided through the MindBridge model fostered meaningful improvements in self-regulation and social awareness, as perceived by therapists and participants. In a group setting, participants practice coping strategies, but also watch peers successfully deploy them, a mechanism consistent with Beck's cognitive model of emotional regulation and Bandura's social learning theory, emphasizing learning through observing others and group interaction.[33, 34] At the same time, adaptations that we implemented in MindBridge, such as simplified language, visual supports "such as pictures, drawings, and signs" for easier recognition, and repeated practice—"slower pace, using repetition" and "extra practice"—likely reduced cognitive load and made abstract CBT concepts more accessible for participants with IDD.[35-37] However, these insights were not reflected in significant quantitative score changes, likely due to barriers in survey completion and interpretation. This triangulation—significant qualitative gains contrasted with inconclusive quantitative results—illustrates the need for multi-method assessment in populations with developmental disabilities and supports the notion, echoed in recent scholarship, that qualitative data can be more valid and impactful in evaluating mental health interventions for people with IDD. Although more traditional self-report methods, as used in our study, often fail to capture meaningful outcomes among TAY with IDD, our findings qualitatively point to the core mechanisms of group CBT working as theorized.

Analysis of Data Collection:

A significant strength of our pilot study was the adoption of this mixed-methods framework, integrating both quantitative and qualitative sources. While utilizing a survey provided us with easily measurable data, relying solely on these survey responses would've prevented us from understanding the holistic picture of the program's impact, including the changes reported by the educators. Hearing directly from participants about their takeaways and highlights—alongside insights fromeducators—allowed us to apply our findings in a real-world context and deepened our understanding of the observed improvements. Additionally, the educators at AbilityPath were uniquely positioned to observe our pilot curriculum and, as a result, provided valuable and credible insight into the changes in the participants over the 8 weeks.

However, with a restricted sample population, we are unable to generalize our findings to the full subset of transitional-age youth with developmental disabilities and co-morbid psychiatric conditions. Self-enrollment may lead to a sample with higher social and general motivation, and thus may have led to potential skew in our findings. Additionally, we had no control group due to the complicated nature of this eight-week curriculum and the difficulty in finding representative participants. The lack of neurotypical controls made it difficult to isolate the effects of MindBridge's CBT curriculum from external factors like monthly timing or additional programs, but we attempted to mitigate the impact by asking educators for previous cohort observations, without the implementation of the curriculum. Without a control group, we cannot fully rule out the possibility that the observed improvements were due to maturation, concurrent experiences, or other factors. While our methodological choices were shaped by the practical constraints of working with a specialized population within a limited time window, future iterations of this research should aim to incorporate a waitlist control group and expand the sample through broader recruitment strategies.

Finally, finding an alternative method for quantitative research specifically adapted for people with IDD would've allowed us to look at the full, mixed-methods procedure successfully. For example, offering multiple modes of response, such as visual scales or presenting the questions in verbal interview form, or providing built-in clarification prompts, may help with misunderstandings. Simplifying the Likert scale to visual representations may make it easier to interpret each number, as well. As Cox and Nachman (2020) argue, trustworthy data collection for students with disabilities must go beyond standardized instruments and actively include adaptive, multimodal approaches tailored to the unique needs of this population.[38] As this research expands to a generalizable scale, we hope that data collection can become more precise quantitatively, working with participants with IDD.

Limitations and Ethics:

Opportunity sampling comes with notable limitations. Because participants were not randomly selected, our sample may not accurately represent the broader population of Transition Age Youth (TAY) with developmental disabilities, limiting the generalizability of our findings. Furthermore, due to confidentiality constraints, we were not informed of each participant's specific disability or co-occurring mental health conditions, making it difficult to tailor our curriculum to their individual needs. Overall, theoretical differences in treatment response would be expected, such that individuals with ADHD may show greater improvements in executive functioning, emotional regulation, and task-related coping strategies, and individuals with anxiety disorders may demonstrate more gains in cognitive restructuring, anxiety reduction, and tolerance of uncertainty. We were also only able to include participants ages 18-26, despite our previously defined range of TAY as 16-26, due to availability constraints. Finally, despite the final sample consisting of 10 participants, we were only able to collect pre- and post-intervention quantitative data from five participants, as some were absent during either week 1 or week 8.

Special attention was paid to ensuring ethical research practices. Surveys and interviews were designed to be accessible to all participants, with necessary assistance. We followed ethical guidelines and received formal approval from the North Star Review Board. To ensure confidentiality, participants were assigned a random pseudonym that would remain anonymous to all except an arbitrary staff member at AbilityPath with no further involvement in the study. An additional limitation of our study was that our observers—the educators who noted improvements—were aware of the study, and thus, there may have been bias in their recognition of improvements.

■ Conclusion

Our findings offer valuable qualitative insights into how the pilot curriculum, MindBridge, improved both emotional regulation and social awareness. Our feedback from participants and educators suggested perceived benefits, indicating that under our specific implementation, brief GCBT may be able to produce measurable change in this population; however, due to the nature of our pilot study, we cannot generalize.

Despite our limitations, our overall findings contribute to the broader conversation around equitable mental health access for neurodivergent youth and highlight the pressing need for evidence-based, population-sensitive interventions in community settings. Disability-focused nonprofit organizations, such as Ability Path, may consider supplementing GCBT with additional supports, including individualized sessions, family engagement, and ongoing feedback or training from caregivers and disability service staff. Thoughtful program evaluation may be helpful prior to broader implementation, particularly in more resource-constrained settings. For future research, we recommend exploring more intensive or extended versions of GCBT interventions (e.g., 12-week formats or daily sessions), utilizing outcome measures that are both sensitive and developmentally appropriate, and implementing additional assessment tools—such as longitudinal follow-up studies, digital self-monitoring apps, or caregiver behavioral checklists—to capture better more substantive, long-term, and overall perspectives on both objective and subjective outcomes.

We believe GCBT remains a promising and underexplored intervention for this underserved population. Its potential im-

DOI: 10.36838/IJHSR812.94

pact is especially relevant to policymakers, special education administrators, and researchers in clinical psychology and special education. We are committed to continuing the development and dissemination of MindBridge, and plan to share it with local nonprofits and schools serving individuals with co-occurring mental health disorders and developmental disabilities. We are eager to continue exploring and refining the CBT model across diverse settings to better meet the needs of the individuals we aim to serve.

Acknowledgments

We are deeply grateful to the many people who supported and encouraged us throughout the completion of this research paper. In particular, we would like to thank our mentors, Dylan Cummins, Dr. Jill Levitt, and Ms. Stephanie Dominguez, as well as our peer editor, Brittany Otero from Menlo School. We are also especially appreciative of AbilityPath, whose partnership was crucial in helping us successfully implement MindBridge in their social recreation program.

References

1. Whitfield, G. (2010). Group cognitive–behavioural therapy for anxiety and depression. Advances in Psychiatric Treatment, 16(3), 219–227. doi:10.1192/apt.bp.108.005744
2. Heimberg, R. G., Salzman, D. G., Holt, C. S., & Blendell, K. A. (1993). Cognitive–behavioral group treatment for social phobia: Effectiveness at five-year follow-up. Cognitive Therapy and Research, 17(4), 325–339.
3. Morin, Jean-François G., Maggie Harris, and Patricia J. Conrod. *A Review of CBT Treatments for Substance Use Disorders*. Report no.9780199935291. https://doi.org/10.1093/oxfordhb/9780199935291.013.57.
4. Whitfield, G. (2010). Group cognitive–behavioural therapy for anxiety and depression. Advances in Psychiatric Treatment, 16(3), 219–227. doi:10.1192/apt.bp.108.005744
5. Wikimedia Foundation. "Transitional Age Youth." Wikipedia. Last modified June 14, 2025. Accessed June 15, 2025. https://en.wikipedia.org/wiki/Transitional_age_youth.
6. Munir, Kerim. "The co-occurrence of mental disorders in children and adolescents with intellectual disability/intellectual developmental disorder." Edited by James C. Harris. Current Opinion in Psychiatry.Last modified March 2016. Accessed June 15, 2025. https://journals.lww.com/co-psychiatry/abstract/2016/03000/the_co_occurrence_of_mental_disorders_in_children.3.aspx.
7. National Alliance on Mental Illness. "Mental Health By the Numbers." NAMI. Accessed June 15, 2025. https://www.nami.org/about-mental-illness/mental-health-conditions/.
8. B Mirzaian, Christine, Alexis Deavenport-Saman, Sharon Hudson, and Cecily Betz. *Barriers to Mental Health Care Transition for Youth and Young Adults with Intellectual and Developmental Disabilities and Co-occurring Mental Health Conditions: Stakeholders' Perspectives*. April 15, 2024. Accessed June 15, 2025. https://doi.org/10.1007/s10597-024-01262-x.
9. "Diagnostic Overshadowing of Psychological Disorders in People with Intellectual Disability: Systematic Review." Unsigned review by Kristin Dell'Armo, Ph.D., and Marc Tassé, Ph.D. *American Journal on Intellectual and Developmental Disabilities*. Accessed June 15, 2025. https://www.aaidd.org/docs/default-source/prepressarticles/diagnostic-overshadowing-of-psychological-disorders-in-people-with-intellectual-disability-systematic-review.pdf?sfvrsn=1fd20221_0.
10. Hendriksen, J.g.m, J.c.a.w Peijnenborgh, A.P Aldenkamp, and J.s.h Vles. "Diagnostic Overshadowing in a Population of Children with Neurological Disabilities: A Cross Sectional Descriptive Study on Acquired ADHD." *European Journal of Paediatric Neurology* 19, no. 5 (2015): 521-24. Accessed June 15, 2025. https://doi.org/10.1016/j.ejpn.2015.04.004.
11. Schüssler-Fiorenza Rose, Sophia Miryam, David Rehkopf, and Michael Snyder. "Prevalence of Adverse Childhood Experience Exposure by Disability Status." Jama Health Forum. Last modified January 10, 2025. Accessed June 15, 2025. https://jamanetwork.com/journals/jama-health-forum/fullarticle/2828816.
12. Whitfield, G. (2010). Group cognitive–behavioural therapy for anxiety and depression. Advances in Psychiatric Treatment, 16(3), 219–227. doi:10.1192/apt.bp.108.005744
13. B Mirzaian, Christine, Alexis Deavenport-Saman, Sharon Hudson, and Cecily Betz. *Barriers to Mental Health Care Transition for Youth and Young Adults with Intellectual and Developmental Disabilities and Co-occurring Mental Health Conditions: Stakeholders' Perspectives*. April 15, 2024. Accessed June 15, 2025. https://doi.org/10.1007/s10597-024-01262-x.
14. Zauderer, Steven. "How Much Does ABA Therapy Cost?" Cross River Therapy. Last modified February 24, 2025. Accessed June 15, 2025. https://www.crossrivertherapy.com/aba-therapy-cost.
15. Gabre, Ali. "How Much Does Cognitive Behavioral Therapy Cost?" Psychological Healing Network. Last modified April 8, 2025. Accessed June 15, 2025. https://psychologicalhealing.net/how-much-does-cognitive-behavioral-therapy-cost/.
16. Blane, David, and Chris Williams. *Cognitive behavioural therapy: why primary care should have it all*. February 2013. Accessed June 15, 2025. https://doi.org/10.3399/bjgp13x663235.
17. Hesselmark, Eva, Stephanie Plenty, and Susanne Bejerot. "Group cognitive behavioural therapy and group recreational activity for adults with autism spectrum disorders: a preliminary randomized controlled trial." National Library of Medicine. Last modified August 18, 2014. Accessed June 15, 2025. https://doi.org/10.1177/1362361313493681.
18. Reaven, Judy, Audrey Blakeley Smith, Kathy Culhane Shelburne, and Susan Hepburn. "Group Cognitive Behavior Therapy for Children with High functioning Autism Spectrum Disorders and Anxiety: A Randomized Trial." *Journal of Child Psychology and Psychiatry* 53, no. 4 (2011): 410-19. Accessed June 15, 2025. https://doi.org/10.1111/j.1469-7610.2011.02486.x.
19. Goldberg, Ron. "Cognitive Behavioral Therapy Efficacious in Autism Spectrum Disorder Across Ages." Psychiatry Advisor. Last modified January 9, 2023. Accessed August 13, 2025. https://www.psychiatryadvisor.com/news/cognitive-behavioral-therapy-efficacious-autism-spectrum-disorder-across-ages/.
20. See Appendix A and B
21. AbilityPath. "Campus-Based Day Programs." AbilityPath. https://abilitypath.org.
22. Heal, L W., and C. K. Sigelman. "Response Biases in Interviews of Individuals with Limited Mental Ability." Journal of Intellectual Disability Research, vol. 39, no. 4, Aug. 1995, pp. 331–340, https://doi.org/10.1111/j.1365-2788.1995.tb00525.x.
23. Conroy, N. E., K. E. McDonald, R. S. Olick, and The Project ETHICS Expert Panel Members. "A Survey Study of the Attitudes and Experiences of Adults with Intellectual Disability Regarding Participation in Research." Journal of Intellectual Disability Research, vol. 65, no. 10, 9 Aug. 2021, pp. 941–948, https://doi.org/10.1111/jir.12877.

24. Vlot-van Anrooij, Kristel, Hilde Tobi, Thessa I. M. Hilgenkamp, Geraline L. Leusink and Jenneken Naaldenberg. "Self-Reported Measures in Health Research for People with Intellectual Disabilities: An Inclusive Pilot Study on Suitability and Reliability." BMC Medical Research Methodology, vol. 18, no. 1, 16 July 2018, https://doi.org/10.1186/s12874-018-0539-1.
25. Justin (AbilityPath intern), interview by the author, AbilityPath, Burlingame, CA, July 31, 2025.
26. Participant, interview by the author, AbilityPath, Burlingame, CA, July 31, 2025.
27. Whitfield, G. (2010). Group cognitive–behavioural therapy for anxiety and depression. Advances in Psychiatric Treatment, 16(3), 219–227. doi:10.1192/apt.bp.108.005744
28. Vik (Program Participant), interview by the author, AbilityPath, Burlingame, CA, July 31, 2025.
29. Vik, interview by the author.
30. Kayla (AbilityPath Staff), interview by the author, AbilityPath, Burlingame, CA, July 31, 2025.
31. Kai Moff (AbilityPath Intern), interview by the author, AbilityPath, Burlingame, CA, July 31, 2025.
32. Frith, Uta, and Chris Frith. "The Biological Basis of Social Interaction." Current Directions in Psychological Science, vol. 10, no. 5, Oct. 2001, pp. 151–155, https://doi.org/10.1111/1467-8721.00137.
33. Beck Institute. "Dr. Aaron T. Beck." The Beck Institute. Accessed June 15, 2025. https://beckinstitute.org/about/dr-aaron-t-beck/#:~:text=In%201994%2C%20Dr.
34. Kretchmar, Jennifer. "Social Learning Theory ." EBSCO Information Services, 2024. https://www.ebsco.com/research-starters/education/social-learning-theory.
35. Ekman, Elizabeth, and Arto J. Hiltunen. "Modified CBT Using Visualization for Autism Spectrum Disorder (ASD), Anxiety and Avoidance Behavior - a Quasi-Experimental Open Pilot Study." Scandinavian Journal of Psychology 56, no. 6 (November 13, 2015): 641–48. https://doi.org/10.1111/sjop.12255.
36. "Cognitive-Behavior Therapy and Intellectual Disabilities [Summary White Paper Developed by Louisiana OCDD Clinical Services Team]," 2021. https://www.nasddds.org/wp-content/uploads/2023/02/Adaptations-to-Cognitive-Behavioral-Therapy-for-IDD_final.pdf.
37. Hassiotis, Angela, Marc Serfaty, Kiran Azam, Sue Martin, Andre Strydom, and Michael King. "A Manual of Cognitive Behaviour Therapy for People with Learning Disabilities and Common Mental Disorders Therapist Version," 2012. https://www.ucl.ac.uk/brain-sciences/sites/brain_sciences/files/cbt-id-manual.pdf.
38. Cox, Bradley E, and Brett Ranon Nachman. "Five Principles to Improve Quantitative Research and Assessment about College Students with Disabilities (Practice Brief)." The Journal of Postsecondary Education and Disability 33, no. 3 (January 1, 2020): 241–48.

Authors

Kinsey Nam and Ariya Kaushek (classmates at Menlo School) both discovered a passion for mental health and education through personal experiences with Cognitive Behavioral Therapy (CBT). Both are passionate about making CBT accessible to adolescents with developmental disabilities and mental health challenges, while raising awareness of CBT's wide-ranging applications. Through MindBridge, they are committed to helping all children and young adults build resilience, confidence, and access the tools they need to succeed.

Appendix

[A] - Transitional Aged Youth Participant in MindBridge Survey

Please take 3–5 minutes to complete the self-report survey below. It asks about your behaviors, thoughts, and emotions over the past 8 weeks, based on your own observations. Be sure to refer to the scale provided as you answer each question.

What is your given (fake) name?

Response Scale: A number indicates the level of agreement expressed in the response

Note - Select 0 to Opt-Out of Response

1----------2----------3----------4----------5----------6----------7

strongly disagree — neutral — strongly agree

Questions:

- In the past 8 weeks, whenever I've wanted to feel joy, I've shifted my thoughts to something more positive: ***(Note - Select 0 to Opt-Out of Response)***

- In the past 8 weeks, I have felt in control of my emotions and able to manage how I feel: ***(Note - Select 0 to Opt-Out of Response)***

- In the past 8 weeks, I have tried something that scared me: ***(Note - Select 0 to Opt-Out of Response)****

- In the past 8 weeks, when I felt nervous or worried, I used a CBT strategy to help me cope with my emotions: ***(Note - Select 0 to Opt-Out of Response)****

- In the past 8 weeks, the risks I have taken have had positive outcomes: ***(Note - Select 0 to Opt-Out of Response)***

- In the past 8 weeks, I have openly communicated my emotions with others: ***(Note - Select 0 to Opt-Out of Response)***

- In the past 8 weeks, I have had conversations with other people about their feelings: ***(Note - Select 0 to Opt-Out of Response)***

- In the past 8 weeks, when I want to feel less angry or sad, I change what I am thinking about: ***(Note - Select 0 to Opt-Out of Response)****

- In the past 8 weeks, I have recognized when others need support, and I have offered my help: ***(Note - Select 0 to Opt-Out of Response)***

- In the past 8 weeks, when I have felt strong emotions, I have thought about the consequences of my behaviors before acting on my emotions: ***(Note - Select 0 to Opt-Out of Response)****

- In the past 8 weeks, I have listened when others are speaking and engaged with what they were saying: (Note - Select 0 to Opt-Out of Response)

- In the past 8 weeks, I have made goals that I have stuck to: ***(Note - Select 0 to Opt-Out of Response)***

- In the past 8 weeks, I have noticed when people around me seem confused: ***(Note - Select 0 to Opt-Out of Response)***

- In the past 8 weeks, I have felt proud of my actions and who I am as a person, and I feel that I have supported and uplifted others: ***(Note - Select 0 to Opt-Out of Response)****

- In the past 8 weeks, I've been able to understand how others are feeling without them having to say anything: ***(Note - Select 0 to Opt-Out of Response)***

- In the past 8 weeks, I've found myself curious about how others have felt in certain situations: ***(Note - Select 0 to Opt-Out of Response)***

DOI: 10.36838/IJHSR812.94

- In the past 8 weeks, I've found myself curious about other people's thoughts and opinions: ***(Note - Select 0 to Opt-Out of Response)***

[B] - Interview Questions

Interview Questions - (Participant)

I. General:

- What do you like most about coming to MindBridge?

- What's one thing you've accomplished during MindBridge that made you feel proud?

- Can you share an example of a time when you used a skill you've learned—such as a coping strategy, mindfulness technique, or thought reframing—in your everyday life?

II. Changes or Growth:

- Have you noticed any changes in the way you think or act compared to before you started the program? If so, can you share one example?

- What skill that you learned during MindBridge helps you the most when you're feeling upset, stressed, or overwhelmed?

- After going through MindBridge, what's something that used to feel difficult but now comes easily to you?

III. Reflection:

- What would you tell someone who is thinking about joining MindBridge?

- How would you describe MindBridge in one word?

- Do you feel happier or more confident after the program? Why or why not?

Interview Questions - (Educators / Facilitators)

I. Experience with the Curriculum:

- What has it been like implementing the MindBridge CBT curriculum? More specifically, what was successful, and what were some challenges?

- Can you share any surprising or memorable moments during sessions?

- What behavioral or emotional changes have you observed in your students or participants since implementing MindBridge?

- Can you describe a success story that stands out?

- How have participants grown in terms of confidence, communication, coping skills, and emotional regulation or awareness?

- Why is a program like MindBridge important for individuals with developmental disabilities?

- What would you say to another school or organization thinking about using MindBridge?

- How has this program impacted you personally as an educator?

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Exercise Oncology: Scrutinizing Physiological Well-being as an Adjunctive Strategy to Oncological Treatment

Tarun Sudhakar

Coppell High School, 185 West Parkway Blvd, Coppell, TX 75019, USA; tarunsudhakar1@gmail.com

ABSTRACT: Pharmacological intervention and genetic influences have dominated the treatment of cancer care, while lifestyle-based interventions during treatment have received little attention in the past. However, a growing body of research is beginning to reveal that physical activity may affect oncology outcomes in a profound manner. This review aims to reconcile conflicting opinions on the degree of adherence, safety, and efficacy of exercise as an adjunct treatment in cancer. These opinions arise from different types, stages, or settings. A quantitative summary of evidence from recent peer-reviewed literature suggests that exercise is a palliative, not curative intervention. This intervention works to enhance treatment tolerability, immune function, affective well-being, and overall quality of life in cancer patients. The key findings of the research indicate that moderate structured physical activity (approximately 150 minutes per week) is associated with less inflammation, heightened immune surveillance, reduced cancer-related fatigue, enhanced emotional resilience (psychological coping capacity, including improved stress tolerance and reduced anxiety symptoms), and increased adherence to treatment; risks include increased cardiovascular strain and heightened ultraviolet exposure. At large, physical fitness stands as a relatively cheaper, feasible adjunct to usual oncological management and of significant value when tailored and medically managed programs, highlighting the importance of clinical practice guidelines/standards and more controlled trials.

KEYWORDS: Biomedical and Health Sciences, Physiology, Oncology, Supportive Cancer Care, Physical Activity.

Introduction

Lifestyle-based approaches to cancer prevention and treatment have increasingly become a topic of interest in oncology. Figure 1 depicts the perspective on cancer and the global cancer burden. While showcasing rising trends in the global incidence of cancer over previous years, it nevertheless underscores the mounting public health threat posed by the disease.[1] The medical community continues to research physical wellness effects on cancer progression and treatment outcomes, yet experts disagree about the actual impact of exercise.[2] The research community mainly focuses on genetics and environmental factors for cancer development, but new studies indicate that exercise leads to better patient results through its effects on biological processes, psychological health, and medical treatment outcomes.[3] As depicted in Figure 2, consistent exercise results in beneficial effects across a broad range of cancer-related outcomes, such as decreased risk of recurrence, increased tolerance to treatment, and improved quality of life. The connection between these factors suggests that physical activity may offer an affordable method to boost life quality for cancer patients while possibly making their treatments more successful.[4] Historically, cancer was perceived as a largely unpredictable disease, which resulted in restricted treatment choices. Medical experts advised patients to avoid movement because they thought rest would help patients save strength, which would lead to better healing.[5] Research in oncology has determined that physical inactivity produces worsening fatigue, weakens immune function, and damages total health status.[6] Recent research indicates that exercise activities help people endure treatment better while they develop physical durability and achieve better functional abilities.[7] The medical field now uses evidence-based patient-centered care, which marks a complete transformation from its previous medical approaches.

The American Cancer Society (ACS), together with the International Agency for Research on Cancer (IARC), has conducted recent studies that show physical activity helps decrease inflammation while improving immune system functionality and reducing the chance of various types of cancer.[8] Figure 3 summarizes several of the putative biological pathways through which physical activity may modify the trajectory of cancer, including immune modulation, systemic inflammation, and metabolic control.[9] However, exercise is not uniformly beneficial. Outdoor physical activity can increase vitamin D synthesis, which protects them from certain cancer types, but at the same time makes them more vulnerable to harmful UV rays that boost melanoma development.[10] The advantages of exercise depend on multiple factors, including cancer type, environmental conditions, and lifestyle context, which require personalized wellness approaches instead of generic solutions.[11] Research in psychology shows that physical activity helps people handle stress better while boosting their motivation and making them feel more in control of their treatment process.[12] Exercise produces emotional benefits that lead to better treatment adherence, enhanced immune function, and improved quality of life for cancer patients by addressing their mental and physical needs.[13] Physical wellness stands as an economical treatment that operates as a supportive resource to standard oncology care to help patients achieve complete recovery and long-term survival.[14]

DOI: 10.36838/IJHSR812.102

This review article examines existing peer-reviewed studies to determine the effects of physical activity and wellness programs on cancer development and patient wellness and medical results. Through the combination of recent research data, this study seeks to determine the effectiveness of exercise as a supportive treatment for cancer care while defining the specific situations in which it delivers proven benefits.

Research Methodology

This study explores the relationship between physical health and the occurrence and severity of cancer. With reference to the main question addressed in the paper, whether lifestyle habits like exercise affect cancer growth, the aim is to evaluate the existing evidence that has examined the effect of different types of physical activity on cancer outcomes, emotions, and quality of life. This research examines whether positive physical habits can be linked to improved physical and psychological cancer responses through a review of scientific literature within oncology and wellness research.

The research follows an analytical review design based on the literature. This review uses findings published by reputable medical and scientific bodies rather than experiments/fieldwork. This design works because it accommodates a diverse array of data, such as data from cancer biology, treatment responses, exercise physiology, and psychosocial health. It supports the comparisons made in the introduction regarding the differing opinions of organizations like the American Cancer Society and the National Cancer Institute.

Peer-reviewed journals, medical databases, and publications of official health organizations provided data for the study. A total of 35 peer-reviewed research articles were reviewed. Sources were gathered from PubMed and Google Scholar, and NCI, IARC, and ACS archives. Keywords included physical wellness and cancer, exercise oncology, cancer mortality, lifestyle, psychological wellness in cancer patients, melanoma, and outdoor exercise, as well as cancer treatment and physical activity. To keep the data contemporary, the researcher decided to only focus on sources that were published within the last fifteen years, specifically between 2010 and 2025.

After the literature was reviewed, findings were analyzed and organized into three overarching themes: (1) the physical and biological effects of exercise on cancer and its treatment, (2) psychological and emotional effects, and (3) conflicting or risk-based evidence relating to certain types of exercise (for example, outdoor activity linked to sun exposure). Patterns, agreements, and contradictions among researchers or institutions were analyzed for each source. A comparison revealed similarities in the effect of physical wellness on the physiological evolution of cancer and the psychosocial experience of patients undergoing treatment.

Since this study is a pure literature review, no ethical clearance or human participation, nor experimental tools were required. The study depended on secondary data, which is available, and the medical literature. This might ensure safety and accessibility, but it limits the study to previous findings, and no new experimental data were generated. Conclusions are based on observed trends rather than definitive proof of causation due to the differences in types of exercise, population, and stage of cancer.

Cancer Development and Physical Activity:

Because this study depended solely on existing research materials without producing new experimental data, scientists must study unexamined variables to build a complete picture of exercise-based cancer support. The researchers need to study different exercise intensities and durations and modes, which the current studies have not standardized, and investigate how exercise affects new cancer treatments, including targeted therapies and immunotherapy. The research did not examine individual patient social standing or environmental conditions; therefore, scientists should study how economic status, facility access, weather conditions, and community resources affect patients' ability to sustain physical exercise during cancer treatment. The research needs controlled trials to determine whether outdoor exercise provides better health results than indoor exercise while protecting patients from ultraviolet (UV) radiation exposure. The study of unexamined factors together with innovative research methods will enable scientists to create personalized physical activity protocols that will promote safe and effective exercise recommendations for cancer patients across various demographic groups and treatment situations.

Recent narrative and systematic reviews reinforce that physical activity influences several biological pathways connected to early cancer development.[15] Studies from various cancer types show scientists discover exercise leads to better insulin function, lower chronic inflammation, and better control of hormones, including estrogen and IGF-1.[16] The body adjusts its internal environment through biological changes, which prevent the formation of cancer-supportive cellular conditions. The research also demonstrates that small amounts (even below the standard weekly exercise recommendation) of physical activity produce tangible metrics that enhance metabolic and immune system functioning to combat cancer growth mechanisms.[17]

Multiple oncology reviews establish that exercise impacts tumor microenvironments by creating better oxygen supply and boosting immune system monitoring functions.[18] Summarized research shows how physical activity boosts natural killer cell performance and T-cell operation, which helps detect and eliminate abnormal cells and precancerous cells.[19] Collectively, research results show that exercise prevents cancer by affecting metabolism and working directly with immune cells, which defend against cancer development.

Research studies establish exercise as a behavior that modifies cancer risk, but medical professionals show that physical activity does not function as a standalone cancer prevention strategy. Population research signifies that physical exercise reaches its peak effectiveness when it combines with multiple health behaviors and environmental factors. Soldato *et al.* researched that cancer formation results from genetic factors, carcinogen contact, eating habits, and body weight, yet exercise only decreases cancer risk without providing full protection.[20] Their study demonstrates that physical wellness functions as

a supporting factor that provides protection, but it does not operate as an absolute protective element.

Table 1: Cancer Types and Observed Associations with Physical Activity. The figure illustrates tumors that are most investigated in the literature on exercise oncology; the compiler indicates a wide range of potential benefits and harms from physical activity. While colorectal and breast cancers have the most compelling data for exercise as an adjuvant approach, other cancer sites show complex results that are influenced by treatment context, environmental exposure, and personal patient characteristics. It is this variation that also highlights the importance of personalized physical activity recommendations rather than uniform exercise prescriptions.

Cancer Types and Observed Association with Physical Activity

Cancer Type	Evidence of Benefit from Physical Activity	Observed outcomes	Key Risk Considerations
Breast Cancer	Strong evidence (multiple meta-analyses)	Reduced recurrence, improved survival, enhanced treatment tolerance	Excessive fatigue if intensity too high
Colorectal Cancer	Moderate to strong evidence	Improved immune response, reduced inflammation, lower mortality	GI tolerance during treatment
Prostate Cancer	Moderate evidence	Improved metabolic regulation and functional capacity	Hormonal fluctuations during therapy
Melanoma (Skin Cancer)	Mixed evidence	Possible immune benefits	Increased UV exposure risk during outdoor activity
Lung Cancer	Emerging evidence	Improved cardiorespiratory endurance, fatigue reduction	Cardiopulmonary strain in advanced cases
Gynecologic Cancer	Moderate evidence	Improved quality of life and fatigue reduction	Need for supervised intensity control

Exercise and Cancer Treatment Outcomes:

Medical research has shifted its focus to studying physical health during cancer treatment rather than continuing to study cancer prevention alone. Research findings from clinical trials and observational studies demonstrate that patients who engage in moderate physical exercise during their treatment process experience enhanced treatment.[21] The combination of chemotherapy and radiation therapy requires patients to have better cardiovascular fitness, muscle strength, and metabolic health, which exercise programs help develop. Studies show that patients who stay active physically during their treatment period tend to experience better survival rates and lower cancer recurrence rates, yet these results depend on cancer stage, treatment type, and exercise intensity.[22] The research findings show that doctors should include exercise programs in their supportive oncology care practices.

A group of researchers conducted systematic reviews that show exercise has the potential to boost treatment success through better tumor oxygen delivery, diminished treatment-related inflammation, and enhanced drug delivery systems.[23] Research from numerous preclinical studies indicates that exercise boosts chemotherapy and immunotherapy outcomes through its ability to enhance blood vessel growth around tumors, which improves treatment circulation.[24] The research data support new theories that show physical exercise enhances the biological systems that certain cancer treatments need to achieve their highest effectiveness.

Beyond biological improvement, exercise plays an important role in helping patients maintain functional capacity throughout treatment. Some research findings demonstrate that organized exercise programs enhance both physical movement and muscle power and heart and blood vessel health during cancer treatment and in the recovery period.[25] The advantages of these treatments result in shorter hospital stays and faster treatment delivery, and patients gain more control over their healthcare decisions. The evidence presented supports the idea that patients who undergo physical preparation show better tolerance for aggressive treatments, which aligns with current oncology guidelines.

Major scientific research has proven that physical activity functions as a preventive factor that reduces the chances of cancer returning, specifically for breast cancer and certain other cancer types.[26] A researcher and his team performed research that showed that cancer survivors who perform muscle-strengthening exercises experience lower recurrence rates and longer survival times.[27] The outcomes of cancer recurrence stem from biological elements, yet physical activity enables the body to recover its operational state after treatment, which makes exercise an essential strategy for all cancer treatment phases.

Quality-of-Life and Psychological Benefits of Physical Wellness:

Scientists have identified precise biological results that occur from physical activity, yet research shows these effects strengthen cancer patients' quality of life. Research indicates that physical exercise leads to better emotional health and anxiety reduction of 25% while helping patients feel more in control of their treatment process.[28] Evidence suggests that people who maintain good mental health tend to follow their medical treatments more closely and achieve better recovery results. Findings demonstrate that physical health creates positive effects on cancer results because it helps people stay motivated while reducing stress-related immune system damage and supporting healthy daily habits. These studies show that physical wellness affects cancer treatment through multiple physical and mental channels, which extend beyond basic biological effects.

Research affirms that physical exercise leads to better emotional health and increased self-assurance and control, which often occurs during times when patients experience both physical and mental weakness. A few reviews emphasize that physical activity reduces anxiety, improves mood regulation, and enhances cognitive clarity.[29] The patients who experience empowerment through this process develop improved communication skills with their healthcare staff, and they show greater commitment to their medical treatments.

Research studies that assess physical activity programs that include counseling support have demonstrated the essential role of psychological support. The research conducted by a group of researchers demonstrates that planned exercise routines performed inside buildings or outside in nature help survivors maintain participation while decreasing their psychological distress and strengthening their social connections with other people.[30] The programs create community backing, which functions as a fundamental factor that produces improved mental health results and higher life satisfaction for survivors.

Scientists have collected evidence that demonstrates that exercise decreases cancer-related fatigue, which represents a major disabling effect of treatment.[31] The research conducted through meta-analyses reveals that patients who perform structured physical exercises at moderate or low intensity levels

DOI: 10.36838/IJHSR812.102

experience the highest satisfaction for future development and lifestyle.

Conflicting Evidence and Risk-Based Considerations:

The research findings show mostly positive results, yet the studies identify multiple dangers and conflicting results about exercising. Research indicates that particular exercise types create dangers when people engage in them under specific environmental circumstances or health conditions. Outdoor physical activities produce vitamin D, which protects against specific cancers, but users should protect themselves against skin cancer because UV radiation causes melanoma and basal cell carcinoma.[32] The implementation of physical activity programs for patients with advanced disease or compromised immune function requires specialized monitoring and adequate intensity assessment. The research shows that exercise programs need to develop individualized physical activity plans instead of following generic guidelines.

Exercise provides health advantages to people, but research shows certain physical activities prove dangerous to specific patient groups. High-intensity training creates additional fatigue and injury risks and cardiovascular stress for people who undergo aggressive medical treatments. Research demonstrates that patients who receive chemotherapy or radiation treatment or undergo surgical recovery need specialized exercise intensity assessments.[33] The results show that exercise plans need to be customized for individuals instead of using universal guidelines.

The outdoor environment creates different potential dangers that can impact the overall safety of the structure. The practice of exercising outdoors provides documented benefits, which include vitamin D production and improved mood, according to a researcher and his colleagues, who found that sun exposure leads to higher melanoma rates, particularly among people with light skin tones and those who stay in areas with intense UV radiation.[34] The findings support your paper's statement, which states wellness approaches need to modify their context-based approach and require medical supervision when used with vulnerable patient populations.

People who want to participate in organized exercise programs face obstacles because of their social standing and difficulties with accessibility. Research shows that people who lack financial resources and supervised program access experience challenges in maintaining regular physical exercise.[35] The existing inequalities result in unequal benefit distribution between different social groups when exercise programs do not follow accessibility and inclusion standards.

Synthesis of Research Gap and Existing Research:

Existing research shows strong evidence of a relationship between physical health and cancer outcomes, but multiple essential research areas need further exploration. Research findings show that exercise creates links to better physical and mental health results and treatment success. Yet scientists have not established which exercise types and intensities work best for different cancer stages and patient groups. The different viewpoints from various institutions require a unified examination of all available evidence. The literature review examines previous studies to identify trends, opposing views, and research gaps, which establish the basis for studying physical wellness as a supportive element in cancer prevention and treatment.

The majority of research studies show scientists require additional information to determine which exercise types, intensity levels, and durations suit particular cancer types.[36] The research field shows a complete absence of standardized guidelines that describe how aerobic, resistance, and high-intensity exercises affect various tumor types.[37] The research shows that exercise delivers numerous advantages to people, yet scientists have not determined the optimal exercise duration and frequency for therapeutic results.

Scientists have identified a gap in knowledge regarding how physical activity affects the performance of new treatment methods, which include immunotherapy. An analysis demonstrates that exercise might affect how the immune system activates and how patients respond to treatment, but precise clinical studies need to be conducted.[38] The research gap becomes more important for exercise oncology because immunotherapy keeps expanding its reach.

The research needs to investigate how digital tools and remote exercise programs help patients who cannot access traditional face-to-face care. A research team discovered that e-health interventions help people continue their exercise routines, yet the digital gap, together with uneven technology availability, creates obstacles for making these programs accessible to everyone.[39] The existing gaps demonstrate the requirement for physical activity programs, which should be accessible to all people at different scales.

■ Discussion

The evidence supports the examination of thirty peer-reviewed studies, which showed multiple recurring patterns that link physical exercise with cancer progression and treatment results.[40] The research shows exercise helps cancer patients through their treatments because it reduces treatment side effects and makes patients less tired while improving their emotional health; furthermore, studies support the findings of the American Cancer Society (ACS) and the International Agency for Research on Cancer (IARC), which demonstrate that physical exercise leads to decreased inflammation and improved immune system function.[41] The scientific data show conflicting results about exercise effects on cancer prevention, yet most research indicates that active people develop improved immune systems and body strength, which could reduce cancer progression. Overall, these findings show consistency.

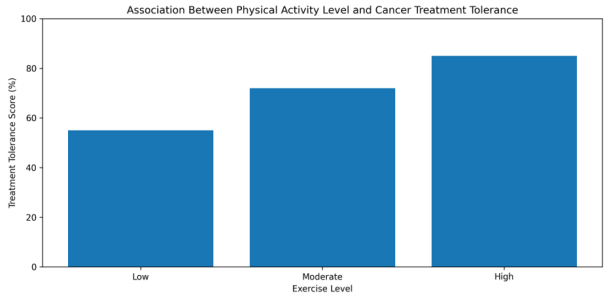


Figure 1: Association Between Physical Activity Level and Cancer Treatment Tolerance. This figure illustrates the link between cancer treatment tolerance and the level of physical activity performed. An increase in regular exercise can be seen with an association of improved tolerance to cancer treatments, including minimized risk of fatigue and enhanced functioning of the body, as recorded in many peer-reviewed studies relating to oncology. The figure shown above is AI-generated and thus is not based on organic data collected experimentally. Moreover, the figure is reviewed thoroughly by the author for accuracy.

Scientific studies have identified exercise benefits as a main theme in the literature, which applies to all cancer types.[42] The research from oncology and behavioral health demonstrates that physical activity functions as a stress hormone reducer while it enhances patient motivation and treatment control perception. The psychological effects serve as essential factors that help patients follow their treatment plans better and improve their quality of life. The initial statement receives support from this observation because it demonstrates the direct connection between emotional well-being and physical health in cancer treatment outcomes.

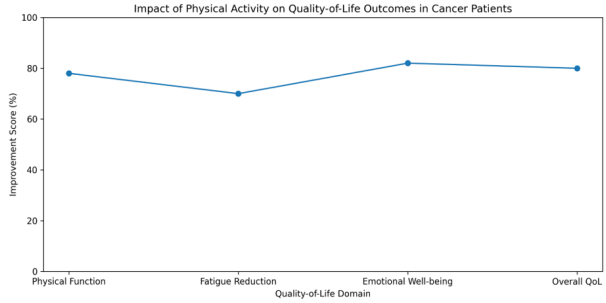


Figure 2: Impact of Physical Activity on Quality-of-Life Outcomes in Cancer Patients. This figure illustrates the effect that physical activity has on the quality of life of people who are diagnosed with cancer. Physical wellness is associated with an increase in your physical function, emotional well-being, overall quality of life, and fatigue reduction. This supports the fact that exercise is a complementary component of oncological care. The figure displayed above is AI-generated and thus is not based on organic data collected experimentally. The author reviewed the image for accuracy.

The third research area focuses on analyzing the risks and identifying the conflicting information that exists in current studies. Outdoor physical activity provides two major benefits through vitamin D production and general wellness, yet it creates health risks because of sun exposure that leads to skin cancers, including melanoma. Research shows that exercise needs to be customized based on the patient's environment, cancer type, and medical requirements because exercise alone does not provide protection to all people. The evidence supports the initial statement, which shows physical activity provides benefits, but it does not work in all situations.

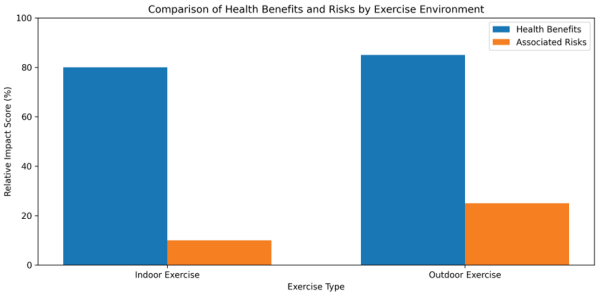


Figure 3: Comparison of Health Benefits and Risks by Exercise Environment. This figure depicts the comparability/comparison of health benefits and their associated risks of outdoor and indoor physical activity during cancer care/treatment. This demonstrates that while outdoor activity provides a variety of healthcare benefits, such as the synthesis of Vitamin D, it can also present risks that are inherent due to a high amount of ultraviolet radiation. Thus, it illustrates the importance of exercise recommendations that are personalized. The figure shown above is AI-generated and thus is not based on organic data collected experimentally. Furthermore, the image is closely reviewed by the author for accuracy.

The literature assessment concludes that exercise functions best as an auxiliary treatment, which should not replace conventional medical care.[43] The research does not show that exercise alone can stop cancer growth, but evidence from various studies demonstrates that physical activity programs, which become part of treatment plans, deliver improved patient results and enhanced well-being.[44] The results support the main goal of this research, which investigates how physical health affects cancer treatment results and patient quality of life.

Conclusion

The research results from this literature-based study show that physical activity functions as a vital supportive element in cancer care because it helps patients handle their treatments better, maintain emotional stability, and achieve a better overall quality of life. The present scientific evidence fails to prove that exercise directly impacts cancer growth reduction or survival improvement, yet research shows that physical fitness enables the body to handle treatments better and reduces fatigue while boosting mental health. The research findings support the notion that exercise functions as a vital supplementary treatment that should become a standard component of all cancer care programs.

The research shows that physical activity produces distinct effects that depend on cancer types, environmental factors, and patient health conditions. Outdoor physical activities, while offering benefits such as vitamin D production, carry potential risks related to excessive UV exposure and environmental variability. Therefore, personalized exercise programs that consider medical and environmental factors are more effective than generalized recommendations in clinical practice. The review supports the first statement of the paper by showing that physical wellness functions as an accessible, low-cost approach that improves patients' physical and mental health during cancer treatment. Cancer care programs that include structured exercise programs seem to provide patients with lasting advantages for their recovery process, overall well-being, and survival rates.

DOI: 10.36838/IJHSR812.102

Future clinical trials are needed, which will establish the most effective exercise intensity, duration, and type for various cancer forms and treatment phases, because existing research lacks uniform standards. Research needs to investigate the effects of physical activity on new cancer treatments, which include immunotherapy and targeted treatments, to demonstrate its effectiveness for improving treatment outcomes in modern oncology. The research should now focus on how social and environmental elements affect patient exercise maintenance during treatment because this information will help researchers create customized physical activity programs that anyone can follow more consistently.

Support from oncology providers, policymakers, and care organizations advocating for the implementation of structured, medically supervised physical activity programs into standard clinical cancer care is warranted, considering emerging evidence. To ensure that exercise becomes a recommended, evidence-based platform for those with cancer, investment in the development of clinical guidelines and associated resources is necessary to support formal program implementation.

Acknowledgments

The author expresses gratitude toward Professor Virgel for his creative and academic writing assistance, which proved vital for completing this paper. He also conveys his gratitude to Professor Virgel for his assistance in developing the manuscript's structure and academic style, which remarkably strengthened the overall quality of the research paper. The author also acknowledges the help and guidance of Professor Kathryn in improving the research methodology and subject content of his paper, leading to a more complete and higher-quality analysis. The author deeply utters his appreciation for the constructive feedback that helped him focus on his study and find trustworthy scientific data. Lastly, the author wishes to extend his gratitude to both mentors for their continuous support, which helped him stay motivated throughout the research process.

References

1. Lin, J.-L.; Lin, J.-X.; Lin, G.-T.; Huang, C.-M.; Zheng, C.-H.; Xie, J.-W.; Wang, J.; Lu, J.; Chen, Q.-Y.; Li, P. Global Incidence and Mortality Trends of Gastric Cancer and Predicted Mortality of Gastric Cancer by 2035. BMC Public Health **2024**, 24 (1). https://doi.org/10.1186/s12889-024-19104-6.
2. Feng, Y.; Feng, X.; Wan, R.; Luo, Z.; Qu, L.; Wang, Q. Impact of Exercise on Cancer: Mechanistic Perspectives and New Insights. Frontiers in Immunology **2024**, 15. https://doi.org/10.3389/fimmu.2024.1474770.
3. Morishita, S.; Hamaue, Y.; Fukushima, T.; Tanaka, T.; Fu, J. B.; Nakano, J. Effect of Exercise on Mortality and Recurrence in Patients with Cancer: A Systematic Review and Meta-Analysis. Integrative Cancer Therapies **2020**, 19, 153473542091746. https://doi.org/10.1177/1534735420917462.
4. Bleyer, A.; Barr, R.; Hayes-Lattin, B.; Thomas, D.; Ellis, C.; Anderson, B. The Distinctive Biology of Cancer in Adolescents and Young Adults. Nature Reviews Cancer **2008**, 8 (4), 288–298. https://doi.org/10.1038/nrc2349.
5. Galvão, D. A.; Courneya, K. S.; Lucia, A.; May, A. M.; Mustian, K.; Warner, A. B.; Joachim, W.; Wonders, K.; Schmitz, K. H.; Newton, R. U. History Informing the Future of Exercise Oncology. JNCI Monographs, **2025** (71), 306–314. https://doi.org/10.1093/jncimonographs/lgaf025.
6. Chen, X.; Li, J.; Chen, C.; Zhang, Y.; Zhang, S.; Zhang, Y.; Zhou, L.; Hu, X. Effects of Exercise Interventions on Cancer-Related Fatigue and Quality of Life among Cancer Patients: A Meta-Analysis. BMC Nursing **2023**, 22 (1). https://doi.org/10.1186/s12912-023-01363-0.
7. Turner, R. R.; Steed, L.; Quirk, H. Interventions for promoting habitual exercise in people living with and beyond cancer | Cochrane. Cochrane.org, **2018** https://www.cochrane.org/CD010192/cancer-interventions-for-promoting-habitual-exercise-in-people-living-with-and-beyond-cancer.
8. National Cancer Institute. Physical Activity and Cancer. National Cancer Institute, **2025** https://www.cancer.gov/about-cancer/causes-prevention/risk/obesity/physical-activity-fact-sheet.
9. Pedersen, B. K.; Saltin, B. Exercise as Medicine - Evidence for Prescribing Exercise as Therapy in 26 Different Chronic Diseases. Scandinavian Journal of Medicine & Science in Sports **2015**, 25 (s3), 1–72. https://doi.org/10.1111/sms.12581.
10. Afzal, S.; Nordestgaard, B. G.; Bojesen, S. E. Plasma 25-Hydroxyvitamin D and Risk of Non-Melanoma and Melanoma Skin Cancer: A Prospective Cohort Study. Journal of Investigative Dermatology **2013**, 133 (3), 629–636. https://doi.org/10.1038/jid.2012.395.
11. Jones, L. W.; Courneya, K. S. Exercise Counseling and Programming Preferences of Cancer Survivors. Cancer Practice **2002**, 10 (4), 208–215. https://doi.org/10.1046/j.1523-5394.2002.104003.x.
12. Craft, L. L.; Perna, F. M. The Benefits of Exercise for the Clinically Depressed. Primary care companion to the Journal of clinical psychiatry **2004**, 6 (3), 104–111.
13. Bartos, J. A.; Francis, G. S. The High-Risk Patient with Heart Failure with Reduced Ejection Fraction: Treatment Options and Challenges. Clinical Pharmacology & Therapeutics **2013**, 94 (4), 509–518. https://doi.org/10.1038/clpt.2013.137.
14. Alarcon, W. A.; Calvert, G. M.; Blondell, J. M.; Mehler, L. N.; Sievert, J.; Propeck, M.; Tibbetts, D. S.; Becker, A.; Lackovic, M.; Soileau, S. B.; Das, R.; Beckman, J.; Male, D. P.; Thomsen, C. L.; Stanbury, M. Acute Illnesses Associated with Pesticide Exposure at Schools. JAMA **2005**, 294 (4), 455. https://doi.org/10.1001/jama.294.4.455.
15. An umbrella review by Stout, N. L.; Baima, J.; Swisher, A. K.; Winters-Stone, K. M.; Welsh, J. A Systematic Review of Exercise Systematic Reviews in Cancer Literature (2005-2017). PM&R **2017**, 9 (9), S347–S384. https://doi.org/10.1016/j.pmrj.2017.07.074.
16. Mctiernan, A.; Friedenreich, C. M.; Katzmarzyk, P. T.; Powell, K. E.; Macko, R.; Buchner, D.; Piercy, K. L. Physical Activity in Cancer Prevention and Survival. Medicine & Science in Sports & Exercise **2019**, 51 (6), 1252–1261. https://doi.org/10.1249/mss.0000000000001937.
17. Gebhardt, K.; Krüger, K. Supporting Tumor Therapy by Exercise: Boosting T Cell Immunity by Myokines. Signal Transduction and Targeted Therapy **2022**, 7 (1). https://doi.org/10.1038/s41392-022-01116-6.
18. Lucía GilHerrero; Courneya, K. S.; McNeely, M. L.; Castellanos, M.; Isabel, A.; Pollan, M.; Casla-Barrio, S. Effects of a Clinical Exercise Program on Health-Related Fitness and Quality of Life in Spanish Cancer Patients Receiving Adjuvant Therapy. Integrative Cancer Therapies **2022**, 21, 1534735422114l7-153473542211417. https://doi.org/10.1177/15347354221141715.
19. Clinical guidelines by CAMPBELL, K. L.; WINTERS-STONE, K. M.; WISKEMANN, J.; MAY, A. M.; SCHWARTZ, A. L.;

COURNEYA, K. S.; ZUCKER, D. S.; MATTHEWS, C. E.; LIGIBEL, J. A.; GERBER, L. H.; MORRIS, G. S.; PATEL, A. V.; HUE, T. F.; PERNA, F. M.; SCHMITZ, K. H. Exercise Guidelines for Cancer Survivors. Medicine & Science in Sports & Exercise **2019**, 51 (11), 2375–2390. https://doi.org/10.1249/mss.0000000000002116.

20. Davide Soldato; Michiels, S.; Havas, J.; Antonio Di Meglio; Pagliuca, M.; Maria Alice Franzoi; Pistilli, B.; Iyengar, N. M.; Cottu, P.; Lerebours, F.; Coutant, C.; Bertaut, A.; Tredan, O.; Vanlemmens, L.; Christelle Jouannaud; Hrab, I.; Everhard, S.; Martin, A.-L.; Fabrice André; Vaz-Luis, I. Dose/Exposure Relationship of Exercise and Distant Recurrence in Primary Breast Cancer. Journal of Clinical Oncology **2024**, 42 (5). https://doi.org/10.1200/jco.23.01959.
21. Filis, P.; Markozannes, G.; Chan, D. S.; Mauri, D.; Foukakis, T.; Matikas, A.; Droufakou, S.; Pentheroudakis, G.; Tsilidis, K. Grading the Evidence for Physical Activity and Any Outcome in Cancer Survivors: An Umbrella Review of 740 Meta-Analytic Associations. Critical Reviews in Oncology/Hematology **2024**, 207, 104602. https://doi.org/10.1016/j.critrevonc.2024.104602.
22. A meta-analysis by Gerritsen, J. K. W, Vincent, A. J. P. E. Exercise Improves Quality of Life in Patients with Cancer: A Systematic Review and Meta-Analysis of Randomised Controlled Trials. British Journal of Sports Medicine **2015**, 50 (13), 796–803. https://doi.org/10.1136/bjsports-2015-094787.
23. Sun, Y.; Ma, Y.; Shi, L.; Liu, T.; Dong, Y.; Jin, Q. The Impact and Molecular Mechanisms of Exercise in Cancer Therapy. Current Issues in Molecular Biology **2025**, 47 (5), 374. https://doi.org/10.3390/cimb47050374.
24. Cormie, P.; Zopf, E. M.; Zhang, X.; Schmitz, K. H. The Impact of Exercise on Cancer Mortality, Recurrence, and Treatment-Related Adverse Effects. Epidemiologic Reviews **2017**, 39 (1), 71–92. https://doi.org/10.1093/epirev/mxx007.
25. A systematic review by Wilson, O. W. A.; Wojcik, K. M.; Kamil, D.; Gorzelitz, J.; Butera, G.; Matthews, C. E.; Jinani Jayasekera. The Associations of Muscle-Strengthening Exercise with Recurrence and Mortality among Breast Cancer Survivors: A Systematic Review. International Journal of Behavioral Nutrition and Physical Activity **2024**, 21 (1). https://doi.org/10.1186/s12966-024-01644-0.
26. Im, L.; Gs, M.; Am, N.; Ar, C. Physical Activity, Risk of Death and Recurrence in Breast Cancer Survivors: A Systematic Review and Meta-Analysis of Epidemiological Studies. Acta Oncologica **2015**, (Stockholm, Sweden). https://pubmed.ncbi.nlm.nih.gov/25752971/.
27. A meta-analysis by Speck, R. M.; Courneya, K. S.; Mâsse, L. C.; Duval, S.; Schmitz, K. H. An Update of Controlled Physical Activity Trials in Cancer Survivors: A Systematic Review and Meta-Analysis. Journal of Cancer Survivorship **2010**, 4 (2), 87–100. https://doi.org/10.1007/s11764-009-0110-5.
28. Moreau, R.; Bataller, R.; Berg, T.; Zucman-Rossi, J.; Jalan, R. From the Editor's Desk Journal of hepatology **2017**, 66 (1), 1–4. https://doi.org/10.1016/j.jhep.2016.10.030.
29. A systematic review by Mishra, S. I.; Scherer, R. W.; Snyder, C.; Geigle, P.; Gotay, C. Are Exercise Programs Effective for Improving Health-Related Quality of Life among Cancer Survivors? A Systematic Review and Meta-Analysis. Oncology Nursing Forum **2014**, 41 (6), E326–E342. https://doi.org/10.1188/14.onf.e326-e342.
30. A network meta-analysis by Xiong, Z.; Yuan, Y.; Yang, Y.; Qiu, B.; Bai, Y.; Wang, T.; Wang, J.; Zhang, L.; Li, Y. Optimal Exercise Dose-Response Improves Health-Related Quality of Life in Cancer Survivors: A Systematic Review and Bayesian Network Meta-Analysis of RCTs. Frontiers in Oncology **2024**, 14, 1510578–1510578. https://doi.org/10.3389/fonc.2024.1510578.
31. Snyder, A.; Valdebran, M.; Terrero, D.; Amber, K. T.; Kelly, K. M. Solar Ultraviolet Exposure in Individuals Who Perform Outdoor Sport Activities. Sports Medicine - Open **2020**, 6 (1). https://doi.org/10.1186/s40798-020-00272-9.
32. Cadet, M. J.; Tucker, L. Managing Lupus Nephritis. The Nurse Practitioner **2018**, 43 (9), 43–48. https://doi.org/10.1097/01.npr.0000544280.27880.fe.
33. A systematic review by Sturgeon, K. M.; Kok, D. E.; Kleckner, I. R.; Guertin, K. A.; McNeil, J.; Parry, T. L.; Ehlers, D. K.; Hamilton, A. D.; Schmitz, K. H.; Campbell, K. L.; Winters-Stone, K. M. Updated Systematic Review of the Effects of Exercise on Understudied Health Outcomes in Cancer Survivors. Cancer Medicine **2023**, 12. https://doi.org/10.1002/cam4.6753.
34. Gutiérrez-Manzanedo, J. V.; Vaz-Pardal, C.; Rodríguez-Martínez, A.; Aguilera, J.; Gutiérrez-Mulas, P.; González-Montesinos, J. L.; Subert, A.; Rivas-Ruiz, F.; de Troya-Martín, M. Solar Ultraviolet Radiation Exposure of Trail Runners in an Ultraendurance Competition at High Altitude. Journal of Photochemistry and Photobiology A: Chemistry **2024**, 460, 116139. https://doi.org/10.1016/j.jphotochem.2024.116139.
35. Chiang, C.-Y.; Chen, M.-Y.; Hsu, C.-W.; Liu, C.-Y.; Tsai, Y.-W.; Liao, H.-C.; Yan, J.-Y.; Chuang, Z.-S.; Wang, Hsin-I.; Pan, C.-H.; Yu, C.-Y.; Yu, G.-Y.; Liao, C.-L.; Liu, S.-J.; Chen, H.-W. Induction of High Affinity Monoclonal Antibodies against SARS-CoV-2 Variant Infection Using a DNA Prime-Protein Boost Strategy. Journal of Biomedical Science **2022**, 29 (1). https://doi.org/10.1186/s12929-022-00823-0.
36. A systematic review by Xiao, K.; Tang, L.; Chen, Y.; Zhou, J.; Yang, Q.; Wang, R. The Effectiveness of E-Health Interventions Promoting Physical Activity in Cancer Survivors: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. Journal of cancer research and clinical oncology **2024**, 150 (2). https://doi.org/10.1007/s00432-023-05546-9.
37. Ladislav Batalik; Katerina Chamradova; Petr Winnige; Filip Dosbaba; Katerina Batalikova; Vlazna, D.; Janikova, A.; Pepera, G.; Hammoda Abu-Odah; Jing Jing Su. Effect of Exercise-Based Cancer Rehabilitation via Telehealth: A Systematic Review and Meta-Analysis. BMC Cancer **2024**, 24 (1). https://doi.org/10.1186/s12885-024-12348-w.
38. Ren, J.; Guo, H.; Pan, J.; Zhang, Y.-F.; Yang, Y.; Wu, X.; Du, S.; Ouyang, M.; Gao, H.-J. Interatomic Spin Coupling in Manganese Clusters Registered on Graphene. Physical Review Letters **2017**, 119 (17), 176806. https://doi.org/10.1103/PhysRevLett.119.176806.
39. Sohail, A. H.; Maan, M. H. A.; Sohail, S. Sex and the Streets: The Open Secret of Sexual Abuse among Pakistan's Two Million Street Children. Child and Adolescent Psychiatry and Mental Health **2021**, 15 (1). https://doi.org/10.1186/s13034-021-00420-3.
40. Yilmaz, A.; Ozkul, A.; Shin, D. S.; Im, S.-B.; Yoon, S.-M.; Kim, B.-T. Morphological Assessment of Cadaveric Radial, Brachial, and Subclavian Arteries: A Neurointerventional Approach. Journal of Korean Neurosurgical Society **2015**, 58 (6), 499–503. https://doi.org/10.3340/jkns.2015.58.6.499.
41. Bohîlțea, R. E.; Margareta Mihai, B.; Munteanu, O.; Ducu, I.; Adrian Dumitru, V.; Gheorghe, C.-M.; Augustin Georgescu, T.; Varlas, V.; Vlădăreanu, R. Early Prenatal Diagnosis of an Atypical Phenotype of Sacral Spina Bifida. Journal of Medicine and Life **2021**, 14 (5), 716–721. https://doi.org/10.25122/jml-2021-0292.
42. A systematic review by Wilson, O. W.; Matthews, C. E.; Wojcik, K. M.; Tarasenko, Y. N.; Butera, G.; Gorzelitz, J.; Schechter, C.; Sheng, J. Y.; Jayasekera, J. The Effects of Post-Diagnosis Rec-

DOI: 10.36838/IJHSR812.102

reational Aerobic Exercise among Breast Cancer Survivors: A Systematic Review/Meta-Analyses. Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology **2025**, 195 (1), 10.1158/1055-9965.EPI24–1798. https://doi.org/10.1158/1055-9965.EPI-24-1798.

43. An umbrella review by Bai, X.-L.; Li, Y.; Feng, Z.-F.; Cao, F.; Wang, D.-D.; Ma, J.; Yang, D.; Li, D.-R.; Fang, Q.; Wang, Y.; Jiang, X.-F.; Huang, D.-H.; Li, X.-Y.; Guo, J.-K.; Zhao, N.; Li, Z.-T.; Ma, Q.-P.; Wang, L.; Wu, Q.-J.; Gong, T.-T. Impact of Exercise on Health Outcomes in People with Cancer: An Umbrella Review of Systematic Reviews and Meta-Analyses of Randomized Controlled Trials. British Journal of Sports Medicine **2025**, bjsports-2024-109392. https://doi.org/10.1136/bjsports-2024-109392.
44. A systematic review by Toohey, K.; Chapman, M.; Rushby, A.-M.; Urban, K.; Ingham, G.; Singh, B. The Effects of Physical Exercise in the Palliative Care Phase for People with Advanced Cancer: A Systematic Review with Meta-Analysis. Journal of Cancer Survivorship **2022s**, 17 (2). https://doi.org/10.1007/s11764-021-01153-0.

■ Author

Tarun Sudhakar is a junior at Coppell High School who shows a particular commitment to biomedical research, particularly in oncology. He is interested in exercise physiology in relation to recovery and prognoses in cancer patients, quality of life improvement, and evidence-based patient-centered care.

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Understanding The Physical Properties of Neutron Stars and Black Holes Through Analysis of Gravitational Waves from Their Collisions

Eshani Lohia

RPG International School, 34 1/1, Diamond Harbor Rd, Kolkata, West Bengal, 700027, India; eshanilohia@gmail.com

ABSTRACT: Two celestial bodies, more than 100 times the mass of the sun, are hurtling towards each other, colliding to send ripples through the very fabric of reality. It is from this chaos that the grandest elements like gold and platinum form, or an even grander singularity is born, where our laws of physics cease to exist. And in short, this is how black holes, as well as neutron stars, collide. Due to recent advances in the field of gravitational wave technology and through observations by LIGO and Virgo, we have received immense new data, which makes it easier to analyze the collision of neutron stars and black holes using gravitational waves. Scientists often use machine learning data models to observe waveforms and determine physical parameters. This paper uses such detections to evaluate physical features about black holes and neutron stars, such as mass, angular momentum, tidal deformability, etc., by analyzing gravitational waves from their collisions. Understanding these features allows us to test theories like that of dense matter and even the limits of physics as we know it.

KEYWORDS: Physics, Astrophysics, Black Hole Collisions, Neutron Star Collisions, Gravitational Waves.

Introduction

When the universe's densest bodies—black holes and neutron stars—collide, they release more energy in mere seconds than our sun can in its entire lifetime. During this brief time, space and time ripples, new elements are born, and even larger black holes are formed as our universe is not just disturbed but reshaped.

These mergers or collisions can occur between binary black hole systems, binary neutron star systems, as well as black holes and neutron stars. These collisions send across gravitational waves— invisible ripples that travel through space—which help us detect physical components like mass, density, spins, tidal deformability, nuclear matter characteristics, etc., of these structures. This is one of the newest observations in astrophysics, with gravitational waves only being discovered in 2015,[1] although first predicted by Einstein in 19182 as a consequence of his general theory of relativity. GW150914 in 2015 was the first black hole collision detected by the Laser Interferometer Gravitational-Wave Observatory (LIGO).[3] This paper focuses on the physical properties of black holes and neutron stars that can be detected through analysis of gravitational waves from their collisions.

We need to understand the physical properties of these collisions to learn more about these events. Studying the properties like tidal deformability of neutron stars literally tells us about ultra-dense matter, while studying properties like spin distributions of black holes gives us a deeper understanding of how they were formed. Though there exist many papers about how gravitational waves have been used to detect specific properties of compact objects—talking about specific case studies or events—there is a lack of papers that specifically emphasize the physical nature of these collisions. This paper aims to do just that and address this lack by instead of focusing on observed data, showing how analyzing gravitational wave data can give us the physical properties like mass, spin, momentum, tidal deformability, etc., of binary black hole, binary neutron star, and black hole–neutron star collisions. It will review the common and unique physics of the events and refer to theories of general relativity,[2] dense matter, etc. Analyzing properties of black holes like mass ratios, spin magnitudes, etc., not only allows us to test the theory of general relativity but also tells us how black holes form and evolve.[3] Measuring radius, tidal deformability, etc., of neutron stars tells us about the equation of state, which is how matter behaves at nuclear and supranuclear densities. We also get information about the origin of heavy metals like gold and platinum through the process of nucleosynthesis. It is almost impossible for any experiment on Earth to recreate these processes, making studying these vital to not only astrophysics but also other basic sciences. Studying this helps us understand how matter is disrupted, and electromagnetic emissions like radiation and light are generated. It starts with a qualitative overview of general relativity (2.1). The next part (2.2) contains the breakdown of the different phases of a merger (inspiral, merger, and ringdown) and then focuses on physical properties derived from the gravitational wave data, like mass, spin, and tidal deformability (2.3). Important events like GW150914,3 GW170817,[4] and GW200115[5] are discussed along with their importance (2.4). It then talks about the astrophysical implications of these collisions, like equations of state, formation, and evolution of black holes (2.5). Finally,(2.6) it looks at future experiments and detectors that can help progress developments in this field and areas that should be focused on. The research used in this paper includes published data, such as results from detections, simulations, and

DOI: 10.36838/IJHSR812.110

prior analyses, to highlight the findings. This work is a review and does not present new gravitational-wave data analysis; it synthesizes existing observational results and theoretical models to highlight the physical interpretation of compact binary mergers.

Results and Discussion

2.1. The Theory of General Relativity:

The Theory of General Relativity by Albert Einstein was first published in 1916,[7] 11 years after his Special Theory of Relativity. A useful quote by John Archibald Wheeler can help sum it up—*Spacetime tells matter how to move; matter tells spacetime how to curve*. This can be explained using an analogy (Figure 1):

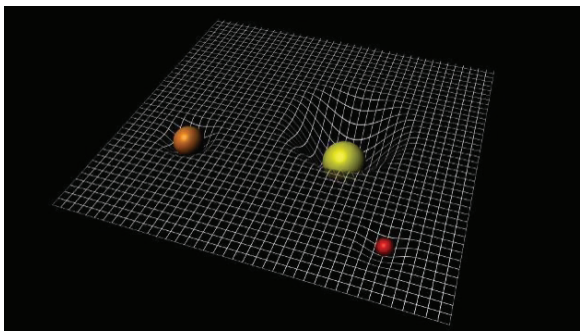


Figure 1:[6] Illustration of spacetime curvature caused by a massive object.

Imagine a stretched sheet of some rubber (spacetime) on which a ball (mass like a star) is kept. This obviously deforms the spacetime around it by "curving" it, and thus an object of a smaller mass (planet) when moving around it is not being "pulled" into it but following geodesics, which is the straightest possible path any object can take in a curved spacetime. The heavier the object, i.e., the more mass it has, the steeper the curves will be, and hence the geodesics will bend more sharply, causing smaller objects to be pulled into it.

Regarding this, Einstein's field equation states (Figure 2):

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$$

where Gμν is the curvature or Einstein tensor telling us how much spacetime is curved at a particular point in time, is the mass-energy or stress-energy tensor, G is Newton's gravitational constant (6.674×10^{-11} $N{\cdot}m^2/kg^2$), and c is the speed of light (3 * 108 m/s).

General Relativity also talks about spacetime and mass-energy in the context of gravitational waves. When two masses orbit each other, spacetime gets deformed as discussed earlier. When the speed of these masses accelerates, they send ripples through space and time in all directions away from the source, i.e., themselves. These ripples are known as gravitational waves.[7]

These waves propagate away from their source at the speed of light and carry information about their origin. They are not distortions traveling through spacetime, but distortions of spacetime. Collisions of black holes and neutron stars provide strong gravitational waves, which are the key components that are detected and analyzed to obtain information about their physical nature by observatories like LIGO (Laser Interferometer Gravitational-Wave Observatory) and Virgo.[9] In fact, though Einstein's Theory of General Relativity (1916) predicted the existence of gravitational waves,[7] it was not detected until 14 September 2015 by LIGO when two massive black holes collided—GW150914.[10]

The theory of conservation of energy makes it impossible to produce gravitational waves by shifting the total mass up and down (monopole) or moving the center of mass back and forth (dipole). Instead, a quadrupole change is what generates gravitational waves, which is when the shape of the mass distribution is changed (stretched or squeezed) over time, like in orbiting binary systems. This wave distortion or strain depends on how fast the quadrupole changes, and the power carried depends on the acceleration of this change. This concept was originally discovered by Einstein in 1918,[11] confirmed indirectly through binary pulsar energy loss,[12] and finally observed by LIGO.[10]

According to weak-field observation, when a source is very far away, the curvature of spacetime is less distinct, so a flat Minkowski spacetime can be considered with a ripple. This can be shown through the following equation (Figure 3):

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$$

where gμν is the spacetime metric, ημν is the Minkowski metric from the theory of Special Relativity, and hμν is the ripple that represents the gravitational wave. It helps provide an easier and simpler way to study gravitational waves without going into much detail about the math behind general relativity. This was derived by Einstein in his 1918 theory.

2.2. Phases of the Merger:

Merger:

A merger between stellar objects refers to the process by which these stellar bodies orbit each other at accelerating speeds and eventually collide due to the emission of gravitational waves by the bodies.[13] At this time, the spacetime curvature is greatest, and the gravitational wave signals have the most amplitude and frequency. These mergers between binary black hole systems, binary neutron star systems, and even black hole–neutron star systems will be studied. There are several phases to these mergers.

Inspiral Phase:

The first and longest phase of the merger is known as the inspiral phase. It is a very significant phase, as the gravitational wave signals from this phase are understood very well, helping scientists predict accurate information on the physical properties of these collisions.[14] In short, it is when the two bodies (neutron stars/black holes) are circling each other with increasing acceleration and are about to collide. This phase produces the gravitational wave "chirp" signal[15] which is a form of gravitational wave that continuously increases its frequency and amplitude over time. It occurs as the two objects orbit each other, they lose energy and angular momentum, radiating these waves. This produces a sort of feedback loop where the loss of energy and momentum causes the orbits to tighten

more, accelerating the motion of the objects and generating more and more energetic gravitational waves, i.e., chirps with increasing amplitude and frequency. This freq of these waves is always double that of the orbit. For stellar mass binaries like binary neutron star systems, this usually occurs in seconds before the merger with detectable frequencies, while in supermassive black hole systems, it can last for minutes but be much less detectable. For neutron star binaries, finite-size tidal effects become important during the late inspiral and introduce additional phase shifts that depend on the tidal deformability parameter Λ, which encodes information about the neutron star equation of state.

Merger Phase:

The merger phase is the most violent and dynamic part of a binary collision, lasting from milliseconds to a few seconds.

As the two objects reach the innermost stable circular orbit (ISCO), after losing their orbital energy due to emission of gravitational waves, they collide and merge into a larger and more compact object.[16] It is a rapid occurrence where both objects move at a huge fraction of the speed of light. The curvature of spacetime is the most extreme here, and any gravitational interactions are non-linear.

For binary black hole mergers, this system includes the deformation of the black hole event horizons and their merging into a single, larger, common event horizon.[17] This also creates a new black hole that rapidly spins away. For binary neutron star mergers, the presence of matter makes the process much more complicated. Matter is heated to high temperatures and densities, which can form a hypermassive neutron star for a limited amount of time before collapsing into a black hole. The projection of matter can also occur, forming a kilonova.[18]

The merger part gives us the most powerful gravitational wave signals. It includes rapid increases in frequency and amplitude, leading to a single peak.[16] At this peak, the signal lasts only for a fraction of a second but radiates a huge amount of energy. This marks the point of peak luminosity, which can produce astonishing amounts of power—for the first-ever detection, GW150914, the peak gravitational wave luminosity exceeded the combined electromagnetic output of all the stars in the observable universe.[10]

The waveform of the merger is more complex than that of the inspiral phase, as it contains a superposition of various frequencies. Thus, modeling the gravitational waveform requires complex numerical relativity simulations, unlike the simpler PN model(a method of approximation used in General Relativity to find the motion and radiation of gravitational waves in binary systems), approximations in the inspiral phase.[14]

A unique feature of this part of the merger is the "kick" or the gravitational recoil that happens in binary black hole mergers when the gravitational waves emitted are anisotropic and carry away linear momentum. In accordance with the principle of conservation of momentum, the newly formed black hole moves in the opposite direction.[19] Thus, the asymmetry in the emission of the gravitational waves is due to the misalignment between the spins of the black holes individually and the orbital angular momentum of the system, leading to precession of the orbital plane.[20]

This "kick" can cause recoil velocities of 5000 km/s, enough to expel a black hole from its original galaxy.[19] When it comes to the collision of two supermassive black holes, typically after the collision of their two galaxies, such "kicks" can cause the remnant to be removed from its galactic center and host galaxy, changing the evolution of both the black hole and the galaxy.[21] In neutron star mergers, a physical process called tidal disruption occurs. This is when a neutron star is completely torn apart by the tidal forces of its merger partner. This process then causes the ejection of matter, forming an accretion disk around the remnant after the collision. This process is extremely pertinent as, unlike in black hole mergers, these radiate electromagnetic material.[18] The material that gets ejected radiates across the electromagnetic spectrum, providing us with signals like a short burst of gamma rays or a kilonova. For example, in the GW170817 event where two binary neutron stars collided, the gravitational wave signal was followed by a short burst of gamma rays as well as a kilonova.[4]

Ringdown Phase:

The ringdown phase is the final phase of the merger. Though it is a brief stage of only fractions of a second to a few seconds, it is when the remnant of the collision or merger finally settles down into a stable stage.[16] After the dramatic merger phase, the remnant is newly formed, unstable, and distorted. During the ringdown phase, this remnant settles down into a stable equilibrium stage and releases a final burst of gravitational waves.

For a black hole, this final burst of gravitational waves includes exponentially damped oscillations called Quasinormal Modes. They are constantly damped due to continuous emission of gravitational radiation and are also known as the black hole's "fingerprint" because they provide a lot of information on the properties of the black hole.[22]

The remnants of neutron stars, however, are much more complex. The collision of neutron stars can lead to a new black hole and also hypermassive neutron stars that collapse after a short time. The gravitational wave signal contains information about this remnant finally settling down.[18] It is short-lived and chaotic, often releasing electromagnetic radiation.[4]

Thus, the three phases of the merger—inspiral, merger, and ringdown—are not completely different occurrences but all linked together to create a process that alters the shape of our universe. They have been crucial in receiving information regarding the physical properties of these collisions, which shall be discussed in detail later. They have also helped us discover many new truths and develop a deeper understanding of our universe. However, challenges remain.

For example, the impact of this "kick" or gravitational recoil in black hole mergers, especially when it comes to their galactic cores, is constantly being questioned.[19] However, with constant upgrades to the LIGO and Virgo observatories,[9] detectors, and key events like the GW170817 binary neutron star merger,[4] we are constantly receiving and decoding new information regarding these collisions. By combining these

DOI: 10.36838/IJHSR812.110

observations throughout the various spectrums, we will be able to paint a newer and more complete picture of this universe.

2.3. Physical Properties:

The detection, followed by the analysis of gravitational wave signals from these collisions, provides us with a lot of information on the physical properties of neutron stars and black holes themselves.

Quantities such as the chirp mass and the overall amplitude and frequency evolution of the signal are tightly constrained by the data itself; parameters including individual component masses, spin orientations, tidal deformability, and neutron-star radii require assumptions about general relativity, waveform models, and the equation of state of dense matter and theoretical modeling. This part of the paper defines how these physical properties are extracted for both the collisions commonly and specifically.

Mass and Mass Ratio:

Mass is the most important parameter we can extract from a gravitational signal from a merger. Since the individual masses cannot be directly determined from the signal, we instead extract a parameter called the chirp mass ($\mathcal{M}$), which is a specific combination of the masses of the two components that is mainly inferred from the gravitational waves in the inspiral phase.[23,24]

First, we need to understand the post-Newtonian (PN) expansion to understand the inference of chirp mass. The Post-Newtonian expansion is a simpler way to tell us about the orbital motion of the two bodies in terms of orbital velocity relative to the speed of light.[23] In simpler terms, it tells us about the increase in frequency of the gravitational wave over time, or gravitational wave chirp. In the lowest order, it can be defined with the following formula (Figure 4):

$$\frac{df}{dt} = \frac{96}{5}\pi^{8/3}\left(\frac{G\mathcal{M}}{c^3}\right)^{5/3} f^{11/3}.$$

Where f — **Gravitational-wave frequency**, equal to twice the orbital frequency of the binary in the inspiral phase; df/dt is the **Frequency evolution (chirp rate)**, describing how rapidly the gravitational- wave frequency increases as the binary loses energy through gravitational radiation. G is **Newton's gravitational constant**, governing the strength of gravity; c is the **Speed of light in vacuum**, setting the relativistic scale of the system; $\mathcal{M}$ is the chirp mass, and π is pi.

Since it is directly proportional to the frequency evolution, it is connected to the rate at which the frequency of the gravitational wave chirp increases. Since it is directly related to the waveform observed by scientists, we can infer the chirp mass with relative accuracy by measuring the derivative of frequency and time of the gravitational wave signal.[24,25] Chirp mass can also be defined using the formula (Figure 5):

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}.$$

where m1 and m2 are the masses of the two components, however, we cannot always determine the individual masses from this equation, as different combinations or ratios of m_1 and m_2 give us the chirp mass. Thus, to get the individual masses and mass ratio, we must use PN formulas or corrections of a much higher order and complex nature.[23] These can include other physical components as well, like spin, higher-order velocity, etc. Since these other parameters cannot be as accurately measured from the waveform as the chirp mass, the degree of accuracy of the masses also decreases.

Another requirement to consider while measuring mass is the cosmological redshift(z). Since gravitational waves are stretched while they travel across the constantly expanding universe, the chirp mass is not the natural chirp mass but the chirp mass that has been influenced by the cosmological redshift.[25] Their relationship can be described by-observed $\mathcal{M}$ = Intrinsic $\mathcal{M}(1 + z)$

Thus, the gravitational waveform is not the only thing that must be considered while determining $\mathcal{M}$, as we also need to consider the redshift of the system. This can be done by multimessenger astronomy, where an electromagnetic complement can help us determine the cosmological redshift. For example, in the GW170817 event, the gravitational-wave signal and the electromagnetic signals from the kilonova due to the neutron star merger gave us an accurate redshift measurement.[26,27] On combining the Gravitational wave estimate of distance and electromagnetic wave estimate of distance, the deterioration between the mass and cosmological redshift was removed. Thus, both the combined electromagnetic wave signals and gravitational wave signals can give us extremely accurate information on the masses of the components of the systems.

Spin of Components:

The spin of the two components involved in the collision heavily influences the shape of the gravitational waveform, helping us decode plenty of information.[28] Spin is a vector quantity, so the magnitude and direction in relation to the orbit impact the radiation emitted and thus the information we infer regarding the collision.

One of the most fundamental parameters to consider here is the effective spin- mass-weighted average of the two components of the spin. If the effective spin is positive, this means that spin is aligned with the orbital angular momentum, and the "hang-up" effect is observed. During this effect, the orbital angular momentum increases, slowing down the inspiral phase of the merger. The gravitational wave signals sent to the LIGO and Virgo detectors last much longer and are thus much easier to detect.[28] If the effective spin is negative, the spin is anti-aligned with the orbital angular momentum; the opposite effect can be observed. During this effect, the orbital angular momentum decreases, increasing the speed of the inspiral phase of the merger. This shortens the gravitational wave signals sent to the LIGO and Virgo detectors, making them harder to detect.[29]

However, by measuring the effective spin across many collisions, scientists can understand the evolution of these mergers and test predictions on different merger models.

Spins that are aligned or anti-aligned with the orbital angular momentum are known as aligned spins. There also exists

a different type of spin caused by precessing spins, where the spins are tilted to some extent with respect to the orbital angular momentum.[30]

In this case, the theory of general relativity states that the orbital plane precesses around the total angular momentum.[28]

Precession is the movement of a spinning body around another such body due to a torque that changes the direction of the first body's axis, like how a spinning top will wobble under gravity due to precession. This can be detected in gravitational wave signals as there are modulations periodically in the amplitude and the phase of the waveform due to a change in the location of the precessing merger in relation to the Earth.[30] This not only gives us information that the spins are not aligned but also removes any degeneracies that exist while interpreting the waveform.

During isolated binary evolutions, the objects form a binary system earlier and end up evolving with each other. Thus, various processes, including tidal locking and stellar winds, align their spins with the orbital angular momentum. They show aligned spins with an effective spin nearly equal to zero or positive in nature.[31]

In dynamic binary evolutions, the objects are formed separately and brought together much later. This can be seen in dense stellar environments like clusters. The pairing of objects in this case is very random, and isotropic spin orientations are seen where there is no specific alignment.[32] There can be a mix of anti- aligned, misaligned, and slightly aligned spins of the mergers. The GW190521 event, which had strong evidence of precession, had a dynamic binary evolution,[33] but the GW170817 event of neutron star merger had relatively small spin effects and an isolated binary evolution.[26]

Radiated Energy and Angular Momentum:

Since the merger of black holes and neutron stars is one of the most energetic events, we already know that a small part of the masses of the components is transformed into gravitational-wave energy.[28] The amplitude and time span of the signal give us the radiated energy and angular momentum of the system, as those parameters tell us how much energy is converted to gravitational wave signals.[23] This radiated energy causes the collision by decreasing the distance between the objects and the overall energy of the system.

Black hole mergers particularly radiate a lot of gravitational wave energy, with the LIGO and Virgo detectors showing that these mergers have radiated energy equivalent to 1–3 solar masses in the fraction of a second.[26] This is literally more light energy than all the stars in the observable universe can emit at the same time. The efficiency of this radiation depends directly on the natural or intrinsic properties of the bodies, like mass, spin, etc. Binaries that have components with equal mass and aligned spins to the orbital angular momentum produce the biggest fraction for mass-gravitational wave energy conversion and have the longest inspiral phase.[28] The remnant thus contains the remaining mass and angular momentum of the system. These can also lead to strong gravitational kicks, already discussed earlier, which have a velocity of thousands of kilometers/second, ejecting black holes completely from their host galaxies.[30]

Final Remnants:

The collision finally leads to the formation of a remnant (a new, larger object that contains the remaining mass and angular momentum of the system).

For black holes, this remnant is usually another larger black hole as described in the Kerr Metric of the Theory of General Relativity.[28] The mass and spin of the remnants are not fixed at the beginning but established much later during merger and ringdown phases as large amounts of energy and angular momentum are radiated away from the system.[34]

As the gravitational wave signals conclude during the ringdown stage, spacetime vibrates around the remnant. This vibration can be expressed in the form of damped oscillations called quasinormal modes (QNMs).[34] Each mode has its own oscillation frequency and damping timeline depending on the remnant's spin and mass. Thus, through these quasinormal modes, the mass and spin of the remnant can directly be measured without the need for information on the previous inspiral phase.[34] This also helps us test the No-Hair theorem, which states that a stationary black hole can be understood just by its mass and its angular momentum.[34] One of the main tests for this is the Inspiral-Merger-Ringdown Consistency Test, which compares the remnant mass and spin predicted in the inspiral and merger phase to the remnant mass and spin of the ringdown phase.[34] If these estimates are constant, the No-Hair theorem from The Theory of General Relativity is true; any significant discrepancy would lead to a new physics beyond The Theory of General Relativity.

The remnants of neutron stars, however, are much more complex. The collision of neutron stars can lead to a new black hole or an HMNS (hyper-massive neutron star), which collapses after a short time.[37] The gravitational wave signal contains information about this remnant finally settling down. It is short-lived and chaotic, often releasing electromagnetic radiation.[38]

Demonstrative Example:

During the inspiral phase of a compact binary merger, the gravitational-wave signal is dominated by the gradual increase in the chirp mass defined above. The rate at which the gravitational-wave frequency increases is directly proportional to the chirp mass and frequency of the signal, so through measuring the frequency and its time derivative directly from the observed waveform, the chirp mass can be determined with high precision, largely independent of other parameters such as spin. For example, in the first gravitational-wave detection, GW150914, the observed chirp mass was approximately 30 $M_{\odot}$, indicating a binary system composed of two stellar-mass black holes significantly heavier than those previously observed through electromagnetic methods. While the chirp mass is tightly constrained, the individual component masses are less precisely determined because they enter the waveform at higher post-Newtonian orders and are partially degenerate with spin effects. In neutron star mergers, additional finite-

DOI: 10.36838/IJHSR812.110

size effects appear during the late inspiral. These tidal effects introduce small but measurable deviations from the point-particle waveform, allowing parameters such as tidal deformability to be constrained alongside the chirp mass. Thus, gravitational-wave observations enable both precise mass measurements and, in the case of neutron stars, direct probes of internal structure.

Properties specif ic to Neutron Stars:

Tidal Deformability (Λ) and Radii:

Tidal deformability is one of the main factors that separates neutron star mergers from black hole mergers.[35] In binary neutron star systems, each star deforms due to the tidal gravitational field of its merger partner.[35] This deformation, known as quadrupole deformation, is small compared to the overall size of the star but has a measurable impact on the gravitational-wave signal, introducing an additional phase shift. These tidal effects become especially important during the final few orbits before merger, when the orbital separation is smallest and gravitational tidal forces are strongest.[36] The degree of tidal deformation is described by the dimensionless tidal deformability parameter, Λ, which characterizes how susceptible a neutron star is to tidal distortion based on its mass and internal structure.[35] Neutron stars described by a stiffer equation of state have higher internal pressure, resulting in larger stellar radii and larger values of Λ, which produce stronger tidal signatures in the gravitational- wave waveform. In contrast, softer equations of state correspond to more compact neutron stars with smaller radii and smaller Λ, leading to weaker tidal effects.[36] In this way, the tidal deformability parameter Λ provides direct information about the properties of ultra-dense nuclear matter.[36] landmark detection of the GW170817 event provided the first observational constraints on tidal deformability.[27] Analysis of the inspiral phase placed an upper limit on Λ, ruling out extremely stiff theoretical models that predicted neutron stars with very large radii and strong tidal distortions.[36]

Since constraints on the equation of state determine the relationship between neutron star mass and radius, measurements of Λ can also be used to infer neutron-star radii.[35] For the GW170817 event, gravitational-wave observations indicated that a neutron star with a mass of approximately 1.4 M⊙ has a radius in the narrow range of about 11.8–13.7 km.[27,36] These findings not only constrain neutron-star radii and equations of state but also provide insight into whether exotic forms of matter, such as deconfined quarks, may exist in neutron-star cores, as well as the maximum mass these stars can support before collapsing into a black hole.[37] Tidal deformability is unique to neutron stars because it arises from their finite size and internal structure. Black holes, by contrast, have no material surface or deformable interior and are fully described by vacuum solutions to Einstein's equations, meaning no tidal deformability effects occur in black hole mergers.[35]

The Final Remnant and Signal

The remnant of a binary neutron star merger depends on the mass of the total system and the nuclear equation of state.[37] It is not common or universal to all mergers. Various possible things could happen- Black Hole- If the mass exceeds the constraint or limit determined by the Equation of State, around $\gtrsim$ 2.8–3.0 M⊙, the remnant immediately collapses to form a black hole with almost no stage in the middle and a very short ringdown gravitational wave signal.[38]

Hypermassive Neutron Star- If the mass is slightly lower compared to the above case, around 2.6–2. M⊙, a hot neutron star is formed, which rotates in different directions. Its mass exceeds the limit on a uniformly rotating star, so though it can be stabilized temporarily by rapid rotation and thermal support, it almost immediately(within milliseconds to seconds) collapses to form a black hole due to cooling down and reduction of angular momentum.[37]

Supramassive Neutron Star: If the mass is slightly lower compared to the above case, around 2.3–2.6 M⊙, a rotating neutron star is formed. Since its mass is above that of the maximum non-rotating limit but less than the differential rotation limit, it survives for longer (it can survive for a few hours). Eventually, it collapses to form a black hole due to the reduction of angular momentum.[38] Stable Neutron Star: If the mass is relatively much lower, typically $\lesssim$ 2.2 M⊙, a new stable neutron star is formed that has a long lifetime before it collapses.[37] When the remnant forms a Hypermassive Neutron Star or Supramassive Neutron Star, a shorter but higher-frequency (2–4 kHz range) gravitational-wave signal is generated.[38] However, unlike the inspiral signal, which is dependent on tidal deformability, the post-merger signal is dependent on the supranuclear-density of the EoS.[37] These post-merger signals are relatively much weaker, so there are certain limits on their detections as seen during the absence of a clear post-merger signal in GW170817.[27] This is trying to be overcome by observatories like Einstein Telescope and Cosmic Explorer, which are currently under development and are sensitive to kHz.[38]

To conclude and summarize, the following table tells us about the physical properties of these compact mergers (Table 1):

Physical Properties Summary:

Table 1:

Property	Waveform	Phase
Chirp Mass	Frequency (df/dt)	Inspiral
Mass Ratio/ Individual Mass	Higher order PN terms	Inspiral/Merger
Aligned Spin	Duration	All
Misaligned Spin	Precession	All
Final Mass	QNM frequency	Merger/ Ringdown
Final Spin	QNM decay time	Ringdown
Radiated Energy	Amplitude/ Duration	Merger/ Ringdown
Tidal Deformability	Phase Shift	Inspiral

2.4. Key Events:

The field of gravitational wave astronomy evolved from Einstein's Theory to reality by the first LIGO detection in 2015.[39] Since then, the LIGO and VIRGO observatories have recorded black hole mergers, neutron star-neutron star mergers, and even black hole-neutron star mergers, which gave insight and

information on the physical properties of these objects.[40] The following key events- GW150914, have not only validated already existing theories but challenged them, introducing new concepts like multi-messenger astronomy.[39–43]

These detections have only been made possible by many observatories like the LIGO Laser Interferometer Gravitational-wave Observatory and Virgo observatory, which can detect minute changes between two mirrors spread across vast distances of several kilometers.[44] The LIGO in Hanford, Washington, and Livingston, Louisiana, has two 4 km L-shaped interferometers where split laser beams measure distortions as small as 10^{-19} m.[44] The Virgo near Pisa, Italy, with 3km arms, works similarly. They began joint observations in 2017, which improved sky localization, helping with follow-up observations.

GW150914- The first detection of gravitational waves by LIGO on 14th September, 2015, when two black holes of 29 and 36 M⊙ merged into a 62 M⊙ remnant, radiating 3 M⊙c^2 in 0.2s.[39] Before this, we only knew of black holes as massive as 20 M⊙, but this showed us that black holes of 29 and 36 M⊙ can exist and merge, releasing ~3 M⊙ of energy in a fraction of a second—more than all the stars in the observable universe at that moment. It confirmed Einstein's Theory of General Relativity, matched the Kerr metric, and proved that interferometers like LIGO can detect waves invisible to telescopes, revolutionizing the field of gravitational wave astronomy.[39]

GW170817- The first neutron star merger on August 17, 2017, by LIGO and Virgo, where both gravitational waves and electromagnetic waves were detected, giving rise to the concept of multi- messenger astronomy.[41,42] In this event, a 100-second signal was followed by gamma rays, kilonova, X-rays, and radio waves, which confirmed the r-process nucleosynthesis in neutron star mergers, constrained neutron star radii (~10–13 km), and showed the effects of tidal deformability.[41,42]

GW190521- The detection of 85 and 66 M⊙ black holes merging to form a 142 M⊙ remnant on May 21, 2019, by LIGO and Virgo, which was the first detected intermediate-mass black hole43 and violated the "mass gap" from stellar evolution theory and confirmed theories of hierarchical mergers where black holes can combine to form larger ones.[43]

GW200105 & GW200115- The detection of neutron star and black hole collisions by the LIGO and VIRGO collaborations on January 5, 2020, and January 15, 2020, respectively.[40] GW200105 had a 9 M⊙ black hole and 1.9 M⊙ neutron star forming an 8.9 M⊙ black hole remnant, while GW200115 had a 6 M⊙ black hole and 1.5 M⊙ neutron star forming a 6.4 M⊙ black hole remnant. This confirmed neutron star–black hole mergers as well as dark mergers due to almost no electromagnetic radiation detected, and expanded our knowledge on these mergers.[40]

The above-mentioned 5 key events revolutionized this field in astrophysics forever by detecting gravitational waves, three merger types (BH–BH, NS–NS, NS–BH), and introducing the concept of multi-messenger astronomy.[39–43] They constrained the equation of state of ultra-dense matter and confirmed the r-process of nucleosynthesis as well as intermediate-mass black holes.39–43 New upcoming detectors and improved sensitivity of the current detectors shall help us detect more gravitational waves and possibly discover a new type of signal.[44]

2.5. Astrophysical Implications:

Detection of Gravitational-Waves and Multi-Messenger Astronomy:

The field of gravitational wave astronomy has become a new way to explore the universe, as these waves have not only confirmed black hole-black hole collisions, neutron star-neutron star collisions, and black hole- neutron star collisions, but also allowed us to test existing theories and gain information on the physical properties of these stellar objects and their mergers.[26,27] Certain events like GW170817 combined this gravitational wave emission with electromagnetic emission, which gave rise to the concept of Multi-Messenger Astronomy and helped us get a better understanding of physical parameters like chirp mass, distance, and the location of the collision.[28,29]

The Equation of State and Tidal Deformability:

Neutron stars are made of ultra-dense nuclear matter, so their structure depends on the nuclear equation of state (EoS), which describes how matter behaves under different conditions.[30] One of the main observable properties related to the EoS is tidal deformability, which tells us the degree of distortion of a star due to its partner's gravitational field.[31] Although this quadrupole deformation may be small, it causes a measurable phase shift in the inspiral gravitational-wave signal, especially in the final orbits before the merger, where the gravitational tidal forces are strongest.[32]

A stiffer EoS produces neutron stars with larger radii and stronger pressure support, resulting in stronger tidal effects, whereas a softer EoS produces smaller, more compact stars with weaker tidal deformation signatures.[30] This information also tells us about the size and stability of neutron stars.

Observations from GW170817 show that a 1.4 M⊙ "canonical" neutron star has a radius of around 11.9 ±1.4 km, ruling out extremely stiff EoSs that predict oversized stars as well as overly soft EoSs that are inconsistent with the existence of 2 M⊙ pulsars.[28,33]

Because neutron stars reach densities several times greater than that of an atomic nucleus, conditions unachievable on Earth, they act as natural laboratories for nuclear physics.[34] Constraints from GW170817 placed limits on the pressure of matter at approximately twice nuclear saturation density, helping refine theoretical models and rule out unrealistic EoSs.[28,35] The EoS also determines whether the merger remnant collapses immediately into a black hole or forms a long-lived neutron star, as well as the amount of matter ejected, linking it to nucleosynthesis and short gamma-ray bursts.[28,36] A remnant with a softer EoS has less pressure support and is more likely to collapse promptly into a black hole, whereas a stiffer EoS allows higher maximum masses and can produce a hypermassive or supramassive neutron star that survives for milliseconds to seconds before collapse, or in some cases remains stable. A stiffer EoS can also lead to greater matter ejection through tidal interactions, which is important for r-process nucleosynthesis.[28]

DOI: 10.36838/IJHSR812.110

R-Process Nucleosynthesis:

Neutron star mergers tell us of the origin of heavy elements like gold, platinum, and even uranium through the rapid neutron-capture process (the r-process).[37] Nucleosynthesis is the process by which atomic nuclei are formed through existing particles like protons, neutrons, etc.[38] In GW170817, neutron-rich debris ejected at ~10% the speed of light underwent nucleosynthesis, with radioactive decay powering the kilonova afterglow.[28,39] The kilonova color change from red to blue indicated many ejecta components which produced lighter nuclei ($Z < 140$) and heavier nuclei, including lanthanides and actinides.[28,39] Spectroscopic observation confirmed strontium, the first element traced to cosmic nucleosynthesis.[40] Though this confirms that neutron star-neutron star collisions are dominant sources of r-process nucleosynthesis, questions remain whether other cosmic phenomena like collapsars also contribute.[41]

Stellar Formation:

Binary black hole mergers form through isolated binary evolution and dynamical assembly in dense stellar clusters.[42]

The isolated binary evolution of two stellar objects evolves together and eventually collapses to form a black hole, though uncertainties exist, like "common envelope," where the outer envelope of one object should expand and engulf its companion, mass transfer, and supernova kicks that can unbind binaries.[42,43] The dynamic channel in denser environments like globular clusters and nuclear star clusters brings black holes together, often leading to hierarchical mergers where remnants can merge again, possibly producing intermediate-mass black holes.[44] However, parameters like merger rate evolution with redshift, eccentricity, and most importantly, spin alignment are required to gain more information.[45]

Mass Gap:

Stellar evolution theory predicts a mass gap of ~50–120 $M_\odot$ where black holes cannot form due to the pair-instability supernova where the core of massive stars is hot enough to form electron–positron pairs, reducing internal radiation pressure that counteracts the impact of gravity, triggering the runaway contraction and explosive burning that destroys the star, leaving no remnant which explains the predicted black hole "mass gap" of 50–120 $M_\odot$.[46,47] This theory was disputed by the GW190521 event, where the masses of the colliding black holes were 66 $M_\odot$ and 85 $M_\odot$ approx, and forced scientists to reconsider.[48] Alternate theories like dense stellar environments, where repeated mergers push remnants into the gap, and alternative stellar pathways, where low metallicity and rotation alter mass loss and allow direct collapse into heavier black holes, have been proposed.[49] Detecting black holes in this mass gap has broader implications for stellar evolution. Population III stars could have formed black holes and distinguished between isolated stellar evolution and dynamical assembly ways.[50]

Neutron Star-Black Hole Collisions:

The fate of a neutron star-black hole merger can be determined depending on whether the neutron star crosses the black hole's event horizon before it is deformed by tidal forces.[51] If disruption happens before the ISCO is reached by the powering kilonovae, gamma-ray bursts, and r-process nucleosynthesis.[52] Otherwise, the neutron star directly moves into the black hole with limited matter ejected. This causes a "dark merger" that can only be detected by gravitational waves.[51]

This depends on physical properties like mass ratio, spin of black holes, and compactness of neutron stars.[53] High spins and lower mass ratios shift the innermost stable orbit inward, giving tidal forces time to disrupt the neutron star and produce ejecta and light.[54] High mass ratio and low spin cause plunging of the neutron into the black hole with no electromagnetic signals. GW200105 and GW200115 confirmed these "dark mergers" with minimal electromagnetic radiation.[55] Thus, the absence of detection of light from a merger reveals information like the mass, spin, as well as constraints on the EoS of ultra-dense matter.[51] Thus, gravitational astronomy has helped us revolutionize our understanding of astrophysics, constrain the EoS of dense matter, uncover nucleosynthesis sites, and reshape stellar evolution theory.[26–55]

2.6. Future:

After the first detection of gravitational waves in 2015, a new window was opened to observe the universe. The current observatories, like LIGO and Virgo, have made revolutionary discoveries that not only confirmed the Theory of General

Relativity, but also provided unprecedented insights into the properties and collisions of black holes and neutron stars.[56,57]

As science progresses, this initial phase of discovery is now transitioning into an era of more precise detections and quantitative measurements in gravitational-wave astronomy.[58]

Next-Generation Observatories:

The new era will be ushered in by third-generation ground-based detectors already in advanced planning and early construction stages. These instruments are designed with a sensitivity an order of magnitude greater than Advanced LIGO.[59]

This enhancement has profound consequences:

- Range: Because the amplitude of gravitational waves decreases inversely with distance, a tenfold sensitivity increase corresponds to a tenfold extension in range.[60]
- Volume: Since observable volume scales with the cube of distance, detection volume grows by a factor of a thousand.[61]
- Events: The detection rates could increase from dozens per year to millions.[62] This massive dataset will shift the focus from individual case studies to statistical population analyses of compact binaries.[63]

Key scientific goals like mapping the mass and spin distributions of compact binaries,[64] inferring merger rates as a function of redshift,[65] and studying binary formation channels and evolutionary pathways.[66]

The two main third-generation observatories are:

- Cosmic Explorer (CE): Two L-shaped surface interferometers with 40 km and 20 km arms are planned. The 40 km instrument will provide broadband sensitivity, while the 20 km interferometer will excel at detecting high-frequency signals from the ringdown phase.[67]

- Einstein Telescope (ET): A subterranean triangular design, with three 10 km arms forming three V-shaped Michelson interferometers. Its "xylophone" configuration consists of interferometers optimized separately for low and high frequencies, providing sensitivity across the full spectrum.[68] ET's low-frequency capability extends inspiral phase tracking and enables detection of intermediate-mass black hole binaries.[69]

Understanding Supranuclear Matter:

Binary neutron star inspirals generate gravitational waves encoding the cold nuclear Equation of State (EOS), while post-merger ringdown phases probe the hot nuclear EOS.[70] Current detectors primarily access inspiral signals, but next-generation sensitivity will allow detection of faint, high-frequency post-merger waves.[71]

Key insights from these detections include constraints on temperature-dependent EOS in ultra- dense matter,[72] probing phase transitions in neutron star matter,[73] and establishing connections between gravitational-wave signals and heavy-ion collision experiments on Earth.[74]

Challenges include noise contamination of faint, short-lived post-merger signals. Advanced signal processing and novel instrument configurations are required to overcome glitches and non-Gaussian noise.[75]

Black Hole Spectroscopy:

A cornerstone of General Relativity is the No-Hair Theorem, which posits that black holes are fully characterized by mass and spin.[76] Gravitational-wave spectroscopy of ringdown emissions provides the means to rigorously test this theorem. Older detectors could only observe the dominant quasinormal mode (QNM).[77] Next-generation detectors will resolve multiple subdominant QNMs and even nonlinear quadratic QNMs.[78] Independent recovery of black hole parameters from multiple modes allows consistency checks of the No-Hair Theorem.[79]

Multi-Messenger and Multi-Disciplinary Future:

Third- generation detectors are also expected to significantly advance multi-messenger astronomy, coordinating gravitational-wave detection with electromagnetic and neutrino observatories.[80] Such synergies will enable precise localization of neutron star collisions and deeper insights into r-process nucleosynthesis.[81]

The nuclear EOS faces discrepancies between astronomical observations and terrestrial nuclear physics experiments. Future detectors will help bridge this divide as gravitational-wave constraints on neutron star structure will inform and refine nuclear theory and EOS modeling,[82] these models can be tested against experimental data from facilities like Jefferson Lab and FRIB400,[83] and machine learning approaches will iteratively improve EOS predictions through a feedback loop between theory, astrophysical measurements, and laboratory data.[84] This integration of astrophysical measurements, nuclear physics experiments, and computational modeling is poised to revolutionize our understanding of matter under extreme conditions, advancing gravitational-wave science into a quantitatively predictive discipline.[85]

Conclusion

To conclude, gravitational wave astronomy has completely changed the way we study the universe by transforming a mere theory in Einstein's Theory of General Relativity into reality by the 2015 LIGO detection.[86] This new beginning in observational astronomy has helped us detect the collisions of black holes and neutron stars and thus infer physical information from them.[87] Thus, this paper aimed to describe the physical properties of neutron stars and black holes that can be detected through the analysis of gravitational waves from their collisions.

The foundations of this are rooted in the Theory of General Relativity, which describes gravity as the curvature of spacetime rather than a conventional force.[88] The detection of gravitational waves confirmed this with a high degree of accuracy and showed that the three phases of a merger- inspiral, merger, and ringdown align with the theory.[89]

Each phase carries its own signals, which carry unique information on not only the merger but also the physical properties of the objects involved in the merger.[90] The inspiral phase tells us about physical parameters like mass and spin with high precision, the merger phase releases huge amounts of energy in the form of gravitational wave emission, and the ringdown phase gives us information on the remnant formed and its stability, providing the most stringent tests of Einstein's theory.[91]

A huge amount of information on the physical properties, such as mass and spin of colliding objects and their remnants, can also be inferred through gravitational waves from the collisions.[92] Since gravitational waves are directly produced from compact objects, we can get accurate information on parameters like masses, spins, radii, and even internal structure, unlike in electromagnetic astronomy, which depended on light interacting with matter.[93] Some events like GW170817 tell us about tidal deformability and the equation of state of nuclear matter, giving us information on physics at densities unattainable on Earth.[94] There have been many key detections that have shaped this field which include GW150914- first detection of gravitational waves from a black hole-black hole collision,[95] GW170817- first detection of neutron star-neutron star collision which introduced the concept of multi messenger astronomy,[96] GW190521- detection of black holes in the "mass gap" which challenged the stellar evolution theory,[97] GW200105 and GW200115- detections of black hole-neutron star collisions.[98]

This also has far-reaching astrophysical implications.[99] Binary neutron star mergers give us a way to connect gravitational physics with nuclear physics and chemistry, giving us a deeper insight into the r-process of nucleosynthesis, nuclear Equation of State, etc.[100] Binary black hole mergers tell us about stellar evolution, binary formation channels, and dynamical processes in dense clusters.[101] Neutron star–black hole collisions shed light on the concept of "dark mergers" without any electromagnetic radiation produced.[102]

DOI: 10.36838/IJHSR812.110

The future of gravitational wave astronomy is also very promising.[103] It includes both updates to current detectors and development of third-generation observatories such as the Einstein Telescope and Cosmic Explorer, which will help improve sensitivity to fainter and more distant gravitational wave signals, possibly helping us detect millions of mergers per year.[104] This also helps in the development of multi- messenger astronomy, which will give us better measurements of properties of compact objects, their evolution, and destruction.[105]

Thus, gravitational-wave astronomy is changing the way we see the universe, but it is only in its beginning stages.[106] There are many persisting questions, like the true equation of state of matter in neutron stars, and how much nuclear matter is ejected into the universe during nucleosynthesis. Though current models have provided plenty of information on this field, many uncertainties, like the ones mentioned, remain. However, since future detectors will be able to detect fainter and a broader range of frequencies, the mysteries can be resolved, and a gap between theory, observation, and experimentation can be closed. Around a decade ago, the signals were a mere theory, but now they are frequently detected, observed, and decoded to gain information and test the laws of physics.[107] As the detection technology advances, we are at the threshold of discoveries that change our understanding of matter, gravity, and tell us the story of how the cosmos has been made.

Acknowledgments

This research paper could only be completed due to the support and guidance of many individuals and institutions. I want to extend my deepest gratitude to my mentors, Dr. Chima McGruder and Ms. Catherine Peretti. Their advice and insights helped me shape this paper and develop a much deeper love for gravitational wave astronomy and astrophysics as a whole. I am also very thankful to the various studies and observations conducted, especially by the LIGO and Virgo observatories. Their groundbreaking work formed the foundation for this paper, and without it, such a deep analysis would not have been possible. I also want to thank the IRIS Intensive Research Program for giving me such a wonderful opportunity to write my own research paper and guiding me every step of the way. "I attest that the ideas, graphics, and writing in this paper are entirely my own. Finally, I would like to express my thanks to my family and friends, as without their constant support and patience, this paper would not have been completed. This review research paper would not have been completed without the collective support of everyone mentioned.

References

1. Abbott, B. P.; Abbott, R.; ***Abbott, T. D.; Abernathy, M. R.***; *et al. Observation of Gravitational Waves from a Binary Black Hole Merger*. ***Phys. Rev. Lett***. 2016, 116 (6), 061102.
2. Einstein, A. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* 1918, 1, 154–167.
3. Abbott, B. P.; Abbott, R.; Abbott, T. D.; ***Abernathy, M. R.***; *et al. GW150914: The Advanced LIGO Detectors in the Era of First Discoveries*. ***Phys. Rev. Lett***. 2016, 116 (13), 131103.
4. Abbott, B. P.; Abbott, R.; Abbott, T. D.; ***Abernathy, M. R.***; *et al. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral.* ***Phys. Rev. Lett***. 2017, 119 (16), 161101.
5. Abbott, R.; Abbott, T. D.; Abraham, S.; ***Acernese, F.***; *et al. Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences.* ***Astrophys. J. Lett.*** 2021, 915 (1), L5.
6. European Space Agency. *Spacetime Curvature*. ESA–C. Carreau, 2015.
7. Einstein, A. *Die Grundlage der allgemeinen Relativitätstheorie.* ***Annalen der Physik*** 1916, 49 (7), 769–822.
8. Wheeler, J. A. *A Journey into Gravity and Spacetime*; W. H. Freeman: New York, 1990.
9. Acernese, F.; Agathos, M.; Agatsuma, K.; ***Aisa, D.***; *et al. Advanced Virgo: A Second-Generation Interferometric Gravitational Wave Detector. Class.* ***Quantum Gravity*** 2015, 32 (2), 024001.
10. Abbott, B. P.; Abbott, R.; Abbott, T. D.; ***Abernathy, M. R.***; *et al. Observation of Gravitational Waves from a Binary Black Hole Merger.* ***Phys. Rev. Lett***. 2016, 116 (6), 061102.
11. Einstein, A. *Über Gravitationswellen*. ***Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*** 1918, 1, 154–167.
12. Hulse, R. A.; Taylor, J. H. *Discovery of a Pulsar in a Binary System*. ***Astrophys. J. Lett.*** 1975, 195, L51–L53.
13. Blanchet, L. *Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries*. ***Living Rev. Relativ***. 2014, 17, 2.
14. Buonanno, A.; Damour, T. *Effective One-Body Approach to General Relativistic Two-Body Dynamics*. ***Phys. Rev. D*** 1999, 59, 084006.
15. Peters, P. C.; Mathews, J. *Gravitational Radiation from Point Masses in a Keplerian Orbit*. ***Phys. Rev***. 1963, 131, 435–440.
16. Pretorius, F. *Evolution of Binary Black-Hole Spacetimes*. ***Phys. Rev. Lett***. 2005, 95, 121101.
17. Thornburg, J. *Event and Apparent Horizon Finders for 3+1 Numerical Relativity*. ***Living Rev. Relativ***. 2007, 10, 3.
18. Shibata, M.; Hotokezaka, K. *Merger and Mass Ejection of Neutron Star Binaries*. ***Annu. Rev. Nucl. Part. Sci***. 2019, 69, 41–64.
19. Campanelli, M.; Lousto, C. O.; Zlochower, Y.; Merritt, D. *Maximum Gravitational Recoil*. ***Astrophys. J. Lett***. 2007, 659, L5–L8.
20. Lousto, C. O.; Zlochower, Y. *Spin-Induced Orbital Precession and the Modulation of Gravitational Waveforms in Binary Black Hole Mergers*. ***Phys. Rev. D*** 2013, 87, 084027.
21. Merritt, D.; Milosavljević, M.; Favata, M.; Hughes, S. A.; Holz, D. E. *Consequences of Gravitational Radiation Recoil.* ***Astrophys. J. Lett.*** 2004, 607, L9–L12.
22. Berti, E.; Cardoso, V.; Will, C. M. *Gravitational-Wave Spectroscopy of Massive Black Holes with the Space Interferometer LISA*. ***Phys. Rev. D*** 2006, 73, 064030.
23. Blanchet, L. *Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries*. ***Living Rev. Relativ.*** 2014, 17, 2.
24. Cutler, C.; Flanagan, É. E. *Gravitational Waves from Merging Compact Binaries: How Accurately Can One Extract the Binary's Parameters from the Inspiral Waveform?* ***Phys. Rev. D*** 1994, 49, 2658–2697.
25. Sathyaprakash, B. S.; Schutz, B. F. *Physics, Astrophysics and Cosmology with Gravitational Waves*. ***Living Rev. Relativ***. 2009, 12, 2.
26. Abbott, B. P.; Abbott, R.; Abbott, T. D.; Abernathy, M. R.; *et al. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral*. ***Phys. Rev. Lett***. 2017, 119, 161101.
27. Abbott, B. P.; Abbott, R.; Abbott, T. D.; Abernathy, M. R.; *et al. Multi-messenger Observations of a Binary Neutron Star Merger.* ***Astrophys. J. Lett.*** 2017, 848, L12.
28. Ajith, P.; Babak, S.; Chen, Y.; Hewitson, M.; *et al. Phenomenological Template Family for Black-Hole Coalescence Waveforms*. ***Class. Quantum Gravity*** 2007, 24, S689–S700.

29. Kidder, L. E. *Coalescing Binary Systems of Compact Objects to (Post)5/2-Newtonian Order. V. Spin Effects*. ***Phys. Rev. D*** 1995, 52, 821–847.
30. Apostolatos, T. A.; Cutler, C.; Sussman, G. J.; Thorne, K. S. *Spin-Induced Orbital Precession and Its Modulation of the Gravitational Waveforms from Merging Binaries*. ***Phys. Rev. D*** 1994, 49, 6274–6297
31. Kalogera, V. *Spin-Orbit Misalignment in Close Binaries with Two Compact Objects*. ***Astrophys. J.*** 2000, 541, 319–328.
32. Rodriguez, C. L.; Chatterjee, S.; Rasio, F. A. *Binary Black Hole Mergers from Globular Clusters: Implications for Advanced LIGO*. ***Phys. Rev. D*** 2016, 93, 084029.
33. Abbott, R.; Abbott, T. D.; Abraham, S.; Acernese, F.; *et al.* GW190521: A Binary Black Hole Merger with a Total Mass of 150 M⊙. ***Phys. Rev. Lett.*** 2020, 125, 101102.
34. Berti, E.; Cardoso, V.; Starinets, A. O. *Quasinormal Modes of Black Holes and Black Branes. Class.* ***Quantum Gravity*** 2009, 26, 163001.
35. Hinderer, T. *Tidal Love Numbers of Neutron Stars*. ***Astrophys. J.*** 2008, 677, 1216–1220.
36. Abbott, B. P.; Abbott, R.; Abbott, T. D.; Abernathy, M. R.; *et al. GW170817: Measurements of Neutron Star Radii and Equation of State*. ***Phys. Rev. Lett***. 2018, 121, 161101.
37. Baiotti, L.; Rezzolla, L. Binary Neutron Star Mergers: A Review of Einstein's Richest Laboratory. Rep. Prog. Phys. 2017, 80, 096901.
38. Paschalidis, V. General Relativistic Simulations of Compact Binary Mergers as Engines for Short Gamma-Ray Bursts. Class. Quantum Gravity 2017, 34, 084002
39. Abbott, B. P.; Abbott, R.; Abbott, T. D.; Abernathy, M. R.; *et al.* Observation of Gravitational Waves from a Binary Black Hole Merger. Phys. Rev. Lett. 2016, 116, 061102.
40. Abbott, R.; Abbott, T. D.; Abraham, S.; Acernese, F.; *et al.* Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences. Astrophys. J. Lett. 2021, 915, L5.
41. Abbott, B. P.; Abbott, R.; Abbott, T. D.; Abernathy, M. R.; *et al.* GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. Phys. Rev. Lett. 2017, 119, 161101.
42. Abbott, B. P.; Abbott, R.; Abbott, T. D.; Abernathy, M. R.; *et al.* Multi-messenger Observations of a Binary Neutron Star Merger. Astrophys. J. Lett. 2017, 848, L12.
43. Abbott, R.; Abbott, T. D.; Abraham, S.; Acernese, F.; *et al.* GW190521: A Binary Black Hole Merger with a Total Mass of 150 M⊙. Phys. Rev. Lett. 2020, 125, 101102.
44. Acernese, F.; Agathos, M.; Agatsuma, K.; Aisa, D.; *et al.* Advanced Virgo: A Second-Generation Interferometric Gravitational Wave Detector. Class. Quantum Gravity 2015, 32, 024001.
45. Abbott, R.; Abbott, T. D.; Abraham, S.; Acernese, F.; Ackley, K.; Adams, C.; *et al.* Population of Merging Compact Binaries Inferred Using Gravitational Waves Through GWTC-3. Phys. Rev. X 2023, 13, 011048.
46. Fowler, W. A.; Hoyle, F. Neutrino Processes and Pair Formation in Massive Stars and Supernovae. Astrophys. J. Suppl. Ser. 1964, 9, 201–319.
47. Heger, A.; Woosley, S. E. The Nucleosynthetic Signature of Population III. Astrophys. J. 2002, 567, 532–543.
48. Abbott, R.; Abbott, T. D.; Abraham, S.; Acernese, F.; *et al.* (LIGO Scientific Collaboration and Virgo Collaboration). GW190521: A Binary Black Hole Merger with a Total Mass of 150 M⊙. Phys. Rev. Lett. 2020, 125, 101102.
49. Mapelli, M. Formation Channels of Merging Stellar Black Holes. Front. Astron. Space Sci. 2021, 8, 697277.
50. Inayoshi, K.; Visbal, E.; Haiman, Z. The Assembly of the First Massive Black Holes. Annu. Rev. Astron. Astrophys. 2020, 58, 27–97.
51. Shibata, M.; Taniguchi, K. Coalescence of Black Hole–Neutron Star Binaries. Living Rev. Relativ. 2011, 14, 6.
52. Metzger, B. D. Kilonovae. Living Rev. Relativ. 2019, 23, 1.
53. Foucart, F. Black Hole–Neutron Star Mergers: Disk Mass Predictions. Phys. Rev. D 2012, 86, 124007.
54. Kyutoku, K.; Ioka, K.; Shibata, M. Ultra-Relativistic Counterparts to Binary Black Hole Mergers. Mon. Not. R. Astron. Soc. 2013, 437, L6–L10.
55. Abbott, R.; Abbott, T. D.; Abraham, S.; Acernese, F.; *et al.* (LIGO Scientific Collaboration and Virgo Collaboration). Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences. Astrophys. J. Lett. 2021, 915, L5.
56. B. P. Abbott; R. Abbott; T. D. Abbott; M. R. Abernathy; *et al.*, Phys. Rev. Lett., 2016, 116, 061102.
57. B. P. Abbott; R. Abbott; T. D. Abbott; M. R. Abernathy; *et al.*, Phys. Rev. X, 2016, 6, 041015.
58. M. Maggiore; C. Van Den Broeck; N. Bartolo; *et al.*, J. Cosmol. Astropart. Phys., 2020, 03, 050.
59. P. R. Saulson, Living Rev. Relativity, 2017, 20, 1.
60. J. Aasi; B. P. Abbott; R. Abbott; T. D. Abbott; *et al.*, Class. Quantum Grav., 2015, 32, 074001.
61. H. Audley; S. Babak; J. Baker; *et al.*, Class. Quantum Grav., 2017, 34, 044001.
62. C. Cutler and K. S. Thorne, in General Relativity and Gravitation, Cambridge University Press, 2002.
63. S. Vitale, Gen. Relativ. Gravit., 2014, 46, 1730.
64. W. M. Farr; S. Stevenson; M. C. Miller; *et al.*, Nature, 2011, 548, 426.
65. T. Regimbau, Res. Astron. Astrophys., 2011, 11, 369.
66. K. Belczynski; D. E. Holz; T. Bulik; R. O'Shaughnessy; *et al.*, Astrophys. J., 2016, 819, 108.
67. M. Evans; R. X. Adhikari; C. Afle; *et al.*, 2021 Astrophys. J., 908, L10.
68. S. Hild; M. Abernathy; F. Acernese; *et al.*, Class. Quantum Grav., 2011, 28, 094013.
69. M. Punturo; M. Abernathy; F. Acernese; *et al.*, Class. Quantum Grav., 2010, 27, 194002.
70. A. Bauswein and H. T. Janka, Phys. Rev. Lett., 2012, 108, 011101.
71. J. A. Clark; A. Bauswein; L. Cadonati; *et al.*, Class. Quantum Grav., 2016, 33, 085003.
72. A. Passamonti; N. Andersson; W. C. G. Ho; *et al.*, Mon. Not. R. Astron. Soc., 2013, 429, 767.
73. L. Rezzolla and K. Takami, Phys. Rev. D, 2016, 93, 124051.
74. B. P. Abbott; R. Abbott; T. D. Abbott; M. R. Abernathy; *et al.*, Astrophys. J., 2017, 850, L39.
75. M. Isi; M. Giesler; W. M. Farr; M. A. Scheel; *et al.*, Phys. Rev. D, 2019, 100, 042003.
76. R. Ruffini and J. A. Wheeler, Phys. Today, 1971, 24, 30.
77. E. Berti; V. Cardoso; A. O. Starinets; *et al.*, Class. Quantum Grav., 2009, 26, 163001.
78. V. Cardoso and P. Pani, Living Rev. Relativity, 2019, 22, 4.
79. M. Giesler; M. Isi; M. A. Scheel; S. A. Teukolsky; *et al.*, Phys. Rev. X, 2019, 9, 041060.
80. M. Branchesi, J. Phys. Conf. Ser., 2016, 718, 022004.
81. B. P. Abbott; R. Abbott; T. D. Abbott; M. R. Abernathy; *et al.*, Astrophys. J. Lett., 2017, 848, L12.
82. A. Schwenk and J. Piekarewicz, Phys. Today, 2021, 74, 34.
83. R. Stroberg; S. R. Stroberg; H. Hergert; J. D. Holt; *et al.*, Ann. Rev. Nucl. Part. Sci., 2021, 71, 307.
84. A. Raithel and F. Özel, Phys. Rev. C, 2019, 99, 045801.
85. S. Vitale and H.-Y. Chen, Phys. Rev. Lett., 2018, 121, 021303.

DOI: 10.36838/IJHSR812.110

86. Abbott, B. P.; Abbott, R.; Abbott, T. D.; Abernathy, M. R.; *et al.* Observation of Gravitational Waves from a Binary Black Hole Merger. Phys. Rev. Lett. 2016, 116, 061102.
87. Acernese, F.; Agathos, M.; Agatsuma, K.; Aisa, D.; *et al.* Advanced Virgo: A Second-Generation Interferometric Gravitational Wave Detector. Class. Quantum Grav. 2015, 32, 024001.
88. Einstein, A. The Foundation of the General Theory of Relativity. Ann. Phys. 1916, 49, 769–822.
89. Blanchet, L. Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries. Living Rev. Relativ. 2014, 17, 2.
90. Maggiore, M. Gravitational Waves: Volume 1: Theory and Experiments; Oxford University Press: Oxford, 2007.
91. Berti, E.; Cardoso, V.; Will, C. M. Gravitational-Wave Spectroscopy of Massive Black Holes with the Space Interferometer LISA. Phys. Rev. D 2006, 73, 064030.
92. Abbott, B. P.; Abbott, R.; Abbott, T. D.; Abernathy, M. R.; *et al.* Properties of the Binary Black Hole Merger GW150914. Phys. Rev. Lett. 2016, 116, 241102.
93. Sathyaprakash, B. S.; Schutz, B. F. Physics, Astrophysics and Cosmology with Gravitational Waves. Living Rev. Relativ. 2009, 12, 2.
94. Abbott, B. P.; Abbott, R.; Abbott, T. D.; Abernathy, M. R.; *et al.* GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. Phys. Rev. Lett. 2017, 119, 161101.
95. Abbott, B. P.; Abbott, R.; Abbott, T. D.; Abernathy, M. R.; *et al.* Tests of General Relativity with GW150914. Phys. Rev. Lett. 2016, 116, 221101.
96. Abbott, B. P.; Abbott, R.; Abbott, T. D.; Abernathy, M. R.; *et al.* Multi-messenger Observations of a Binary Neutron Star Merger. Astrophys. J. Lett. 2017, 848, L12.
97. Abbott, R.; Abbott, T. D.; Abraham, S.; Acernese, F.; *et al.* GW190521: A Binary Black Hole Merger with a Total Mass of 150 M⊙. Phys. Rev. Lett. 2020, 125, 101102.
98. Abbott, R.; Abbott, T. D.; Abraham, S.; Acernese, F.; *et al.* Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences. Astrophys. J. Lett. 2021, 915, L5.
99. Kalogera, V.; Belczynski, K.; Kim, C.; O'Shaughnessy, R.; Willems, B. Formation of Double Compact Objects. Phys. Rep. 2007, 442, 75–108.
100. Lattimer, J. M.; Prakash, M. The Physics of Neutron Stars. Science 2004, 304, 536–542.
101. Rodriguez, C. L.; Chatterjee, S.; Rasio, F. A. Binary Black Hole Mergers from Globular Clusters: Implications for Advanced LIGO. Phys. Rev. D 2016, 93, 084029.
102. Foucart, F. Black Hole–Neutron Star Mergers: Theoretical Status. Front. Astron. Space Sci. 2020, 7, 46.
103. Abbott, R.; Abbott, T. D.; Abraham, S.; Acernese, F.; *et al.* Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO, Advanced Virgo, and KAGRA. Living Rev. Relativ. 2020, 23, 3.
104. Reitze, D.; Adhikari, R. X.; Ballmer, S.; Barsotti, L.; *et al.* Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO. Bull. Am. Astron. Soc. 2019, 51, 35.
105. Punturo, M.; Abernathy, M.; Acernese, F.; Astone, P.; *et al.* The Einstein Telescope: A Third-Generation Gravitational Wave Observatory. Class. Quantum Grav. 2010, 27, 194002.
106. Barack, L.; Cardoso, V.; Nissanke, S.; Sotiriou, T. P.; *et al.* Black Holes, Gravitational Waves and Fundamental Physics: A Roadmap. Class. Quantum Grav. 2019, 36, 143001.
107. Schutz, B. F. Gravitational Wave Astronomy. Class. Quantum Grav. 2011, 28, 125023.
108. Centrella, J.; Baker, J. G.; Kelly, B. J.; van Meter, J. Black-Hole Mergers and Gravitational Waves. Rev. Mod. Phys. 2010, 82, 3069–3119.

Author

Eshani Lohia, a high school student from West Bengal, India, is deeply passionate about horse riding, football, Formula 1, and reading. Above all, she loves astrophysics and believes her life's purpose is to uncover the mysteries of the universe and make a meaningful impact in physics.

IJHSR International Journal of High School Research

is a publication of

terra science and education

N.Y. based 501.c.3 non-profit organization
dedicated for improving K-16 education