IJHSR

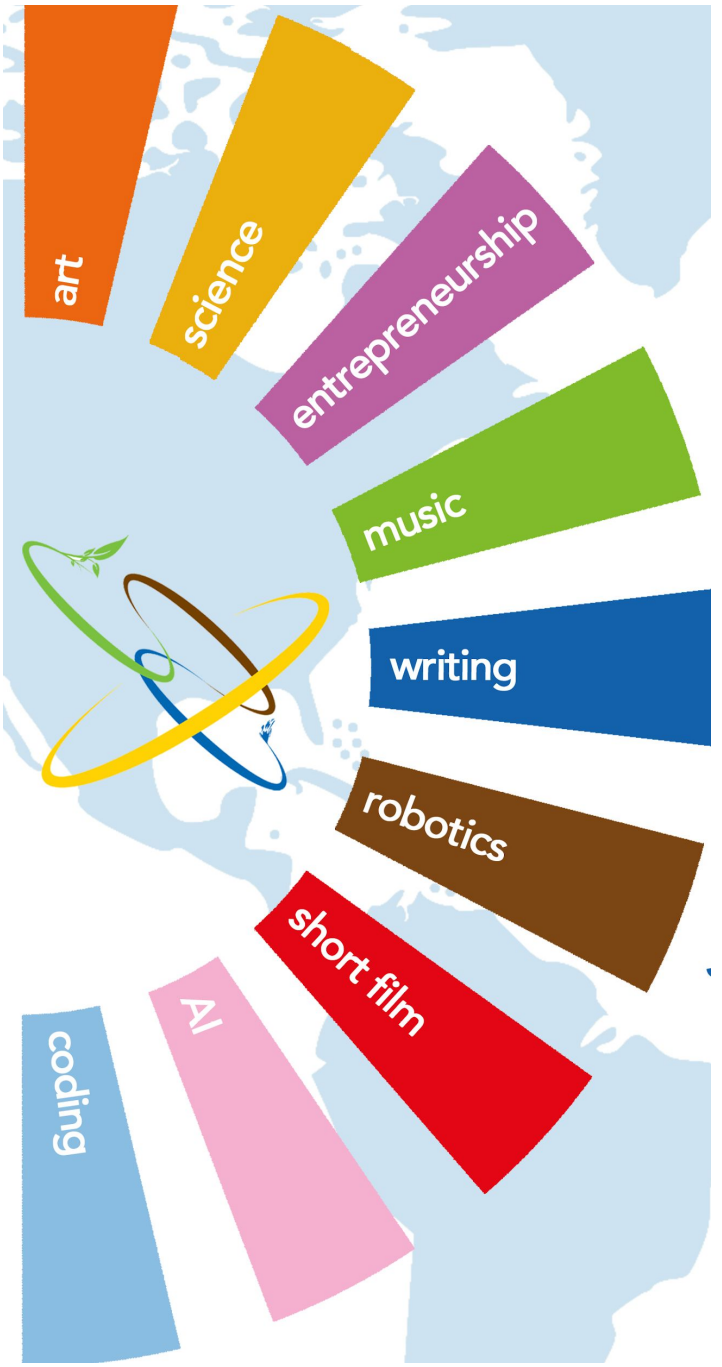
International Journal of High School Research

February 2026 | Volume 8 | Issue 3
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

February 2026| Volume 8| Issue 3

Editorial Board

International Journal of High School Research

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Class 100 Filtration Mechanism for Household-level Remediation to Eliminate Groundwater Contaminants

Adhit Mandal

The Shri Ram School, Aravali, Gurgaon, Haryana, India, 122002; adhit.mandal29@gmail.com

ABSTRACT: This research paper focuses on providing potable drinking water for low-income households, where treated tap water from municipal sources is unavailable. It leverages multi-stage, ultra-filtration for a range of physical, chemical, and microbiological contaminants in groundwater, resulting from manmade activities like industrial pollution, untreated sewage water, fecal sludge, waste landfills, mining, grey water from homes not connected to the drainage system, and natural reasons like the presence of minerals, resulting from the dissolution of rocks. The water does not meet quality standards, with high levels of suspended particles, fluoride, chloride, arsenic, nitrates, sulfates, phosphates, metals (iron, manganese, lead, mercury, cadmium, chromium), organic contaminants (pesticides, herbicides, oil, hydrocarbons), and microbes. Using easily available, affordable materials and leveraging a Class 100 filtration mechanism, contaminated water samples were passed through separate multi-stage ultrafiltration columns with layers of coarse sand grains, slag wool, pink sand, activated charcoal, and ceramic cones. When water flowed slowly through the filtration column, the physical, chemical, and microbiological contaminants were removed. This mechanism is low-maintenance, does not require electricity, and the media need to be cleaned after 30-40 days, with replacement of only the activated charcoal, thus providing a practical approach to address multiple contaminants in drinking water.

KEYWORDS: Chemistry, Materials Chemistry, Class 100 Water Filtration, Water Contamination, Water Purification.

Introduction

The sustainability of drinking water sources has become a growing challenge, particularly in developing nations, due to increasing population, inadequate water management, and the inevitable crisis of climate change. According to estimates by the World Health Organization (WHO) and the United Nations Children's Fund (UNICEF) in 2021, approximately 2 billion people worldwide, or a quarter of the global population, lacked access to clean water.[1] As per another estimate in 2024, 4.4 billion people across 135 low- and middle-income countries — over half of the world's population — do not have safe household drinking water, with fecal contamination as the primary limiting factor affecting them.[2] In a large country like India, about 30% of urban households, mostly those living in slums, and 90% of rural households still depend entirely on untreated surface water or groundwater.[3]

Groundwater quality has deteriorated due to various anthropogenic activities, including industrial pollution, sewage, waste landfills, and mining. Additionally, there are naturally occurring or geogenic reasons for water contamination, resulting from the presence of natural minerals that are caused by the dissolution of soluble rock products. Groundwater may contain several hazardous contaminants and does not meet the standards specified by international organizations, such as the WHO,[4] or the country-specific guidelines published by regulatory agencies. These standards cover a range of parameters, including physical and chemical metrics, microbiological parameters, and other aspects of water quality.

The kind of contaminants and the degree of contamination varies across countries and regions within a country, with quality issues like elevated levels of fluoride, arsenic, iron, manganese, chlorides, nitrates, sulfates, phosphates, heavy metals (lead, mercury, cadmium, chromium), organic contaminants (pesticides, herbicides, oil, hydrocarbons), suspended particles, microbes, and hardness minerals like calcium and magnesium.

In several countries, especially in rural areas, village ponds play a crucial role in maintaining the quality of groundwater, which is often drawn from hand pumps and borewells. The pond water becomes contaminated because households and community water points in many places are not connected to the drainage network. Hence, grey water (from bathing, washing clothes, and utensils) stagnates outside houses or at water points. Greywater contamination is also an issue in urban slums. Fecal sludge, biosolids, and untreated wastewater from nearby industrial activity further deplete the water quality in ponds. Most often, the users are not even aware of the presence of the contaminants and their side effects.

Prolonged consumption of contaminated drinking water has several repercussions. High levels of dissolved solids can impact health by causing gastrointestinal issues, kidney problems, and even affecting the taste and odor of the water, making it less palatable and sometimes leading to dehydration. High fluoride levels can cause fluorosis, which is typically diagnosed at a more advanced stage and is irreversible. Dental fluorosis causes loss of luster and shine of the dental enamel. Skeletal fluorosis leads to severe pain associated with rigidity and restricted movements of the cervical and lumbar spines, knees, pelvis, and shoulder joints, often leading to crippling deformity. Arsenic causes skin diseases and cancer. Chlorides combined with sodium cause high blood pressure. Nitrates in water cause

DOI: 10.36838/v8i3.1

thyroid and birth defects like methemoglobinemia or blue baby syndrome, where the ability of red blood cells to carry oxygen is reduced and can lead to serious illness or death. Sulfate, when combined with magnesium, may cause gastrointestinal issues. Consumption of heavy metals in drinking water can cause neurological problems, kidney damage, and cancer. The pathogens cause waterborne diseases like cholera, typhoid, and dysentery.[5-9]

The goal of this research is to identify suitable natural and sustainable filtration media for creating a household-level water treatment system, especially in low-resource settings, such as rural areas and urban slums, to deliver water quality comparable to that of households with a municipal treated water supply.

Methods

Section 1: Contaminants in water samples:

Water samples, drawn from hand pumps and being used for drinking purposes, were collected from villages and urban slums in India. These were collected from areas where water issues exist, such as high dissolved solids and foul odors, community water points not linked to drainage systems, stagnant and dirty water close to the water points, and industrial activity nearby that releases untreated wastewater into the groundwater.

I took four different samples of water from urban and rural sources, tested the samples for all characteristics (physical, chemical, metal, microbiological) through an accredited lab, and compared them to the specifications published by the Bureau of Indian Standards (BIS) IS: 10500.[10] The first water sample was taken from the municipal water in urban areas, which is treated water. In ideal situations, all other samples of water should meet at least these standards. The quality parameters of this sample were compared to those of water samples from different sources, which are expected to have high contaminants due to factors such as natural reasons, proximity to stagnant, dirty water, including grey water, fecal sludge, and untreated water from industrial activity. The samples chosen were as follows:

- Sample 1 (S1) – Municipal water supply to urban households
- Sample 2 (S2) – Hand pump water from a village where stagnant grey water and sludge are present close by, because the main drainage system is not adequately connected to all houses
- Sample 3 (S3) – Hand pump water in an urban slum where stagnant grey water is present close by, caused by a lack of connection of the community water point with the main drainage system
- Sample 4 (S4) – Hand pump water in a village where untreated water from industrial activity has seeped into the groundwater. The industrial activity in the vicinity comprises metal and alloy factories, dyeing units, plastic manufacturing facilities, chemical processing plants, and electronics and battery manufacturing plants. In several cases, smaller units occasionally do not adhere entirely to wastewater treatment standards and thus contribute to groundwater contamination.

Section 2: Results of testing of water samples before filtration:

The characteristics that did not meet BIS standards in any one of the samples are being discussed further in this report. The characteristics that met the BIS standards have not been discussed further. The test results for S1, municipal water supply to urban households, met all the specifications as per BIS standards for the parameters tested. S2, S3, and S4, however, had different kinds of contaminants, some at alarmingly high levels.

Physical characteristics: The test results for S2, collected from a hand pump in a village with stagnant grey water and sludge in the vicinity, indicated a foul smell. The odor in the other two contaminated samples was agreeable. The total dissolved solids (TDS) were higher than the BIS standard of 500 maximum in all three contaminated samples, as showcased in Figure 1. The levels of TDS were 3256 mg/L, 985 mg/L, and 1590 mg/L in S2, S3, and S4, respectively, as against the prescribed standard of a maximum of 500 mg/L.

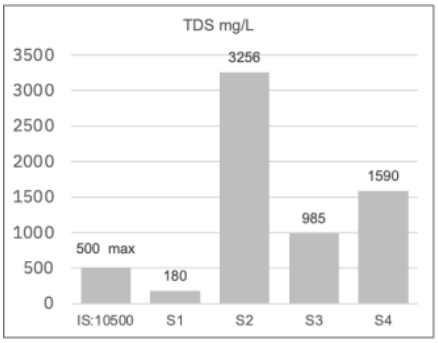
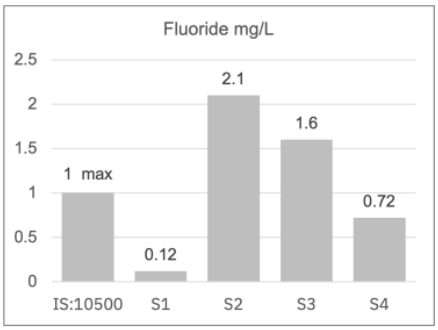
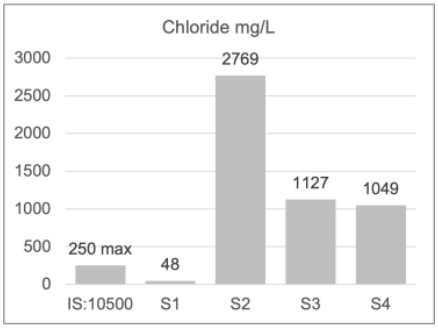


Figure 1: TDS levels in drinking water samples vis-à-vis BIS standards. TDS in S2 was more than 6 times, in S3 almost twice, and in S4 more than thrice the industry standard (maximum 500 mg/L) prescribed, while the municipal water sample S1 adhered to the standard.

Chemical analysis: According to the laboratory test results on the water samples, the key contaminants in S2, S3, and S4 were fluoride, chloride, nitrate, and sulfate.





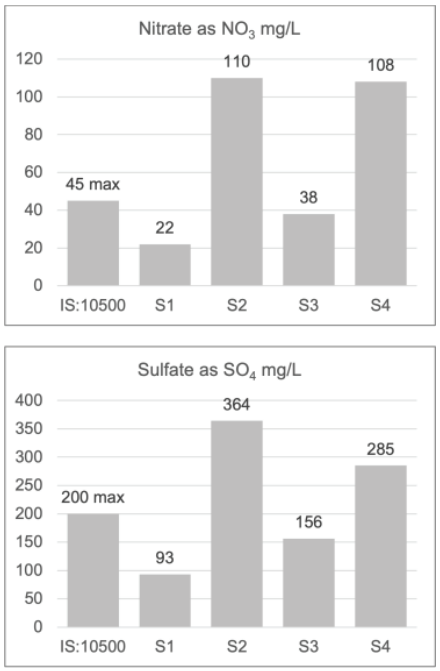


Figure 2: Results of chemical (fluoride, chloride, nitrate, and sulfate) analysis of drinking water samples. While S1 from the municipal source met all parameters for chemical analysis, S2 was contaminated with all four chemicals, S3 was contaminated with fluoride and chloride, and S4 was contaminated with chloride, nitrate, and sulfate.

As evident from Figure 2, the level of contamination of S2 was the highest amongst all samples – fluoride at double the level (2.1 mg/L), chloride at more than ten times (2769 mg/L), nitrate at more than double (110 mg/L), and sulfate at 1.8 times (364 mg/L) the maximum levels prescribed as per the industry standards. Major chemical contaminants in S3 were fluoride (1.6 mg/L) and chloride (1127 mg/L), with the latter at more than four times the industry standards. S4 had almost four times the prescribed limit of chloride (1049 mg/L), more than twice the limit of nitrate (108 mg/L), and high levels of sulfate (285 mg/L).

Metal analysis: The results of the presence of metals are showcased in Figure 3. The test results showed that iron and lead levels in all samples were way above industry standards. Iron levels were 1.1 mg/L, 0.9 mg/L, and 1.5 mg/L in S2, S3, and S4, respectively, as compared to the prescribed standard of a maximum of 0.3 mg/L. The levels of lead were 0.05 mg/L, 0.03 mg/L, and 0.05 mg/L in S2, S3, and S4, respectively, in comparison to the prescribed standard of a maximum of 0.01 mg/L.

Moreover, alarming levels of cadmium in S2 were found, 3.09 mg/L as compared to the industry standard of 0.01 mg/L. Chromium and uranium levels were unusually high in S3, chromium levels at 4.1 mg/L, with the maximum limit prescribed at 0.05 mg/L, and uranium at 4.1 mg/L, with the maximum limit prescribed at 0.03 mg/L. This was mainly an outcome of the untreated water from industrial activity seeping into the groundwater.

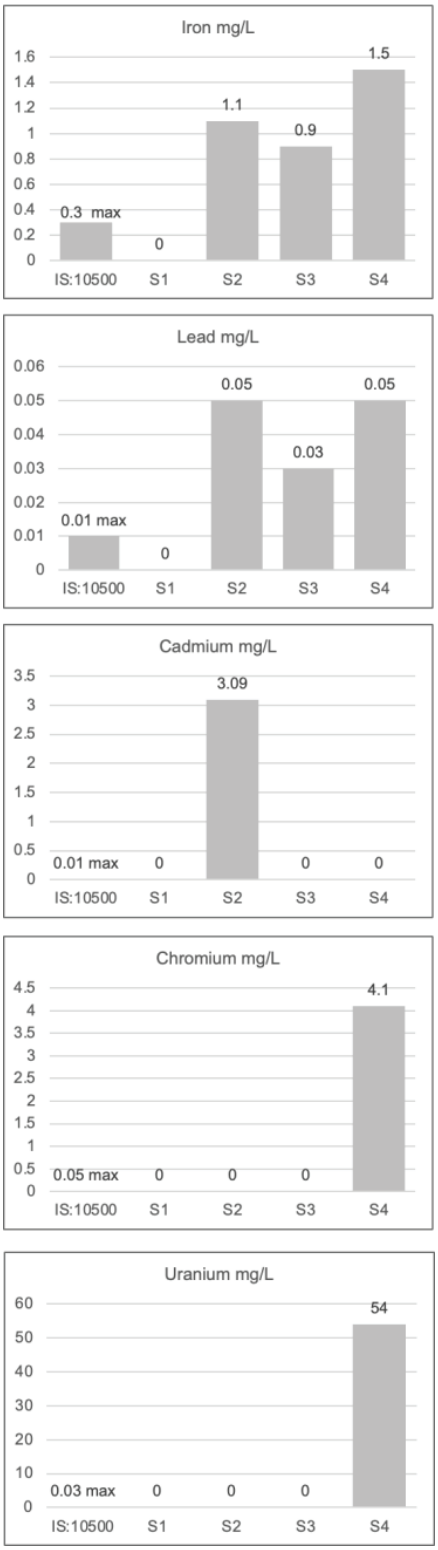


Figure 3: Results of metal analysis of drinking water samples. Iron and lead levels were higher than acceptable norms in all three contaminated samples. Cadmium level in S2 was 300 times the permissible limit. S4 was found to have abnormally high levels of chromium and uranium contamination.

Microbiological analysis revealed coliform contamination in S3, indicating that the water was not suitable for drinking purposes.

DOI: 10.36838/v8i3.1

Section 3: Filtration mechanism for addressing multiple contaminants:

To address the issue of varying contaminants across locations, my solution is based on Class 100 water filtration, a multi-stage ultrafiltration system for water, utilizing principles of physical, chemical, and microbiological processes. "Class 100" implies a system designed to minimize particulates to a very high degree and remove microscopic contaminants.

Using natural materials for Class 100 water filtration requires carefully selecting materials that are most effective in removing physical and chemical contaminants, as well as bacteria, at a micro scale.

Figure 4: Materials used in the filtration column for water purification. Layers of coarse sand, slag wool, pink sand, activated charcoal, and ceramic cones, available in bulk locally in all environments, were chosen for further experimentation.

Five layers of different materials were selected to create a filtration column, with a focus on adsorption. The layers included coarse sand grains, slag wool, pink sand, activated carbon, and ceramic cones, as exhibited in Figure 4. The ultrafiltration mechanism works through physical filtration, where pollutants are trapped in the filtering media as water passes through, and absorption filtration, where contaminants are adsorbed into the filtration media. The adsorbing media were selected based on their high thermal stability, small pore diameters, high exposed surface area, and hence high surface capacity for adsorption.

The filtration column for a household of 4-5 members can be designed to be approximately 2 feet tall. Each layer is approximately 4 inches thick, ensuring a water flow rate of approximately 1.5 to 2 liters per hour. I experimented with various thicknesses of each layer. I found that a thickness of approximately 4 inches is suitable for slowing down the water flow and providing ideal contact time with the filtration media. The raw contaminated water is fed into the filtration column from the top and moves downward through the layers due to the force of gravity. The treated, purified water is collected from the outlet at the bottom of the filtration system.

Section 4: Filtration Column layers (in descending order):

Layer 1 - Coarse sand grains (approximate grain size of 1mm)

These are available in bulk locally in all environments. For this project, river sand was used, which has a high surface area, surface charges (especially from clay and oxide coatings), the presence of functional groups for chemical binding, and pore spaces that trap microbes and particulates. 70-90% of river sand is composed of crystalline silica (SiO_2). While the silica itself is not highly adsorptive, its large surface area and porosity help trap suspended solids. The surface can become slightly negatively charged, enabling the adsorption of positively charged metal ions, such as lead and chromium. The small amounts of clay minerals present in sand, such as kaolinite and montmorillonite, make it highly adsorptive due to its high surface area, negative surface charges that attract positively charged ions (like heavy metals), and the presence of functional groups (e.g., OH^-, $Si\text{-}O^-$, $Al\text{-}O^-$) that bind contaminants. Sand also has trace amounts of iron and aluminum oxides and hydroxides (Fe_2O_3, FeOOH, AI_2O_3, $Al(OH)_3$).

The river sand was rinsed thoroughly in clean water and then dried in the sun.

The sand media helps in:

• Removing organic matter by trapping suspended particles (like plant material, algae) as well as dissolved organic compounds (adsorbs natural organic matter from decaying plants and animals).

• Removing bacteria and pathogenic microorganisms. While sand does not kill bacteria, it traps them, and over time, a biofilm of microorganisms that grows in the sand can break down the organic material. This gets removed when the media of the column is cleaned as part of the maintenance cycle.

• Removing organic sludge that accumulates from surface runoff or decaying biological material (like fecal sludge).

• Partially capturing small amounts of oil and grease, particularly if they are bound to other particles or sediments.

• Adsorbs heavy metals and is effective for the removal of chromium (VI) and lead.

• Adsorbing arsenic and phosphates.

Layer 2: Slag wool

Slag wool is a by-product of steel production. It is a fibrous material made from molten slag that has been spun into fibers. It features a dense fiber matrix, a porous structure, and a high surface area, making it ideal for adsorption. Its chemical composition includes SiO_2 35–50%, calcium oxide (CaO) 10–30% which enhances chemical reactivity with heavy metals like lead and copper, aluminum oxide (Al_2O_3) that facilitates adsorption sites via surface hydroxyl (–OH) groups, magnesium oxide (MgO) which facilitates ion exchange and neutralization reactions, and iron oxides (Fe_2O_3/FeO), that can adsorb anions like phosphate and arsenate. It also assists in physical filtration by acting as a fine mesh, especially right below the sand layer.

It helps in the following:

• Effective adsorption of certain metals, especially heavy metals like lead, cadmium, zinc, copper, nickel, chromium, and arsenic, via ion exchange

• Removing phosphates that bind to the iron, aluminum, and calcium oxides

• Removing suspended solids like silt, sand, and larger particulates that could cloud water by acting as a physical barrier to sedimentation, thereby reducing turbidity

• Trapping small organic particles (like plant material and algae)

• Absorbing oils, greases, dyes

• Partially removing bacteria and microorganisms (if they are attached to larger particles); however, it does not kill pathogens or remove dissolved biological contaminants.

Layer 3: Pink sand

Pink sand is composed mainly of quartz, feldspar ($KAlSi_3O_8$, $NaAlSi_3O_8$, $CaAl_2Si_2O_8$), calcium carbonate ($CaCO_3$), and traces of iron oxides (Fe_2O_3, Fe_3O_4). This enhances its adsorption capabilities compared to river sand. In filtration processes, it helps in the following:

• Slowing down the flow of water in the filtration column significantly. This is one of the major reasons for adding this layer.

• Reducing suspended solids and organic matter further – the fine-grain size of pink sand provides a large surface area for good bacteria to attach to and thrive, aiding in water filtration

• Partially removing bacteria

• Adsorbs traces of heavy metals, especially iron, manganese, and arsenic, as well as phosphates

• pH buffering by neutralizing acidic water

Before placing a layer of activated charcoal, a thick bundle of Grade 1 filter papers was tightly packed to ensure that grains of pink sand do not get pushed down into the activated charcoal below.

Layer 4: Activated carbon

Activated carbon is an adaptable adsorbent due to its properties, including a large surface area, high pore volume, diverse pore structure, extensive adsorption capacity, and a high degree of surface reactivity. It is prepared from various carbonaceous source materials, such as agricultural waste like coconut shells, wood (mostly bamboo cane and acacia), agricultural residues (like rice husk, betel nut husk, sugarcane bagasse), seeds (mango, papaya), and shells (like tamarind shell, cashew nutshell, acorn shell, banana peel, palm kernel shell). The higher adsorption capability of activated carbon depends on porous characteristics such as surface area, pore size distribution, and pore volume. The porous structure of activated carbon forms during the carbonization process, and it further develops during the activation process. The pore system of activated carbon differs from one another, and individual pores vary in size, ranging from less than a nanometer to thousands of nanometers.

In my filtration column, I have used activated charcoal made from coconut shells, which has a higher density of micro-pores as compared to other forms of activated carbon, meaning it has a higher surface area and porosity. This layer helps to:

• Adsorb fluoride, arsenic, chlorides, heavy metals (like iron, lead, mercury, copper, zinc, cadmium, nickel), trace metals (chromium, uranium).

• Remove dissolved organic compounds (pesticides, petroleum-based products, industrial chemicals, natural organic matter, organic dyes from many industries, such as the textile, cosmetics, leather, printing, rubber, and food industries), chlorine and chlorinated compounds, volatile organic compounds (benzene, toluene, xylene), and pharmaceutical residues.

• Absorb odor-causing compounds (mostly industrial pollutants) and inorganic compounds like sulfates, improving taste.

• Address hardness and salinity issues.

• Adsorb microbial contaminants to some extent within the microscopic pores of the carbon.

Layer 5: Ceramic cones.

Ceramic cones are made from locally sourced clay materials, including kaolin clay, alumina, silica, feldspar, and even sawdust and rice husks. They are fired at controlled temperatures to create porous structures ideal for filtration. In this filtration column, cones made of kaolin clay are used. This layer enables mechanical filtration, adsorption, and even antimicrobial action. It is ideal for filtering out dissolved inorganic compounds (such as salts and nitrates) and trapping bacterial contaminants in its microscale pores. Ceramic cones can filter out particulates as small as 0.2–0.5 microns. They are durable and can be cleaned multiple times, extending their usability. Moreover, this layer provides support for all the previous layers of media, ensuring that the smaller particles do not clog the outlet pipe.

For further research, these layers were used to create a filtration column in a clean plastic container, as exhibited in Figure 5.

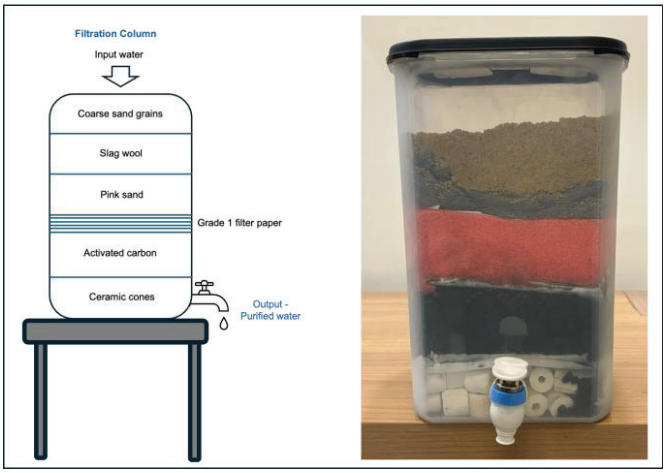


Figure 5: Depiction of the filtration column created for water purification. The layers of materials were assembled to create a filtration column for treating contaminated water, with each layer being approximately 4 inches in thickness, providing ideal contact time with the filtration media.

During the experiments, it was observed that the efficacy of the filtration column decreased after 30 to 45 days. Hence, all the media needs to be replaced or cleaned. The layer of activated charcoal in the filtration column was replaced, while the rest of the media was boiled separately in water, dried in the sun, and then reused.

DOI: 10.36838/v8i3.1

Result and Discussion

The three contaminated water samples were filtered through three separate filtration columns. The water collected from the respective outlets of the columns was tested separately through an accredited lab. The pH levels post-filtration were examined, but since they were well within the range prescribed by BIS across all three samples, they were not tabulated in the subsequent analysis.

Post filtration, the foul odor in S2 was eliminated. The TDS levels, which were higher than the industry standards across S2, S3, and S4, were brought down to 126 mg/L, 53 mg/L, and 102 mg/L, below the industry standard of 500 mg/L. From Figure 6, it is evident that the TDS levels decreased by approximately 95% in all three samples.

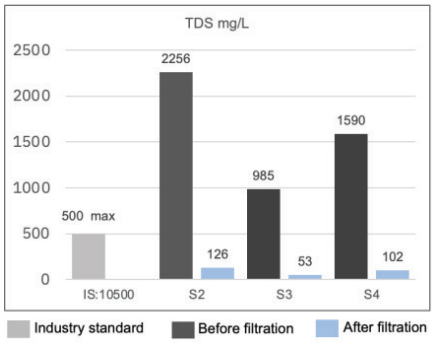


Figure 6: Comparison of TDS levels in drinking water samples before and after filtration. Post filtration of each contaminated through the filtration columns, it was observed that the TDS levels were brought down to levels that were much lower than the prescribed industry standard.

A chemical analysis post-filtration indicated a significant drop in contamination levels. S2, which had high levels of fluoride, chloride, nitrate, and sulfate, witnessed a drastic reduction in contamination levels, meeting industry standards. Fluoride levels went down from 2.1 mg/L to 0.65 mg/L. Chloride levels significantly reduced from 2769 mg/L to 48 mg/L. Nitrate levels fell from 110 mg/L to 42 mg/L. Sulfate levels fell from 364 mg/L to 140 mg/L. These results are showcased in Figure 7.

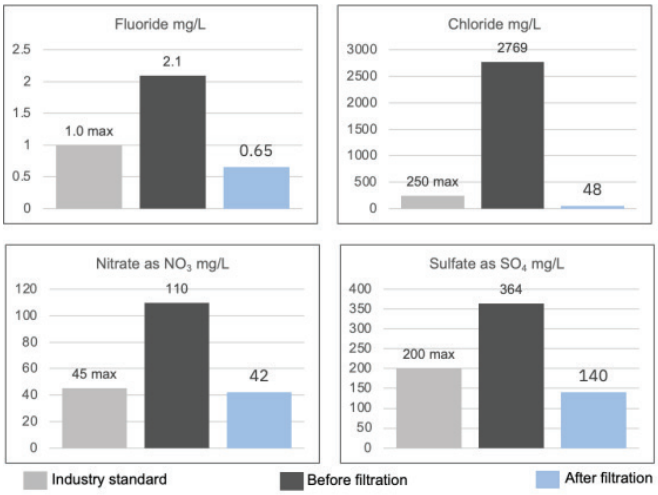


Figure 7: Comparison of contamination levels of chemicals in S2 before and after filtration. The high contamination levels of fluoride, chloride, nitrate, and sulfate in S2 were all addressed post-filtration and fell to levels below the prescribed industry standards.

S3, which had high levels of fluoride and chloride, showed an 82% and 96% decrease, respectively, with both levels meeting industry standards post-filtration, as evident from Figure 8. Fluoride levels went down from 1.6 mg/L to 0.3 mg/L. Chloride levels significantly reduced from 1127 mg/L to 41 mg/L.

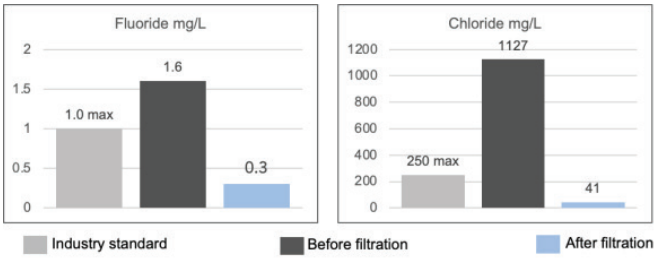


Figure 8: Comparison of contamination levels of chemicals in S3 before and after filtration. Fluoride and chloride levels in S3, which were higher than the prescribed norms, were treated effectively, post-filtration.

S4, which had high levels of chloride, nitrate, and sulfate, experienced a drastic reduction in contamination levels, meeting industry standards, as shown in Figure 9. Chloride levels fell from 2769 mg/L to 48 mg/L, nitrate levels fell from 110 mg/L to 42 mg/L, and sulfate levels fell from 285 mg/L to 128 mg/L.

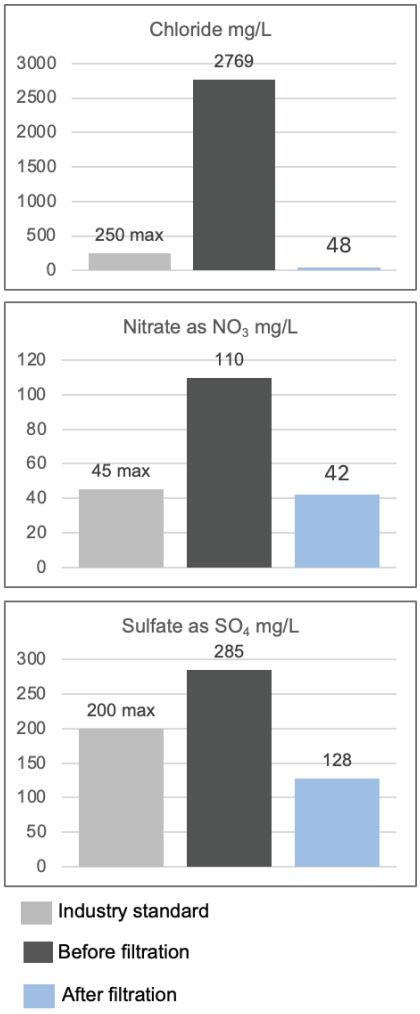


Figure 9: Comparison of contamination levels of chemicals in S4. S4, which had high levels of chloride, nitrate, and sulfate, showed a significant decrease in the levels of these chemicals, with levels meeting industry standards post-filtration.

Thus, the metal analysis results also show a significant reduction in contamination levels. The iron and lead levels, which were high in all three samples, fell to levels that meet industry standards. However, the chromium, cadmium, and uranium levels, although significantly lower post-filtration, did not meet the BIS standards. Chromium level in S4 fell by 75% from 4.08 mg/L to 1.02 mg/L; however continued to be higher than the industry standard of 0.05 mg/L maximum. Cadmium level in S2 fell by 97% from 3.09 mg/L to 0.08 mg/L; however continued to be higher than the industry standard of 0.01 mg/L maximum. Uranium level in S4 fell by 98% from 53.9 mg/L to 1.08 mg/L, but continued to be higher than the BIS standard of 0.03 mg/L maximum. Hence, the water samples S2 and S4 were unsafe for consumption due to very high contamination levels of heavy metals. To reduce these contaminants to meet industry standards, certain activating agents will need to be researched further and added to the filtration column.

S3, which had the presence of *E. coli* and coliforms, showed that these contaminants were eliminated post-filtration, as indicated by the laboratory test results.

Conclusion

Multiple factors influence groundwater quality. These include local geology, land use, climatic conditions, patterns and frequencies of rainfall, and anthropogenic activities such as the use of fertilizers and pesticides in agriculture, the disposal of domestic sewage and industrial effluents, and the extent of groundwater resource exploitation. Low-income households cannot afford commercially available solutions, such as reverse osmosis systems, for installation in their homes. High levels of dissolved solids, fluoride, chloride, arsenic, nitrates, sulfates, phosphates, metals (such as iron, manganese, lead, mercury, cadmium, and chromium), organic contaminants (including pesticides, herbicides, oil, and hydrocarbons), and microbes in drinking water have toxicological and epidemiological implications.

Home-built water filtration solutions, based on easily available and affordable filtration media, can help resolve the issue of drinking water quality for low-income households, for whom commercial solutions are unaffordable. Given that the type of contaminants and degree of contamination vary across regions, a home filtration solution must utilize media that address multiple contaminants to provide a safe source of water. This is evident from the results of the water samples, which had multiple contaminants, tested at an accredited laboratory before and after filtration through a series of media in a filtration column.

A major advantage of the filtration mechanism examined in my project is that it is gravity-based and does not require any electricity. The materials used in the column are low-cost, easily available, require minimal maintenance, and address a wide variety of contaminants, including physical, chemical, heavy metals, and microbiological. Initial material cost for a single filtration column to filter roughly 10 liters per day works out to approximately INR 700-800 (less than USD 10). The replacement cost for approximately 30 days is only for activated charcoal, the cost of which is approximately INR 100-150 (USD 1.2-1.8). The contaminants in the different water samples used in the project were varied. By leveraging low-cost, readily available, and low-maintenance media for water treatment, households in resource-constrained communities can address the issue of multiple contaminants in water for human consumption.

Going forward, some areas will require further research. There are potential limitations to the filtration mechanism for high concentrations of certain contaminants, such as heavy metals. In case the contamination levels, especially of heavy metals, are way too high, this filtration mechanism is unlikely to bring them down to the WHO standards, as was the case in two of my sample results post-filtration. The cadmium level in S2 and the chromium and uranium levels in S4, although they fell significantly, did not reach the levels specified in the BIS standards. Additional research and experiments are needed to determine the type of agents required and whether they can be incorporated into this project or if specific technical interventions are necessary.

Acknowledgments

I would like to acknowledge the guidance from Dr. Soumik Banerjee, Ph.D., Chief Scientific Officer, Powell Laboratories, Kolkata, India, and his inputs on the scientific rationale of the media being used for filtration of groundwater.

References

1. WHO/UNICEF Joint Monitoring Programme for water supply, sanitation, and hygiene. (2021). *Progress on household drinking water, sanitation, and hygiene 2000-2020: five years into the SDGs*. World Health Organization (WHO) and the United Nations Children's Fund (UNICEF).
2. Mapping Safe Drinking Water Use in Low- and Middle-income Countries; journal-article; 2021; Vol. 372, pp 928–932. https://doi.org/10.1126/science.abf2946.
3. Agapitova, N., Navarrete Moreno, C., Barkataky, R., & World Bank. (2017). *WaterLife: Improving access to safe drinking water in India* [Report]. The World Bank. https://documents1.worldbank.org/curated/en/586371495104964514/pdf/115133-WP-P152203-PUBLIC-17-5-2017-12-28-1-WaterlifeCaseApril.pdf
4. WHO. (2025). WHO drinking water specifications. In *WHO Drinking Water Specifications*. https://cpcb.nic.in/who-guidelines-for-drinking-water-quality/
5. Saha, D., Central Ground Water Board, Singh, B. P., Srivastava, S. K., Dwivedi, S. N., Mukherjee, R., & Gupta, S. (2014). Concept Note on Geogenic Contamination of Ground Water in India with a special note on Nitrate. In *Central Ground Water Board*. Ministry of Water Resources, Govt. of India.
6. Stati. (n.d.). *WATER QUALITY DATA OF GROUND WATER UNDER NWMP - 2022*.
7. Prasad Mookerjee, S. (n.d.). Operation and Maintenance: Drinking water quality monitoring and surveillance. In *Chapter 8*.
8. Kanel, S. R., Jr., Das, T. K., Varma, R. S., Kurwadkar, S., Chakraborty, S., Joshi, T. P., Bezbaruah, A. N., & Nadagouda, M. N. (2023). Arsenic contamination in groundwater: Geochemical basis of treatment technologies. *In ACS Environ. Au* (p. 135–152). https://doi.org/10.1021/acsenvironau.2c00053

DOI: 10.36838/v8i3.1

9. Nagarnaik, P., Maldhure, A. V., Alam, N., CSIR-NEERI, Bhattacharya, A., Srivastava, A., Sharma, V., Rajasekhar, D., Muralidharan, A., Tewari, P. K., Purkait, M. Kr., Philip, L., UNICEF, KPMG, E&Y, R., S., Budhraja, R., Mahajan, V., Lal, B., & Sheel, V. (n.d.). *Handbook on Drinking Water Treatment Technologies*.
10. BUREAU OF INDIAN STANDARDS & Drinking Water Sectional Committee. (2012). DRINKING WATER — SPECIFICATION. In the *Indian Standard* [Report]. https://cpcb.nic.in/wqm/BIS_Drinking_Water_Specification.pdf (Original work published 1983)

Authors

Adhit Mandal is a student at The Shri Ram School Aravali, Gurgaon, India. He is a passionate environmentalist and researcher in the areas of renewables and sustainable materials. He aspires to join a prestigious college course in Materials Science Engineering. He loves participating in competitions that give him an opportunity to research innovative solutions for real-life global issues.

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

A Systematic Review of Emerging Liquid Biopsy Techniques for the Detection of Melanoma

Anushka Pakhare

Juanita High School, 10610 NE 132nd St, Kirkland, WA, 98034, USA

ABSTRACT: Melanoma cancer is in the top 10 most diagnosed cancers worldwide yearly, causing around 57,000 deaths annually. Liquid biopsy in melanoma provides a novel staging and management tool that, in conjunction with tissue biopsies, can revolutionize treatment and care. It offers sufficient information about the melanoma in the patient's body by looking at various biomarkers present in bodily fluids. This systematic review provides a recent overview of rising liquid biopsy techniques for the detection and treatment response of melanoma. A total of 30 papers were compiled from 2019-2024. Results depict that ctDNA is the prominent biomarker used in the detection and treatment monitoring of stage III and IV melanoma. NGS and PCR-based assays are prominent genetic analysis tools for mutation analysis. There is a lack of use of liquid biopsy in early-stage melanoma for detection and treatment surveillance. Liquid biopsy is the future of the management, treatment, and survival of melanoma cancer patients; therefore, this needs to be addressed in more research.

KEYWORDS: Translational Medical Sciences, Disease Detection and Diagnoses, Melanoma Cancer, Liquid Biopsy, Biomarkers.

Introduction

In individuals ages 25-39, melanoma cancer is the 3rd most diagnosed cancer in the United States.[1] In 2013, 7,990 people died of melanoma, 1.3% of all cancer deaths globally.[2] By providing adequate screening and early treatment, these deaths could have been prevented, as melanoma cancer patients' 5-year survival rate is 99% if detection can occur before the cancer spreads to the lymph nodes. Therefore, it is important to prioritize regular check-ups for any suspicious symptoms because early detection might improve prognosis even more. As with most cancers, the earlier they are diagnosed, the better the outcome.

Between 2015 and 2019, melanoma mortality decreased by 4% every year because of the introduction of new immunotherapy drugs on the market.[3] This shows that providing treatment can help significantly improve the survival rate, but it can only be done if the patient is diagnosed early. If the disease is detected early and monitored during treatment using advanced methods such as liquid biopsy to identify and monitor a specific set of mutations that occur in DNA in the blood, the patient's mortality rate could decrease.

Liquid biopsy studies biomarkers in body fluids and fluid connective tissues such as blood, urine, tears, and CSF rather than regular tissues. Biomarkers are molecular and genetic indicators of normal or abnormal processes in a person's body, in this case, from cancerous and normal cells. Some biomarkers used in oncology are circulating tumor DNA (ctDNA), circulating cell-free DNA (cfDNA), circulating tumor cells (CTC), and tumor mutational burden (TMB).[4] Liquid biopsy is an essential tool that can be used to gain information on tumor possibility and mutations in your genetic code at a molecular level that imaging through a PET scan can't show us.

There are multiple reasons why blood tests are done on known cancer patients. Firstly, because of guided therapy. In guide therapy, if whole cells can be recovered, then they can be sequenced for specific mutations. Additionally, if ctDNA, cfDNA, or CTC levels go up, then doctors can know that the patient has relapsed. Commonly, when cfDNA levels it might trigger a full-body PET scan to look for recurrence.

One of the many benefits of liquid biopsy is personalized treatment. Personalized medicine is a new frontier extremely beneficial to cancer patients because each patient is treated according to their genetic profile. This can potentially improve the management of patients, minimize side effects, and increase survival rates. The benefit of using ctDNA or cfDNA from liquid biopsies is that the treatment will be able to combat unique problems for each specific patient, instead of having a generalized approach to various problems. This can solve not just the overarching problem, but also the small ones, in a patient, ensuring they live long, healthy lives. Another benefit, as mentioned above, is that it is noninvasive. Normal biopsies, while they are tiny and not very invasive, still require surgery to collect tissue for analysis and therefore put the patient at risk for bleeding, infections, and scarring. But with a liquid biopsy, similar critical information can be accessed completely noninvasively.

However, there are drawbacks and challenges. One challenge is that the concentration of tumor DNA available in the blood is not high enough. Cells, during apoptosis, shed only a little bit of their DNA into the blood because it is packaged into bound vesicles called apoptotic bodies. The rest is degraded by phagocytic cells. Therefore, amplification is required after extraction, usually based on a PCR test, to increase the targeted sequence amount so that mutations can be detected. With liquid biopsies, false negatives are also common in clinical settings

DOI: 10.36838/v8i3.9

because of the low amount of tumor DNA that is not detected by the biopsy. Thus, the doctors there would probably have to conduct a tissue biopsy, defeating the noninvasive benefit of liquid biopsies. To find rare mutations in ctDNA or cfDNA, detected in liquid biopsy, Next Generation Sequencing (NGS) is conducted. However, it is quite expensive. It was found that for Targeted Panel Testing, sequencing cancer-specific genes costs around 2100 dollars. For whole genome and comprehensive genome profiling, it costs even more, around 4950 and 3420 dollars. Thus, not all clinics worldwide can afford to conduct NGS for each melanoma cancer patient. Before the implication of any treatment based on data from liquid biopsy, there must be a thorough analysis of the results with proper health authorities and prominent scientists, which has not yet been implemented and will take time, too.

While liquid biopsy isn't an emerging technique for melanoma, there are others that are being studied more recently for detection. For example, by using artificial intelligence to analyze pictures of different skin lesions, doctors can evaluate their diagnosis with another source. Additionally, total body photography allows doctors to diagnose high-risk patients who most likely have a family history of melanoma. Dermoscopy, a process that uses a dermoscope to magnify sub-surface skin structures, is also becoming more sensitive and efficient to use to identify malignant melanoma in patients. Not necessarily a new technique, but multimodal approaches have become more common, so evidence can be corroborated by multiple sources. By using these new imaging and photography in conjunction with liquid biopsy, NGS, and immunotherapies for melanoma cancer detection and treatment, patients' survival can be key to happen.

Methods

Research on melanoma and liquid biopsy has been a prominent topic in the scientific community in the last 30 years. By targeting specific genes and using a broad range of technological tools for collecting data, scientists are able to understand how to prevent, locate, and treat melanoma cancer. This paper analyzes a small portion of the most recent research done.

A list of 30 papers on melanoma, liquid biopsy, biomarkers, and new treatments in the field, ranging from 2019 to 2023, was compiled for a literature search. Some key terms for search include "melanoma and liquid biopsy" and "melanoma and ctDNA." This information was summarized into a table with headings of methods, target genes, biomarkers, stage of cancer, and more. I've also stated the limitations and the future of liquid biopsy in melanoma cancer in the discussions. The papers that were selected adhered to specific criteria: non-systematic reviews and had to be only in the English language.

Results

Table 1: Compilation of scientific papers. Findings illustrate that ctDNA is the most prominent biomarker used. NGS and PCR are the most common tools, and analysis of PD-L1 is not common.

Author (Year)	Method	Target Gene	Research Question/ Topic	Biomarker	Tumor Stage	Sample Size	Preanalytical factors
Yin He, Xiaosheng Wang (2023)	Multiomics	-	Identifying biomarkers associated with immunotherapy response in melanoma by multi-omics analysis	TMB, PD-L1 expression	-	472	-
Gabriel Velez (2021)	Using vitreous biopsies for identifying biomarkers for uveal melanoma by examining independent cohort of patients	-	Can vitreous biopsy find biomarkers associated with uveal melanoma?	SCFR, HGF, and HGFR (Proteins)	3 and 4	8	-
Deborah H Im (2022)	Illumina sequencing and targeted sequencing	BAP1 and GNAQ	Does aqueous humor of uveal melanoma have sufficient ctDNA to perform genetic analysis?	ctDNA	1	37	-
John J. Park (2021)	ddPCR (digital droplet PCR) and NGS	TP53, GNAQ	Can baseline ctDNA in metastatic UM strongly correlate with baseline LDH level and disease volume?	ctDNA	Advanced stage	17	EDTA
Russell J. Diefenbach (2022)	NGS panels	TERT	Can a custom NGS panel detect mutations in the TERT promoter region in ctDNA?	ctDNA	3 and 4	21	EDTA
Claudia Sabato (2022)	RT-qPCR and PCR	-	Can a blood test based on microRNA non-invasively detect melanoma?	microRNA	3 and 4	19	EDTA
Carlos Alberto Aya-Bonilla (2020)	RT-PCR and ddPCR	19 different genes	Can multimarker derived CTC scores be used for prognostic and treatment response in metastatic melanoma?	CTC	-	43	-
Ekaterina Galanzha (2019)	In vivo blood test	-	Can a noninvasive in vivo blood test detect CTC in melanoma bloodstream by using our PAFC Cytophone?	CTC	3 and 4	18	-
Rongzhi Huang (2020)	Targeted panel of genes	Different genes	Can the immune-related genes prognosis biomarker be an effective potential prognostic classifier in the immunotherapies and surveillance of melanoma?	CMTM6	-	905	-
Sandra Fitzgerald (2023)	NGS and ddPCR	BRAF and NRAS	Dynamic ctDNA Mutational Complexity in Patients with Melanoma Receiving Immunotherapy	ctDNA	3 and 4	29	EDTA
Jenny H. Lee (2019)	Cox regression analyses and ddPCR	BRAF, NRAS, KIT	Can ctDNA that is detected before completing surgical resection in patients with AJCC stage IIIB/C/D (high-risk stage III)	ctDNA	3	174	EDTA

			with a BRAF, NRAS or KIT mutant melanoma show that its is an independent predictor of worse MSS in patients receiving no systemic adjuvant therapy?				
Elin S. Gray (2015)	ddPCR	BRAF and NRAS	Can we use ctDNA to monitor treatment response in metastatic melanoma?	ctDNA	M1a or b	48	EDTA
Russell J. Diefenbach (2020)	NGS	30 different genes	Can stage 3 or stage 4 melanoma be monitored by NGS from ctDNA analysis?	ctDNA	3 and 4	91	
Jesper Geert Pedersen (2020)	ddPCR	BRAF, NRAS, TERT	Can levels of ctDNA, MCP1, and TNFα predict disease progression in metastatic melanoma patients with checkpoint inhibitors?	ctDNA	3 and 4	16	EDTA
Claudia H. D. Le Guin (2021)	Illumina sequencing	GNAQ/ GNA11	Can detection of cfDNA in plasma provide a diagnostic lead time diagnosis of metastases or tumor recurrence in uveal melanoma?	cfDNA	-	151	-
Akira Kaneko (2021)	ddPCR and Capp Sequencing	BRAF	Can liquid biopsy CAPP sequencing and ddPCR detect tumor presence and mutations?	cfDNA	Early stage	15	-
Zeynep Eroglu MD (2023)	multiplex PCR	-	Predictive and prognostic value of a personalized, tumor-informed ctDNA assay for the detection of molecular residual disease (MRD)	ctDNA	3 and 4	69	-
Jan Braune (2020)	ddPCR	BRAF and NRAS	Can early changes in ctDNA predict responses to treatment and ctDNA for detecting tumor burden for melanoma	ctDNA	3 and 4	62	EDTA
Andrea Forschner (2020)	Illumina sequencing and ddPCR	BRAF and MEK	Can BRAF V600 mutant ctDNA can be used to reliably determine progressive disease under targeted therapy and whether patients' prognoses are different if ctDNA is detectable before initiating targeted therapy?	ctDNA	Advanced	19	EDTA
Marina Berger (2021)	NGS	BRAF and MEK	Can NGS gene panels be used for treatment monitoring in melanoma cancer through ctDNA analysis?	ctDNA	3 and 4	31	Strek
Selena Y Lin (2020)	qRT PCR	BRAF	Can blood molecular profiling of circulating tumor cells (CTCs) enable monitoring of patients with metastatic melanoma	CTC	3 and 4	75	Strek
			during checkpoint inhibitor immunotherapy (CII) + in combination with targeted therapies?				
Anthony Lucci (2023)	CellSearch	-	investigated how frequently and how early circulating tumor cells (CTCs) were identified prior to the surveillance imaging detection of melanoma progression	CTC	3	325	-
Ioana Gencia (2020)	Real time PCR	-	Compare miRNA expression between primary melanomas from different sites	microRNA	3 and 4	32	-
John J. Park (2021)	NGS and ddPCR	GNAQ, GNA11 and CYSTLR2	Can circulating tumor DNA reflect Uveal Melanoma responses to protein Kinase C inhibition?	ctDNA	-	17	EDTA
Pawel Sobczuk (2022)	qPCR	BRAF	Can we evaluate the clinical utility of plasma circulating tumor DNA analysis for BRAF mutation?	ctDNA	3 and 4	46	-
Russel J. Diefenback (2022)	NGS	TERT, BRAF. NRAS	Detection of melanoma mutations using circulating tumor DNA (ctDNA) is a potential alternative to using genomic DNA from invasive tissue biopsies.	ctDNA	3 and 4	21	EDTA
Lydia Warburton (2020)	ddPCR and NGS	BRAF	Can ctDNA be a reliable biomarker for disease progression after treatment was stopped?	ctDNA	Advanced	70	EDTA
Joesph W. Po (2019)	Flow cytometry	PD-L1	Can a reliable melanoma circulating tumor cell (CTC) detection method evaluate PDL1 on CTC's?	CTC	4	14	-
Carolyn Hall (2018)	Cox regression analyses	-	Determine if circulating tumor cells (CTCs) are associated with shortened (180-day) progression-free survival (PFS) after a baseline CTC assessment in stage IV melanoma patients.	CTC	4	93	-

Table 2: FDA-approved single and combination drug treatments for melanoma cancer. Results depict combination drugs to be more utilized currently.

Drug/treatment	Description	Single or combination	Month and year approved by the FDA
Talimogene laherparepvec/ IMLYGIC	Genetically modified herpes virus to create immunoprotein called GM-CSF	Single	October, 2015
Ipilimumab/YERVOY	Supports activation and proliferation of T-cells strengthening the immune system	Single	March, 2011
Nivolumab and retalimab- rmb/ OPUDUALAG	Uses blocking antibodies against PD1 and a lymphocyte activating gene (LAG-3) blocking antibody	Combination	March, 2022

DOI: 10.36838/v8i3.9

For melanoma, as found in this review, ctDNA is the most common biomarker used. 16/30 or 53.3% of the results depicted that ctDNA was the biomarker used for detection in their studies, showing that currently it is the most prominent. In earlier reports, the expression of PD-L1 has been more prominent, but that is not the case now: the majority of the studies focused on BRAF and NRAS genes. The results also show that ctDNA is more sensitive to III and IV late-stage cancers, as shown in 81.3% of the 16 studies. (Table 1) However, in the early detection of melanoma, as shown in the studies gathered, the sensitivity is very low for stage I or II ctDNA (3 out of 16). This is because there is less presence of cancerous cells; therefore, fewer cells are releasing DNA. You need a large amount of ctDNA in the blood for a liquid biopsy to have a high sensitivity. Additional work must be done to be able to use liquid biopsy to aid the genetic profiling of patients with earlier stages of melanoma, so that more personalized treatments can be made and given to them. It was also found that the most common genetic analysis technique used was next-generation sequencing (NGS) and droplet digital polymerase chain reaction (ddPCR) in the studies collected (Table 1).

One way that scientists and doctors are using the immune system is checkpoint inhibition. Checkpoint inhibition is the blockage of proteins that inhibit the immune system from attacking cancerous cells in the body. It is a type of immunotherapy, a new treatment that focuses on using the immune system to stop the spread of cancer cells. Specifically, in terms of melanoma, the inhibition of the immune checkpoint inhibitor PD-1 is a key treatment that can be done by two specific drugs, pembrolizumab and nivolumab, on the market. Nivolumab is depicted in Table 2 as one of 2 in the up-and-coming combination therapy.

FDA-Approved Drug Treatments:

As shown, in previous years, the use of a single drug for melanoma treatment was common. 50% of the therapies presented show that it is a combination therapy, and they are all in the last 5-7 years. On the other hand, single-drug treatment was more common in the last 10-15 years. Therefore, combination therapies have become more prevalent. This is because of the need to combat multiple aspects of melanoma tumors in patients' treatment. These various aspects could include increasing immune system strength, targeting and blocking specific genes, and creating proteins or antibodies. In relation to liquid biopsy, the use of combination therapies allows for targeting multiple cells, proteins, or kinases that can be recognized through sequencing after liquid biopsy. Liquid biopsy can monitor this treatment as well, and adjustments to delivery, drugs, or amount can be made accordingly.

Preanalytical Factors:

Preanalytical factors were also considered for bias in this systematic review. The amount of plasma in which tumor DNA was extracted, a range between 2-40 mL, affected the study's sensitivity. The type of tube they used to store the plasma, Strek or EDTA, also introduced bias. Additionally, the type of DNA extraction done, the different protocols conducted during this process, and how they quantified their data are preanalytical factors that introduce bias in their study.

■ Discussion

The application of invasive tissue biopsy in clinical settings is a prognostic tool that, with initial tissue biopsy diagnosis to detect the depth of melanoma, can become a revolutionary tool for monitoring and treating patients. By understanding the depth of tissue biopsy, the amount of radiation to treat the patient can be determined, and liquid biopsy can be used to check for relapses.

Firstly, liquid biopsy must be more sensitive to late-stage melanoma because of the increased amount of ctDNA in the fluids. Because 83% of the ctDNA studies were utilizing stage 3 and 4 melanomas, this shows that to procure enough DNA for sequencing and analysis, they needed fluid in which there was enough DNA for a liquid biopsy to detect. Connecting back to the introduction, this demonstrates how liquid biopsy can be an essential tool for monitoring treatment progress in late-stage cancers, as ctDNA levels are high enough for analysis. This shows the advantage of using liquid biopsies rather than using tissue biopsies to check progress, which will take time, money, and risk.

Secondly, the increased prevalence of BRAF, MEK, and NRAS gene sequencing over combating the protein PDL1 shows the scientific community heading toward combating melanoma at a genetic level. 14 studies explicitly stated using those genes in their process, while only 1 explicitly stated combating PDL1. A key part of personalized medicine, as stated in the introduction, is creating treatments according to the patient's own DNA. Therefore, more research is being done toward installing personalized medicine in melanoma detection and treatment, as sequencing genes is the first step.

While these studies are thorough on liquid biopsy, biomarkers, and combination therapies, there are some limitations. These studies lacked the use of biomarkers in earlier stages of melanoma. This is because their low concentration in the plasma extracted does not enable them to be used. Therefore, there should be an increased focus on early diagnosis in future studies that use liquid biopsy as the primary diagnostic tool. Liquid biopsy is the future of the detection of melanoma cancer.

■ Conclusion

The main results of this study show that the most common biomarker used for diagnosis and prognosis is ctDNA, which is extracted by the liquid biopsy technique for stage III or IV melanoma patients. In addition, the implementation of FDA-approved combination drug treatments in melanoma over singular treatments has risen in recent years. In conclusion, liquid biopsy is the future of melanoma cancer.

■ Acknowledgments

Anushka Pakhare would like to thank her mentor, Dr. Maria Neofytou.

References

1. Melanoma Research Alliance. Melanoma: facts & stats about skin cancer. Melanoma Research Alliance. Published 2023.Accessed February 21, 2024. Available at: https://www.curemelanoma.org/about- melanoma/melanoma-101/melanoma-statistics-2
2. National Cancer Institute. Melanoma of the skin - cancer stat facts. SEER. Published 2018. Accessed [date]. Available at:https://seer.cancer.gov/statfacts/html/melan.html
3. Fitzgerald S, Blenkiron C, Stephens R, Mathy JA, Somers-Edgar TJ, Rolfe G, Martin C, Jackson CB, Eccles MR, Robb TJ, Rodger J, Lawrence B, Guilford P, Lasham A, Print CG. Dynamic ctDNA mutational complexity in patients with melanoma receiving immunotherapy. Mol Diagn Ther. 2023; 27(4): 537-550. doi:10.1007/s40291-023-00651-4.
4. Nikanjam M, Kato S, Kurzrock R. Liquid biopsy: current technology and clinical applications. J Hematol Oncol 2022; 15(1). doi:10.1186/s13045-022-01351-y.
5. Aya-Bonilla CA, Morici M, Hong X, McEvoy AC, Sullivan RJ, Freeman J, Calapre L, Khattak MA, Meniawy TM, Millward M, Ziman MR, Gray ES. Detection and prognostic role of heterogeneous populations of melanoma circulating tumour cells. Br J Cancer 2020; 122(7): 1059-1067. doi:10.1038/s41416-020-0750-9.
6. Berger M, Thueringer A, Franz D, Dandachi N, Talakić E, Richtig G, Richtig E, Rohrer M, Koch L, Wolf IH, Koch C, Rainer BM, Koeller M, Pichler M, Gerritsmann H, Kashofer K, Aigelsreiter A. Circulating tumor DNA as a marker for treatment response in metastatic melanoma patients using next-generation sequencing—a prospective feasibility study. Cancers (Basel). 2021; 13(12): 3101. doi:10.3390/cancers13123101.
7. Braune J, Keller L, Schiller F, Graf E, Rafei-Shamsabadi D, Wehrle J, Follo M, Philipp U, Hussung S, Pfeifer D, Mix M, Duyster J, Fritsch R, von Bubnoff D, Meiss F, von Bubnoff N. Circulating tumor DNA allows early treatment monitoring in BRAF- and NRAS-mutant malignant melanoma. JCO Precis Oncol. 2020; (4): 20-31. doi:10.1200/po.19.00174.
8. Bornfeld N, Bechrakis N, Jabbarli L, Richly H, Lohmann DR, Zeschnigk M. Early detection of metastatic uveal melanoma by the analysis of tumor-specific mutations in cell-free plasma DNA. Cancer Med. 2021; 10(17): 5974-5982. doi:10.1002/cam4.4153.
9. Diefenbach RJ, Lee JH, Menzies A, Carlino MS, Long GV, Saw RP, Howle JR, Spillane AJ, Scolyer RA, Kefford RF, Rizos H. Design and testing of a custom melanoma next-generation sequencing panel for analysis of circulating tumor DNA. Cancers. 2020; 12(8): 2228. doi:10.3390/cancers12082228.
10. Spillane J, Scolyer RA, Kefford RF, Rizos H. Design and testing of a custom melanoma next-generation sequencing panel for analysis of circulating tumor DNA. Cancers. 2020; 12(8): 2228. doi:10.3390/cancers12082228.
11. Diefenbach RJ, Lee JH, Stewart A, Menzies AM, Carlino MS, Saw RPM, Stretch R, Long GV, Scolyer RA, Rizos H. Anchored multiplex PCR custom melanoma next-generation sequencing panel for analysis of circulating tumor DNA. Front Oncol 2022; 12: 820510. doi:10.3389/fonc.2022.820510.
12. Duvivier H. Anti PD-1 therapy in melanoma. Cancer Treatment Centers of America. Published 2021. Accessed March 23, 2024. Available at: https://www.cancercenter.com/cancer-types/melanoma/treatments/pd-1-blockers.
13. Erogl Z, Krinshpun S. Circulating tumor DNA-based molecular residual disease detection for treatment monitoring in advanced melanoma patients. American Cancer Society. Published 2023. Accessed March 12, 2024. Available https://acsjournals.onlinelibrary.wiley.com/action/showCitFormats?doi=10.1002%2Fcncr.34716
14. Kelkenberg S, Sinnberg T, Garbe C, Biskup S, Battke F. Circulating tumor DNA correlates with outcome in metastatic melanoma treated by BRAF and MEK inhibitors – results of a prospective biomarker study. OncoTargets Ther. 2020; 13: 5017-5032. doi:10.2147/ott.s248237.
15. AIM at Melanoma Foundation. FDA approved drugs. AIM at Melanoma Foundation. Accessed March 3, 2024. Available at:https://www.aimatmelanoma.org/how-melanoma-is-treated/fda-approved-drugs/#1598840228178-ee1bae38-4236
16. Galanzha EI, Menyaev YA, Yadem AC, Sarimollaoglu M, Juratli MA, Nedosekin DA, Foster SR, Jamshidi-Parsian A, Siegel ER, Makhoul I, Hutchins LF, Suen JY, Zharov VP. In vivo liquid biopsy using Cytophone platform for photoacoustic detection of circulating tumor cells in patients with melanoma. Sci Transl Med. 2019; 11(496) doi:10.1126/scitranslmed.aat5857.
17. Gencia I, Baderca F, Avram S, Gogulescu A, Marcu A, Seclaman E, Marian C, Solovan C. A preliminary study of microRNA expression in different types of primary melanoma. Bosn J Basic Med Sci. 2019; 20(2). doi:10.17305/bjbms.2019.4271.
18. Gray ES, Rizos H, Reid AL, Boyd SC, Pereira MR, Lo J, Tembe V, Freeman J, Lee JHJ, Scolyer RA, Siew K, Lomma C, Cooper A, Khattak MA, Meniawy TM, Long GV, Carlino MS, Ziman M. Circulating tumor DNA to monitor treatment response and detect acquired resistance in patients with metastatic melanoma. Oncotarget. 2015; 6(39). doi:10.18632/oncotarget.5788.
19. Hall CS, Ross M, Bowman Bauldry JB, Upshaw J, Karhade MG, Royal R, Patel S, Lucci A. Circulating tumor cells in stage IV melanoma patients. J Am Coll Surg. 2018; 227(1):116. doi:10.1016/j.jamcollsurg.2018.04.026.
20. Heidrich I, Ačkar L, Mossahebi Mohammadi P, Pantel K. Liquid biopsies: Potential and challenges. *Int J Cancer*. 2021; 148(3): 528-545. doi:10.1002/ijc.3321721.
21. He Y, Wang X. Identifying biomarkers associated with immunotherapy response in melanoma by multi-omics analysis. ComputBiol Med. 2023; 167: 107591. doi:10.1016/j.compbiomed.2023.107591.
22. Huang R, Mao M, Lu Y, Yu Q, Liao L. A novel immune-related genes prognosis biomarker for melanoma: associated with tumor microenvironment. Aging. 2020; 12(8): 6966-6980. doi:10.18632/aging.103054.
23. Im DH, Peng CC, Xu L, Kim ME, Ostrow D, Yellapantula V, Bootwalla M, Biegel JA, Gai X, Prabakar RK, Kuhn P, Hicks J, Berry JL. Potential of aqueous humor as a liquid biopsy for uveal melanoma. Int J Mol Sci. 2022; 23(11): 6226.doi:10.3390/ijms23116226.
24. Kaneko A, Kanemaru H, Kajihara I, Mijiddorj T, Miyauchi H, Kuriyama H, Kimura T, Sawamura S, Makino K, Miyashita A, Aoi J, Makino T, Masuguchi S, Fukushima S, Ihn H. Liquid biopsy-based analysis by ddPCR and CAPP-Seq in melanoma patients. JDermatol Sci. 2021; 102(3): 158-166 doi:10.1016/j.jdermsci.2021.04.006.
25. Lin SY, Chang SC, Lam S, Ramos RI, Tran K, Ohe S, Salomon MP, Asgar A, Lim CT, Fischer TD, Foshag LJ, Boley CL, O'Day SJ, Hoon DSB. Prospective molecular profiling of circulating tumor cells from patients with melanoma receiving combinatorial immunotherapy. Clin Chem. 2019; 66(1): 169-177. doi:10.1373/clinchem.2019.307140.
26. Lucci A, Addanki S, Chiang YJ, Meas S, Sarli VN, Upshaw JR, Manchem M, Patel SP, Wargo JA, Gershenwald JE, Ross MI. Presence of circulating tumor cells predates imaging detection of relapse in patients with stage III melanoma. Cancers.2023; 15(14): 3630. doi:10.3390/cancers15143630.
27. National Cancer Institute. Melanoma of the skin - cancer stat facts. SEER. Published 2018. Available at:https://seer.cancer.gov/statfacts/html/melan.html. Accessed February 14, 2024.

DOI: 10.36838/v8i3.9

28. Nikanjam M, Kato S, Kurzrock R. Liquid biopsy: current technology and clinical applications. J Hematol Oncol. 2022; 15(1). doi:10.1186/s13045-022-01351-y.
29. American Academy of Dermatology Association. Skin cancer. Available at: https://www.aad.org/media/stats-skin-cancer#:~:text=The%20five%2Dyear%20survival%20rate. Published 2022. Accessed March 10, 2024.
30. Park JJ, Diefenbach RJ, Byrne N, Long GV, Scolyer RA, Gray ES, Carlino MS, Rizos H. Circulating tumor DNA reflects uveal melanoma responses to protein kinase C inhibition. Cancers. 2021; 13(7): 1740. doi:10.3390/cancers13071740.
31. Pedersen JG, Madsen AT, Gammelgaard KR, Aggerholm-Pedersen N, Sørensen BS, Øllegaard TH, Jakobsen MR. Inflammatory cytokines and ctDNA are biomarkers for progression in advanced-stage melanoma patients receiving checkpoint inhibitors. Cancers. 2020; 12(6): 1414. doi:10.3390/cancers12061414.
32. Po JW, Ma Y, Bavanthi Balakrishna, Brungs D, Azimi F, Paul de Souza, Becker TM. Immunomagnetic isolation of circulating melanoma cells and detection of PD-L1 status. PLoSONE. 2019; 14(2). doi:10.1371/journal.pone.0211866.
33. Sabato C, Noviello TMR, Covre A, Coral S, Caruso FP, Besharat ZM, Splendiani E, Masuelli L, Battistelli C, Vacca A, Catanzaro G, Po A, Anichini A, Maio M, Ceccarelli M, Di Giacomo AM, Ferretti E. A novel microRNA signature for the detection of melanoma by liquid biopsy. J Transl Med. 2022; 20 (1). doi:10.1186/s12967-022-03668-1.
34. Gray ES, Reid AL, Bowyer S, Calapre L, Siew K, Pearce R, Cowell L, Frank MH, Millward M, Ziman M, Long GV, Carlino MS, Menzies AM, Sandhu SK, Scolyer RA, Rizos H. Circulating tumor DNA to monitor treatment response and detect acquired resistance in patients with metastatic melanoma. *Journal of Clinical Oncology*. 2022; 40(9): 858-866. doi:10.1200/JCO.21.01729.
35. Velez G, Nguyen HV, Chemudupati T, Ludwig CA, Toral M, Reddy S, Mruthyunjaya P, Mahajan VB. Liquid biopsy proteomics of uveal melanoma reveals biomarkers associated with metastatic risk. *Molecular Cancer*. 2021; 20 (1): 39. doi:10.1186/s12943-021-01336-4.
36. Warburton L, Calapre L, Pereira MR, Reid A, Robinson C, Amanuel B, Ziman M, Millward M, Gray E. Circulating tumour DNA in advanced melanoma patients ceasing PD1 inhibition in the absence of disease progression. *Cancers*. 2020; 12 (11): 3486. doi:10.3390/cancers12113486.

Author

Anushka Pakhare is currently a junior at Juanita High School. She is interested in the field of biology and medicine and will pursue those interests in her future.

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Emerging Cannabidiol Research in Neurological Disorders: Insights into Epilepsy, Multiple Sclerosis, and Depression

Juwon Kim

Seoul Academy, 422 Yeoksam-ro, Gangnam-gu, Seoul, 06200, Republic of Korea; joellekim07099@gmail.com
Mentor: Charity N Bhebhe

ABSTRACT: The cannabis plant contains 100+ distinct cannabinoids, and the two most researched are cannabidiol (CBD) and Δ9-tetrahydrocannabinol (Δ9-THC). Cannabinoids offer new treatment options for neurological disorders and their symptoms, including epilepsy, multiple sclerosis (MS), and depression. This review summarizes the recent findings in evaluating the therapeutic potential of CBD in various neurological disorders, such as epilepsy, MS, and depression, and the mechanisms behind its anticonvulsant, anti-inflammatory, and analgesic properties. CBD significantly reduces epileptic seizures in children and young adults with only mild side effects. In MS, studies on the therapeutic potential of cannabinoids have shown mixed results due to limitations in study design or other sources of variability, but they suggest that a combination of CBD and THC can reduce spasticity. In depression, higher doses of CBD appear beneficial and well tolerated, yet individual differences among patients and the limited number of clinical trials indicate that further research is needed to better understand these effects.

KEYWORDS: Biomedical and Health Sciences, Immunology, Cannabidiol, Epilepsy, Multiple Sclerosis (MS), Depression.

Introduction

Progress has been made in the exploration of the medicinal use of cannabis in humans. In particular, there has been promising evidence that cannabis could treat the symptoms of neurological disorders, including epilepsy, MS, Parkinson's Disease, Huntington's Disease, Autism Spectrum Disorder, and Complex Motor Disorders.[1] For example, a prominent single-case study of a girl named Charlotte, who was diagnosed with SCN1A-confirmed Dravet Syndrome, provided evidence that CBD administration may significantly decrease the number of epilepsy seizures. By month three of high-concentration CBD extract, Charlotte had a >90% reduction in tonic-clonic seizures. The term "Charlotte's Web" gained recognition following its use in her case, contributing to the popularization of medical marijuana for seizures.[2] The use of cannabis has doubled in the last 20 years, with medical use rising between 2013 and 2020 in the United States, where the plant is widely legalized or decriminalized, emphasizing the need to expand our research on the medicinal use of cannabis.[3] The present review aims to analyze and examine the potential of cannabinoids in treating neurological disorders, focusing on future research and therapeutic applications. As the medical use of cannabis products continues to grow in the United States, more people are becoming interested in its medicinal effects.

Cannabis, also commonly called marijuana or weed, refers to a type of plant that can produce psychoactive effects when consumed.[4] Cannabis is also a diverse plant genus known for its industrial purposes and medical compounds, with a long history of human use. This family of plants consists of three distinct species: *Cannabis sativa*, *Cannabis indica*, and *Cannabis ruderalis*, all of which have psychoactive effects.[5] CBD and THC are both natural compounds synthesized in cannabis and identified as phytocannabinoids with similar chemical structures. Phytocannabinoids are mainly present in the trichomes of female cannabis plants, while male leaves produce less psychoactive substances. Both THC and CBD are biosynthesized in the plant from olivetolic acid, which is converted to cannabigerolic acid (CBGA) by the enzyme CBGA synthase using geranyl diphosphate as a substrate. CBGA is the central precursor of Δ9-THCA and CBDA. The two acids are then decarboxylated (neutralized) into their neutral forms, THC and CBD.[6] THC serves as the primary psychoactive component in cannabis, while CBD does not induce euphoria. Compared to THC, CBD's flexible structure with a free-moving hydroxyl group and open-ring conformation allows it to adopt multiple shapes and interact with many different receptors.[7] (Figure 1) Both THC and CBD have been researched as potential avenues for the treatment of anxiety, depression, sleep apnea, and neurological disorders.[8] Although both compounds have been proposed to act, at least in part, through the endocannabinoid system (ECS), they produce markedly different effects.

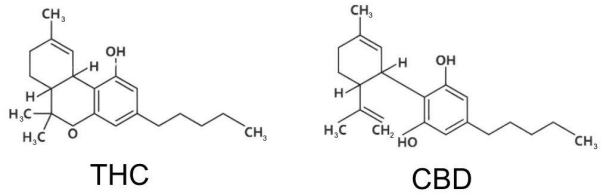


Figure 1: Molecular structures of THC and CBD.

Since THC directly binds to cannabinoid receptors (CB1 and CB2) linked to the brain's reward system or dopaminergic pathway, the substance induces psychoactive effects. Thus, it can also cause adverse events such as paranoia and hallucinations if consumed in excess.[9] However, since CBD interacts with the ECS by indirectly influencing the CB1 and CB2 receptors rather than binding directly to them, it has less se-

DOI: 10.36838/v8i3.15

vere effects when consumed. CBD exhibits a low affinity for cannabinoid receptors and seems to exert its therapeutic effects in other sites. It interacts with ion channels to exert anticonvulsant effects, modulates cyclooxygenase and lipoxygenase enzymes to produce analgesic outcomes, targets the periaqueductal gray area to mediate antinociceptive responses, and engages 5-hydroxytryptamine 1A (5-HT1A) receptors to facilitate anxiolytic effects.[10] The mechanisms behind the interaction between ECS, THC, and CBD are discussed in the neurochemistry section. CBD is recognized for its antiepileptic, anti-inflammatory, antioxidant, and analgesic properties, making it a promising therapeutic agent for neurological and inflammatory conditions such as multiple sclerosis (MS), Dravet syndrome, Lennox-Gastaut syndrome, Parkinson's disease, post-traumatic stress disorder (PTSD), fibromyalgia, and other disorders associated with dysregulated immune responses. The therapeutic effects of CBD are thought to be mediated through multiple pathways that include modulation of the endocannabinoid system as well as actions at non-ECS targets such as ENT1–A2A adenosine signaling, GPR55, TRPV1, and 5-HT1A receptors.[11]

The History of Cannabis Use:

Originally from Central and Southeast Asia, the cannabis plant has been used for centuries for entheogenic and religious purposes, as well as in traditional medicine. Early historical records indicate its application in ancient China for alleviating rheumatic pain, fatigue, and inflammatory conditions, and its use in textile manufacturing.[12] In ancient Rome, the Roman historian Pliny recorded the use of *Cannabis sativa* roots for relieving pain. In India, it has been revered as a sacred plant for both medicinal and spiritual purposes since approximately 1000 BCE, subsequently disseminating through Persia, Europe, and the Americas.[13] During the 19th and early 20th centuries, cannabis began to be acknowledged in Western medicine, with notable contributions from individuals such as William B. O'Shaughnessy and Jacques-Joseph Moreau. Both experimented with their patients and demonstrated that the plant had analgesic and anticonvulsant properties. However, its popularity diminished due to ethical and economic issues, leading to the International Drug Control Treaty in 1925 followed by the Marijuana Tax Act in the U.S. in 1937.[14] Variable efficacy among patients, the emergence of alternative medications, and increasing legal restrictions—partly due to rising recreational use—contributed to its removal from the U.S. Pharmacopeia in 1941.[15] Many experimentations involving the medicinal use of the drug were terminated accordingly.

Despite the challenges, a resurgence of interest occurred in the late 20th century following the clarification of THC's chemical structure by the scientists Yehiel Gaoni and Raphael Mechoulam.[16] The finding was then followed by the discovery of cannabinoid receptors and the identification of endocannabinoids, which sparked renewed scientific exploration into their medicinal applications.

Neurochemistry and Mechanisms of Cannabis/Cannabinoids/ECS:

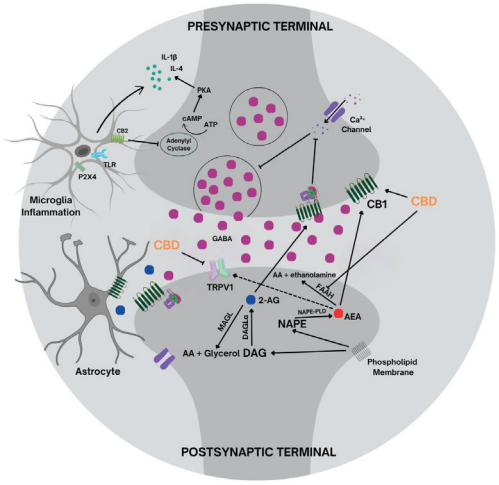


Figure 2: Molecular mechanisms of CBD on cannabinoid signaling pathways in the ECS. These interactions could take place not only at GABAergic synapses but also at other neurotransmitter synapses, such as those involving glutamate. Cannabinoids act on CB1 and CB2 (GPCRs) as well as other non-cannabinoid signaling pathways that are involved in neuroinflammation. CBD promotes the desensitization of TRPV1 channels and the modulation of neurotransmission. Through these interactions, CBD affects neuroinflammatory and excitatory processes associated with epilepsy and MS.

ECS serves as a regulatory system that maintains homeostasis within the body by activating various physiological processes, including pain perception, emotional regulation, appetite control, and memory function.[17] This system operates through endogenous cannabinoids, N-arachidonoylethanolamine (AEA), and 2-arachidonoylglycerol (2-AG), which act as ligands on CB1 and CB2 cannabinoid receptors. CB1 receptors are G protein-coupled receptors (GPCRs) that, for the most part, are located on neuronal terminals of the brain cells, whereas CB2 receptors are primarily associated with the immune system.[18] 2-AG is synthesized from diacylglycerol (DAG) by diacylglycerol lipase-α (DAGLα), and AEA is formed from N-acyl-phosphatidylethanolamine (NAPE) by NAPE-specific phospholipase D (NAPE-PLD). AEA and 2-AG are synthesized depending on membrane phospholipid precursors through activity-dependent activation of specific phospholipase enzymes. 2-AG crosses the postsynaptic membrane and binds to the CB1 receptors located in the presynaptic membrane. Activated CB1 receptors are responsible for the suppression of neurotransmitter release by inhibiting voltage-gated Ca2+ channels, which reduce presynaptic Ca2+ influx, and by coupling to Gi/o proteins.[19] The coupling inhibits adenylyl cyclase, which reduces the production of cyclic adenosine monophosphate (cAMP) and activation of protein kinase A (PKA) subsequently. Reduced PKA activity decreases the phosphorylation of proteins involved in neurotransmitter release. 2-AG is degraded by the monoacylglycerol lipase (MAGL) and, to a lesser extent, by alpha/beta domain-containing hydrolase 6 and 12 (ABHD6,12). On the other hand, AEA mainly activates membrane-bound receptors, including CB1 and transient receptor potential vanilloid 1 (TRPV1).[20]

TRPV1 increases its sensitivity during neuroinflammatory conditions, and CBD acts as a full agonist for TRPV1, desensitizing the channel and reducing neuronal hyperexcitability.[21] The system also involves astrocytes, which are specialized glial cells, as they produce 2-AG when CB1 receptors are activated[22] (Figure 2).

CBD and THC can both alleviate pain, but they do so through distinct mechanisms involving different cannabinoid receptors. THC acts as a partial agonist of CB1 receptors, which inhibits the release of neurotransmitters including glutamate, GABA, acetylcholine, and serotonin (5-HT) when activated. The inhibition causes excessive dopamine release due to a reduction in the release of inhibitory neurotransmitters like GABA in areas such as the ventral tegmental area. This leads to increased dopamine release in the nucleus accumbens, a key reward center in the brain, and induces a high.[23] Likewise, the inhibition of excitatory neurotransmitters like glutamate reduces the transmission of pain signals, indirectly reducing pain. On the contrary, CBD works as a negative allosteric modulator on the CB1 receptor, reducing the euphoric effects of THC.[24] (Figure 2) CBD can alleviate neuropathic pain by indirectly activating CB2 receptors through the elevation of endocannabinoid levels. CB2 receptors are mainly located in immune cells and the spleen. Their expression is limited under normal physiological conditions but increases in response to reactive (proinflammatory) microglia after inflammation or injury. Reactive microglia induce the expression of Toll-like receptors and purinergic P2X4 receptors, subsequently leading to the production of inflammatory cytokines and chemokines that may contribute to neurodegenerative diseases. The stimulation of the CB2 receptors inhibits the neuroinflammatory signaling pathways by a negative feedback mechanism and promotes the return of microglia to a homeostatic, anti-inflammatory state.[25] CB2 receptors can improve inflammation by modulating various immune cells, such as eosinophils, macrophages, neutrophils, and lymphocytes.[26]

Since 2-AG also serves as a major source of arachidonic acid, its breakdown by MAGL links the endocannabinoid system to pro-inflammatory eicosanoid production. Thus, inhibition of MAGL by JZL184 reduces arachidonic acid availability and consequently decreases the synthesis of pro-inflammatory mediators.[27] CBD also competes with AEA to indirectly reduce neuropathic and inflammatory pain by binding to fatty acid amide hydrolase (FAAH), as shown in Figure 2. FAAH is an enzyme responsible for AEA degradation. This results in elevated levels of AEA that could interact with CB2 receptors and enhance the modulation of pain perception and neuroinflammation. Additionally, CBD shows strong anti-inflammatory effects by increasing adenosine signaling responses. This is induced because CBD blocks the equilibrative nucleoside transporter 1, which usually removes adenosine from the extracellular space. By preventing this uptake, CBD increases extracellular adenosine levels, leading to greater activation of A2A adenosine receptors on immune cells and a subsequent reduction in pro-inflammatory mediators such as TNFα.[28] Also, CBD acts as an antagonist of the lipid-activated receptors, GPR55s, which are in both inhibitory and excitatory neurons of the hippocampus. By blocking GPR55, CBD restores the balance between excitation and inhibition by increasing the excitability of inhibitory interneurons and reducing excitatory.[29] The mechanism is especially important in seizure conditions such as Dravet syndrome.

Methods

This review aims to analyze recent clinical trials to assess CBD's potential in treating symptoms of epilepsy, MS, and depression. A literature search was conducted to compile a list of clinical trials where researchers evaluated the efficacy of CBD in these various neurological disorders. A search for relevant literature was conducted in PubMed with a focus on papers published in the last 15 years, by searching the keywords cannabinoids, epilepsy, multiple sclerosis, depression, entourage effect, and clinical. This method yielded 9 relevant papers, and the papers were categorized together based on their relation to similar cases. There are three groups for analysis: multiple sclerosis, epilepsy, and depression. They were research publications freely accessible to the public. Diagrams (Figure 1 and Figure 2) showing the molecular structure of CBD and its mechanisms within the endocannabinoid system were created by the author using Canva to visually provide key signaling pathways. Furthermore, a comparative analysis of similar articles was performed, identifying both consistencies and inconsistencies between the data to examine potential differences in interpretations or methodological limitations. CBD appears to be a promising treatment for epileptic seizures and MS spasticity, as different trials demonstrate its effectiveness in reducing neuroinflammation and activities of pro-inflammatory cytokines such as interleukins (IL-4, IL-5, IL-13, and IL-1β). Although clinical trials are limited, CBD additionally exhibits antidepressant-like effects. Furthermore, this paper explores the boundaries of contemporary research and speculates on the methodological improvements for future studies on CBD.

Results and Discussion

1. CBD and Epilepsy:

Many existing studies have successfully proven the efficacy of CBD in reducing epileptic seizures.[30] Epilepsy is a chronic neurological disorder accompanied by severe seizures. There are different types of seizures, such as tonic (sudden muscle stiffness or tension) and tonic-clonic seizures (a loss of consciousness and muscle contractions). Due to ethical reasons, CBD experiments are typically conducted with subjects with Drug-resistant epilepsies such as Lennox-Gastaut Syndrome, Dravet Syndrome, and Tuberous Sclerosis Complex, who are mostly already treated with a variety of antiepileptic medications. The optimal dosage of CBD varies, considering CBD's interaction with other drugs. Data generated from 39 randomized clinical trials and 13 meta-analyses suggest the optimal dose of purified CBD oral solution to be 20 mg/kg/day in children and adolescents with Drug-resistant epilepsies.[31] A pharmacokinetic study concluded that CBD was shown to reduce seizure frequency in both children and young adults with epilepsy. Key findings included higher doses of CBD resulting in greater seizure reduction, and CBD reached its steady state

DOI: 10.36838/v8i3.15

after approximately 2–6 days of treatment.[32] Although CBD is well tolerated overall, it has adverse events such as diarrhea, somnolence, decreased appetite, and alanine transaminase or aspartate aminotransferase elevation.[33]

Since most human clinical trials testing CBD involve participants whose symptoms are not relieved through current therapy, it is important to understand drug-drug interaction in these subjects and the use of CBD as an adjunctive treatment. Another meta-analysis demonstrated the high efficacy of CBD treatment in reducing seizures, compared with the placebo, 10, 20, and 50 mg/kg/day being the most effective doses.[34] Moreover, the study stated that CBD as adjunctive therapy in the co-administration of clobazam showed a greater reduction in seizure therapy. Clobazam is mostly used to treat LGS and DS. When administered, clobazam is biotransformed into the active metabolite N-desmethylclobazam. The cytochrome CYP2C19 plays a role in the metabolism of N-CLB, transforming it into an inactive compound. CBD inhibits the function of CYP2C19, significantly increasing the concentration of N-CLB in plasma. N-CLB can bind to GABA-A receptors, increasing GABA to reduce seizures.

There are several other conventional anti-seizure medications, including valproate, steroids, ACTH, baclofen, tizanidine, and clonazepam, that could interact with CBD. Co-administration of valproate and CBD could increase liver enzyme levels, requiring monitoring of liver function.[35] A mouse model suggested that CBD may inhibit p-glycoprotein and increase topiramate levels.[36]

Most clinical trials investigating the effects of CBD on epilepsy primarily focus on young infants, children, or young adults, leaving a gap in research on older adult populations. Age differences play a crucial role in CBD's pharmacokinetics and pharmacodynamics. Therefore, understanding how CBD interacts with anti-seizure medications in older adults with Drug-resistant epilepsies remains an important area for future research.

2. CBD and Multiple Sclerosis:

Another major neurological disorder for which CBD shows great potential in treatment is MS spasticity. MS is characterized as a neuroinflammatory disease associated with demyelination and autoimmune responses.[37] There is still a lack of clinical trials testing CBD on patients with MS, and it is crucial to investigate individual variability considering genetic and environmental factors. Nine disease-modifying therapies are available for relapsing-remitting MS, including interferons, glatiramer acetate, teriflunomide, sphingosine 1-phosphate receptor modulators, fumarates, cladribine, and 3 types of monoclonal antibodies. One additional therapy, ocrelizumab, is approved for primary progressive MS.[38] CBD does not treat the disease itself but alleviates spasticity similarly to baclofen, tizanidine, gabapentin, pregabalin, serotonin-noradrenaline reuptake inhibitors, tricyclic antidepressants, dantrolene, and benzodiazepines.[39]

MS spasticity is mostly measured by the spasticity numerical rating scale (NRS), a 0-10 self-reported rating. Modified Ashworth Scale, ranging from 0-4, is also used to evaluate the spasticity of MS patients. Common adverse events of CBD usage in MS are mouth dryness, fatigue, headache, dizziness, loss of appetite, and stomachache.

The ongoing CANSEP trial is the first Canadian randomized clinical trial that investigates the safety and efficacy of CBD in adults with MS. The primary outcomes are measured using the mean Numeric Rating Scale, which scores from 1 to 10, assessing spasticity reduction. Secondary outcomes are measured through various clinical assessments such as the Mean Opinion Score Pain Effects Scale, Modified Ashworth Scale, Pittsburgh Sleep Quality Index, and Epworth Sleepiness Scale. These tests aim to document any AEs experienced by the participants. CANSEP's objective is to compare the distinct and combined effects of THC and CBD, provided in the form of softgels manufactured by PurCann Pharma, by administering varying dosages of each drug.[40]

There have been inconsistent findings of CBD use on multiple sclerosis across various studies. For example, a randomized, double-blind, placebo-controlled trial in Denmark indicated no significant difference between the placebo group and those receiving active treatment with either oral THC (max. Dose 22.5 mg/day) or CBD (max. Dose 45 mg/day) alone or in combination for treating neuropathic pain and spasticity.[41] In this study, cannabis-based medicine was an add-on to patients' ongoing treatment with analgesics and antispastic medicine. However, the study did confirm that the CBD group experienced fewer adverse events compared to the THC treatment groups, and the cannabis-based medicine had a small effect on neuropathic pain and spasticity.

On the contrary, a systematic review article examining the effects of add-on nabiximols, also known as Sativex, consisting of 27 mg THC and 25 mg CBD per ml, has confirmed its efficacy in alleviating spasticity. The review evaluated its efficacy across seven criteria: results from placebo-controlled trials, both short-term and long-term treatment outcomes, gait improvement, the Modified Ashworth Scale, sleep disturbance, the Barthel Activities of daily living, and the Subject Global Impression of Change. Except for the Modified Ashworth Scale and Barthel Activities of Daily Living, significant differences in the response rates were observed, underscoring the efficacy of nabiximols as an adjunctive treatment for symptomatic relief in MS spasticity and other related symptoms such as gait control, sleep disturbance, and a feeling of improvement.[42] Because the Barthel Activities of Daily Living reflects overall functional independence rather than spasticity specifically, improvements in spasticity may not correspond to significant changes in Barthel ADL scores. In patients with multiple sclerosis, motor dysfunction from the underlying disease already influences daily motor function and thereby the Barthel Index, which may explain why several trials reported no improvement or even worsening scores despite treatment.

Nonetheless, both studies have limitations for consideration. The Danish trial indicates that a higher dose of CBD, which could be well tolerated, could show potential in treating spasticity, while the oral formulation of these drugs has low bioavailability. There are several ways to improve the bioavailability of CBD: lipid-based formulations, which are consumed

with fats (sesame oil and coconut butter), polymeric encapsulation of CBD in Poly Lactic-co-Glycolic Acid (PLGA), structural modification of CBD with cyclodextrins, or modified carbohydrates.[43] Future clinical settings testing the efficacy of each method are essential in optimizing CBD formulations. Interestingly, the Danish trial suggested that, except for the THC and CBD treatment for spasticity, the maximal possible effects were clinically insignificant. This is consistent with the observations made by Kleiner *et al*. (2023) as nabiximols contain both THC and CBD. However, the presence of THC was associated with a higher frequency of adverse events among patients.

3. CBD and Depression:

A person is diagnosed with depression when they experience a persistent depressive mood and loss of interest in activities. Patients with depressive disorder take antidepressant medications such as fluoxetine (Prozac), citalopram (Celexa), and escitalopram (Lexapro). Despite the available antidepressant pharmacological options, there are limitations of existing treatments, as sometimes patients develop resistance to drugs like Selective Serotonin Reuptake Inhibitors. Therefore, it is important to explore alternative therapeutic agents that present fewer or no adverse events.

Antidepressant-like effects induced by CBD in the forced swimming test in mice substantiated its dependency on serotonin levels in the central nervous system. The forced swimming test is often used to test the efficacy of antidepressant drugs by measuring the time until the subject becomes immobile, which implies behavioral despair and helplessness. The co-administration of CBD with serotonergic (fluoxetine, FLX) (a type of Selective Serotonin Reuptake inhibitor) resulted in elevated serotonin levels, which correspondingly reduced the duration of immobility observed in the test. CBD's mechanism appears to involve enhancing serotonin neurotransmission and activating postsynaptic 5-HT1A receptors, supported by evidence that serotonin depletion (para-chlorophenylalanine treatment) prevents its antidepressant effects.[44] This data suggests that CBD may offer an alternative to classical antidepressant medication.

One common symptom of depression is anhedonia, or a lack of interest or pleasure in activities that were enjoyed.[45] To evaluate the effects of CBD on depressive-like Wistar-Kyoto (WKY) rats, 30 mg/kg of oral CBD was administered to these animals and they were assessed using Saccharin Preference Test, which measures the animal's preference for saccharin and ability to experience pleasure, and the Novel Object Exploration test which demonstrated increased exploration of the novel object and locomotion at a dosage of 45 mg/kg and increased locomotion at 15 mg/kg. Prohedonic effects were observed in both tests, further suggesting that CBD provides antidepressant-like effects. However, effective doses of CBD in WKY rats varied, reflecting underlying individual variability in the optimal dose required for specific treatment.[46]

One potential explanation of the antidepressant effects of CBD is that CBD interacts with serotonin (5-HT1A) receptors, CB1, G-protein coupled receptor 55 (GPR55), and PPAR-γ. It is essential to recognize the mechanisms behind the antidepressant effects of CBD, as there are few available human clinical trials, and depression is largely dependent on individual differences. A randomized, double-blind, placebo-controlled trial substantiated CBD's antidepressant effects in 31 treatment-resistant young individuals (12-25 years) and found a reduction in mean scores using the Overall Anxiety Severity and Impairment Scale (OASIS).[47] However, contradictory results were seen in bipolar depression by a randomized, double-blind, placebo-controlled pilot study, as there was no significant difference in Montgomery-Asberg Depression Rating Scale (MADRS) scores between the placebo and CBD groups.[48] The inconsistent reports may be partly explained by the different scalings and subtypes of depression implemented by each study. Even though many preclinical studies provided strong evidence for the antidepressant properties of CBD, there are only a handful of clinical studies that explore the primary outcomes of CBD on depression in limited age groups. This points to an urgent need for more studies, especially longitudinal, to allow the development of better treatment strategies for subjects with depression. Additionally, clinical studies on depression advocate higher CBD doses (usually 600 mg/day) as they come with fewer side effects and are well-tolerated.

Another emerging cannabinoid-related therapeutic effect in conditions like depression and anxiety is the "entourage effect." It suggests that the combined presence of terpenes and cannabinoids enhances the overall therapeutic effects beyond what each component could achieve separately. This concept was first proposed by Mechoulam and Ben-Shabat, and this introduced the synergy between cannabinoids and terpenes to maximize pharmacological benefits.[49] Several reviews confirmed this cannabis synergy and supported botanical drug development.[50,51] A study explored the entourage effect by showing that acute *Cannabis sativa* L. leaf extract (THCA, CBCA, β-caryophyllene, α-humulene, limonene, etc.) induced an antidepressant-like effect in depressive rodent models. Furthermore, *Cannabis sativa* L. leaf extract and inflorescence extracts (THCA, CBCA, CBGA, myrcene, limonene, β-caryophyllene, etc.) demonstrated anti-inflammatory effects by decreasing the expression of inflammatory mediators and pro-inflammatory cytokines.[52] Thus, further future clinical studies of the entourage effect are necessary to optimize cannabis appliances containing mixtures of terpenes.

4. Future Directions:

CBD is shown to significantly reduce epileptic seizures in children and young adults and is well-tolerated with mild adverse events such as diarrhea, decreased appetite, and somnolence. Drug-drug interactions are also important in their effectiveness, and seizure frequency seems to decrease even more when CBD is used in combination with clobazam. Thus, it is necessary to explore the positive or negative impact of CBD in combination with other first-line epilepsy medications. Clinical trials examining the use of CBD in elderly populations with epilepsy are still limited and require further attention. Studies investigating the use of cannabinoids in MS

DOI: 10.36838/v8i3.15

have shown inconsistent results: one clinical trial reported a reduction in spasticity frequency in MS, whereas the other found no significant differences between placebo and treatment in either spasticity frequency or overall quality of life. Research design limitations may be responsible for such differences, including the use of inappropriate spasticity scaling methods, ineffective administration methods, or different drug formulations. Despite these inconsistencies, both studies suggested that a combination of CBD and THC is effective in reducing spasticity. However, it is important to note that THC-containing formulations are associated with a greater risk of adverse events. Future research should focus on developing testing methods with higher bioavailability to obtain more accurate results. Clinically defining depression is challenging, and severity depends on individuals; therefore, conducting case studies is important to develop personalized treatments for individual patients. Many trials have recommended higher doses of CBD since CBD is well-tolerated, and these higher doses are considered acceptable in clinical use. However, there is still a limited number of trials in testing CBD in depression, highlighting the need for further clinical research.

The integration of CBD into existing neurological treatment options faces challenges, ranging from legal restrictions to research biases and patient-specific responses. Global cannabis regulations create significant barriers, especially in the United States, where CBD's legality varies between federal and state laws. The 2018 Farm Bill legalized CBD with less than 0.3% THC at the federal level, and the Food and Drug Administration (FDA) has approved only one CBD-based medication, Epidiolex, for epilepsy. State-by-state variations in medicinal cannabis policies further complicate clinical access, restricting patient inclusiveness and physician prescriptions. Internationally, there are still countries permitting medical cannabis under strict guidelines or even completely banning cannabis-derived products.

Furthermore, industry-sponsored studies often structure their design, sample selection, and statistical reporting in ways that favor positive outcomes, raising concerns about bias in clinical findings. Additional confounds can arise from individual variability, including differences between cannabis users and non-users, first-line versus adjunctive CBD therapy, duration of consumption, and age. For instance, individuals with psychiatric conditions may experience greater pre-existing symptoms that could overshadow the effects of CBD. Side effects are also more common in cannabis non-users than in cannabis user groups, as well as in older individuals.[53] Product variability also influences the therapeutic effects of CBD. It was found that about 70% of eighty-four CBD products (from 31 companies) ordered online were found to be mislabeled, reflecting discrepancies between the labeled and actual contents.[54] This can discourage them from purchasing these products freely. Lastly, CBD has a non-selective receptor binding profile, which requires further *in vivo* studies to confirm the exact mechanisms of the action of CBD.

The apprehension surrounding the administration of significant quantities of CBD, along with the possibility of severe adverse events, deters patients from engaging in such new treatment options. For example, CBD may be detrimental to the developing embryo, as demonstrated in studies using zebrafish models of early development.[55] Therefore, it is important to research and list possible adverse events under certain conditions and to develop guidelines for CBD use. In preclinical trials, researchers frequently conduct animal studies, predominantly using mice, to evaluate the effects of CBD. Nonetheless, the discrepancies between animal and human subjects, including variations in CBD dosage and neurological architecture, may influence the outcomes in human populations. Therefore, future investigations must incorporate well-designed clinical trials that assess higher doses of CBD while considering individual parameters. Moreover, further research is needed to fully understand the benefits and potential side effects of combining cannabis-based treatments with conventional pharmacological approaches.

Another emerging frontier in cannabis research is expanding knowledge on the other cannabinoids present in the cannabis plant. For example, CBN, co-produced with THC after decarboxylation of Delta 9 THCA, has been shown to serve as a sleep aid. Furthermore, studies into the interactions between the terpenes derived from cannabis leaves and the main cannabinoid components are important, as they may exhibit synergic effects at reduced dosages and associated risks. To do so, the identification of chemical structures and properties of each cannabis component is vital.

■ Conclusion

This paper presents recent clinical trials that evaluate the therapeutic potential of CBD in individuals with epilepsy, MS, and depression, and demonstrate its effects. As the field of CBD research continues to expand, the proposals presented here aim to guide future studies toward discovering broader therapeutic applications of CBD. CBD is effective in reducing epileptic seizures, and its combination with THC may further reduce MS spasticity. Further personalized research is essential to optimize CBD treatment in individuals with depression. Moreover, there are important factors to consider before conducting clinical trials, such as drug formulation, study population selection, detailed evaluation of each subject, outcome measurements, safety and adverse events monitoring, legal considerations, drug-drug interactions, and bias control. Identification of other cannabinoids and terpenes in cannabis plants is crucial for discovering new medications for neurological disorders.

■ Acknowledgments

I would like to thank my mentor, Charity Bhebhe, for her mentorship and support. I also thank the entire IJHSR team for organizing my research and providing valuable opportunities.

■ References

1. Chayasirisobhon, S. The Role of Cannabidiol in Neurological Disorders. *Perm. J.* **2021**, *25* (2), 1–1. https://doi.org/10.7812/TPP/20.156.

2. Maa, E.; Figi, P. The Case for Medical Marijuana in Epilepsy. *Epilepsia* **2014**, *55* (6), 783–786. https://doi.org/10.1111/epi.12610.
3. Rhee, T. G.; Rosenheck, R. A. Increasing Use of Cannabis for Medical Purposes Among U.S. Residents, 2013–2020. *Am. J. Prev. Med.* **2023**, *65* (3), 528–533. https://doi.org/10.1016/j.amepre.2023.03.005.
4. Lafaye, G.; Karila, L.; Blecha, L.; Benyamina, A. Cannabis, Cannabinoids, and Health. *Dialogues Clin. Neurosci.* **2017**, *19* (3), 309–316. https://doi.org/10.31887/DCNS.2017.19.3/glafaye.
5. Lapierre, É.; Monthony, A. S.; Torkamaneh, D. Genomics-Based Taxonomy to Clarify Cannabis Classification. *Genome* **2023**, *66* (8), 202–211. https://doi.org/10.1139/gen-2023-0005.
6. Bonini, S. A.; Premoli, M.; Tambaro, S.; Kumar, A.; Maccarinelli, G.; Memo, M.; Mastinu, A. Cannabis Sativa: A Comprehensive Ethnopharmacological Review of a Medicinal Plant with a Long History. *J. Ethnopharmacol.* **2018**, *227*, 300–315. https://doi.org/10.1016/j.jep.2018.09.004.
7. Shahbazi, F.; Grandi, V.; Banerjee, A.; Trant, J. F. Cannabinoids and Cannabinoid Receptors: The Story so Far. *iScience* **2020**, *23* (7), 101301. https://doi.org/10.1016/j.isci.2020.101301.
8. Mick, G.; Douek, P. Clinical Benefits and Safety of Medical Cannabis Products: A Narrative Review on Natural Extracts. *Pain Ther.* **2024**, *13* (5), 1063–1094. https://doi.org/10.1007/s40122-024-00643-0.
9. Chrobak, A. A.; Woroń, J.; Siwek, M. Green Rush and Red Warnings: Retrospective Chart Review of Adverse Events of Interactions between Cannabinoids and Psychotropic Drugs. *Front. Pharmacol.* **2024**, *15*, 1500312. https://doi.org/10.3389/fphar.2024.1500312.
10. Duarte, R. A.; Dahmer, S.; Sanguinetti, S. Y.; Forde, G.; Duarte, D. P.; Kobak, L. F. Medical Cannabis for Headache Pain: A Primer for Clinicians. *Curr. Pain Headache Rep.* **2021**, *25* (10), 64. https://doi.org/10.1007/s11916-021-00974-z.
11. Singh, K.; Bhushan, B.; Chanchal, D. K.; Sharma, S. K.; Rani, K.; Yadav, M. K.; Porwal, P.; Kumar, S.; Sharma, A.; Virmani, T.; Kumar, G.; Noman, A. A. Emerging Therapeutic Potential of Cannabidiol (CBD) in Neurological Disorders: A Comprehensive Review. *Behav. Neurol.* **2023**, 1–17. https://doi.org/10.1155/2023/8825358.
12. Brand, E. J.; Zhao, Z. Cannabis in Chinese Medicine: Are Some Traditional Indications Referenced in Ancient Literature Related to Cannabinoids? *Front. Pharmacol.* **2017**, *8*. https://doi.org/10.3389/fphar.2017.00108.
13. Crocq, M.-A. History of Cannabis and the Endocannabinoid System. *Dialogues Clin. Neurosci.* **2020**, *22* (3), 223–228. https://doi.org/10.31887/DCNS.2020.22.3/mcrocq.
14. Pisanti, S.; Bifulco, M. Modern History of Medical Cannabis: From Widespread Use to Prohibitionism and Back. *Trends Pharmacol. Sci.* **2017**, *38* (3), 195–198. https://doi.org/10.1016/j.tips.2016.12.002.
15. Zuardi, A. W. History of Cannabis as a Medicine: A Review. *Rev. Bras. Psiquiatr.* **2006**, *28* (2), 153–157. https://doi.org/10.1590/S1516-44462006000200015.
16. Mechoulam, R.; Gaoni, Y. The Absolute Configuration of Δ1-Tetrahydrocannabinol, the Major Active Constituent of Hashish. *Tetrahedron Lett.* **1967**, *8* (12), 1109–1111. https://doi.org/10.1016/S0040-4039(00)90646-4.
17. Manzanares, J.; Julian, M.; Carrascosa, A. Role of the Cannabinoid System in Pain Control and Therapeutic Implications for the Management of Acute and Chronic Pain Episodes. *Curr. Neuropharmacol.* **2006**, *4* (3), 239–257. https://doi.org/10.2174/157015906778019527.
18. Howlett, A. C.; Abood, M. E. CB 1 and CB 2 Receptor Pharmacology. In *Advances in Pharmacology*; Elsevier, 2017; Vol. 80, pp 169–206. https://doi.org/10.1016/bs.apha.2017.03.007.
19. Howlett, A.; Blume, L.; Dalton, G. CB1 Cannabinoid Receptors and Their Associated Proteins. *Curr. Med. Chem.* **2010**, *17* (14), 1382–1393. https://doi.org/10.2174/092986710790980023.
20. Zou, S.; Kumar, U. Cannabinoid Receptors and the Endocannabinoid System: Signaling and Function in the Central Nervous System. *Int. J. Mol. Sci.* **2018**, *19* (3), 833. https://doi.org/10.3390/ijms19030833.
21. Etemad, L.; Karimi, G.; Alavi, M. S.; Roohbakhsh, A. Pharmacological Effects of Cannabidiol by Transient Receptor Potential Channels. *Life Sci.* **2022**, *300*, 120582. https://doi.org/10.1016/j.lfs.2022.120582.
22. Hegyi, Z.; Oláh, T.; Kőszeghy, Á.; Piscitelli, F.; Holló, K.; Pál, B.; Csernoch, L.; Di Marzo, V.; Antal, M. CB1 Receptor Activation Induces Intracellular Ca2+ Mobilization and 2-Arachidonoylglycerol Release in Rodent Spinal Cord Astrocytes. *Sci. Rep.* **2018**, *8* (1), 10562. https://doi.org/10.1038/s41598-018-28763-6.
23. Bloomfield, M. A. P.; Ashok, A. H.; Volkow, N. D.; Howes, O. D. The Effects of Δ9-Tetrahydrocannabinol on the Dopamine System. *Nature* **2016**, *539* (7629), 369–377. https://doi.org/10.1038/nature20153.
24. Laprairie, R. B.; Bagher, A. M.; Kelly, M. E. M.; Denovan-Wright, E. M. Cannabidiol Is a Negative Allosteric Modulator of the Cannabinoid CB_1 Receptor. *Br. J. Pharmacol.* **2015**, *172* (20), 4790–4805. https://doi.org/10.1111/bph.13250.
25. Bie, B.; Wu, J.; Foss, J. F.; Naguib, M. An Overview of the Cannabinoid Type 2 Receptor System and Its Therapeutic Potential. *Curr. Opin. Anaesthesiol.* **2018**, *31* (4), 407–414. https://doi.org/10.1097/ACO.0000000000000616.
26. Rakotoarivelo, V.; Mayer, T. Z.; Simard, M.; Flamand, N.; Di Marzo, V. The Impact of the CB2 Cannabinoid Receptor in Inflammatory Diseases: An Update. *Molecules* **2024**, *29* (14), 3381. https://doi.org/10.3390/molecules29143381.
27. Pihlaja, R.; Takkinen, J.; Eskola, O.; Vasara, J.; López-Picón, F. R.; Haaparanta-Solin, M.; Rinne, J. O. Monoacylglycerol Lipase Inhibitor JZL184 Reduces Neuroinflammatory Response in APdE9 Mice and in Adult Mouse Glial Cells. *J. Neuroinflammation* **2015**, *12* (1), 81. https://doi.org/10.1186/s12974-015-0305-9.
28. Carrier, E. J.; Auchampach, J. A.; Hillard, C. J. Inhibition of an Equilibrative Nucleoside Transporter by Cannabidiol: A Mechanism of Cannabinoid Immunosuppression. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (20), 7895–7900. https://doi.org/10.1073/pnas.0511232103.
29. Kaplan, J. S.; Stella, N.; Catterall, W. A.; Westenbroek, R. E. Cannabidiol Attenuates Seizures and Social Deficits in a Mouse Model of Dravet Syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (42), 11229–11234. https://doi.org/10.1073/pnas.1711351114.
30. Perucca, E. Cannabinoids in the Treatment of Epilepsy: Hard Evidence at Last? *J. Epilepsy Res.* **2017**, *7* (2), 61–76. https://doi.org/10.14581/jer.17012.
31. Borowicz-Reutt, K.; Czernia, J.; Krawczyk, M. CBD in the Treatment of Epilepsy. *Molecules* **2024**, *29* (9), 1981. https://doi.org/10.3390/molecules29091981.
32. Wheless, J. W.; Dlugos, D.; Miller, I.; Oh, D. A.; Parikh, N.; Phillips, S.; Renfroe, J. B.; Roberts, C. M.; Saeed, I.; Sparagana, S. P.; Yu, J.; Cilio, M. R.; the INS011-14-029 Study Investigators. Pharmacokinetics and Tolerability of Multiple Doses of Pharmaceutical-Grade Synthetic Cannabidiol in Pediatric Patients with Treatment-Resistant Epilepsy. *CNS Drugs* **2019**, *33* (6), 593–604. https://doi.org/10.1007/s40263-019-00624-4.
33. Fazlollahi, A.; Zahmatyar, M.; ZareDini, M.; Golabi, B.; Nejadghaderi, S. A.; Sullman, M. J. M.; Gharagozli, K.; Kolahi, A.-A.; Safiri, S. Adverse Events of Cannabidiol Use in Patients With Epilepsy: A Systematic Review and Meta-Analysis. *JAMA Netw.*

DOI: 10.36838/v8i3.15

Open **2023**, *6* (4), e239126. https://doi.org/10.1001/jamanetworkopen.2023.9126.
34. Talwar, A.; Estes, E.; Aparasu, R.; Reddy, D. S. Clinical Efficacy and Safety of Cannabidiol for Pediatric Refractory Epilepsy Indications: A Systematic Review and Meta-Analysis. *Exp. Neurol.* **2023**, *359*, 114238. https://doi.org/10.1016/j.expneurol.2022.114238.
35. Gilmartin, C. G. S.; Dowd, Z.; Parker, A. P. J.; Harijan, P. Interaction of Cannabidiol with Other Antiseizure Medications: A Narrative Review. *Seizure* **2021**, *86*, 189–196. https://doi.org/10.1016/j.seizure.2020.09.010.
36. Socała, K.; Wyska, E.; Szafarz, M.; Nieoczym, D.; Wlaź, P. Acute Effect of Cannabidiol on the Activity of Various Novel Antiepileptic Drugs in the Maximal Electroshock- and 6 Hz-Induced Seizures in Mice: Pharmacodynamic and Pharmacokinetic Studies. *Neuropharmacology* **2019**, *158*, 107733. https://doi.org/10.1016/j.neuropharm.2019.107733.
37. Tafti, D.; Ehsan, M.; Xixis, K. L. Multiple Sclerosis. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2025.
38. McGinley, M. P.; Goldschmidt, C. H.; Rae-Grant, A. D. Diagnosis and Treatment of Multiple Sclerosis: A Review. *JAMA* **2021**, *325* (8), 765. https://doi.org/10.1001/jama.2020.26858.
39. Beard, S.; Hunn, A.; Wight, J. Treatments for Spasticity and Pain in Multiple Sclerosis: A Systematic Review. *Health Technol. Assess.* **2003**, *7* (40). https://doi.org/10.3310/hta7400.
40. Zertal, A.; Alami Marrouni, K.; Arbour, N.; Jutras-Aswad, D.; Pomey, M.-P.; Rouleau, I.; Prat, A.; Larochelle, C.; Beaulieu, P.; Chamelian, L.; Sylvestre, M.-P.; Morin, D.; Ouellette, J.-S.; Fréjeau, N.; Duquette, P. Efficacy of Cannabinoids Compared to the Current Standard Treatments on Symptom Relief in Persons with Multiple Sclerosis (CANSEP Trial): Study Protocol for a Randomized Clinical Trial. *Front. Neurol.* **2024**, *15*, 1440678. https://doi.org/10.3389/fneur.2024.1440678.
41. Hansen, J. S.; Gustavsen, S.; Roshanisefat, H.; Kant, M.; Biering-Sørensen, F.; Andersen, C.; Olsson, A.; Chow, H. H.; Asgari, N.; Hansen, J. R.; Nielsen, H. H.; Hansen, R. M.; Petersen, T.; Oturai, A. B.; Sellebjerg, F.; Sædder, E. A.; Kasch, H.; Rasmussen, P. V.; Finnerup, N. B.; Svendsen, K. B. Cannabis-Based Medicine for Neuropathic Pain and Spasticity—A Multicenter, Randomized, Double-Blinded, Placebo-Controlled Trial. *Pharmaceuticals* **2023**, *16* (8), 1079. https://doi.org/10.3390/ph16081079.
42. Kleiner, D.; Horváth, I. L.; Bunduc, S.; Gergő, D.; Lugosi, K.; Fehérvári, P.; Hegyi, P.; Csupor, D. Nabiximols Is Efficient as Add-On Treatment for Patients with MultipleSclerosis Spasticity Refractory to Standard Treatment: A SystematicReview and Meta-Analysis of Randomised Clinical Trials. *Curr. Neuropharmacol.* **2023**, *21* (12), 2505–2515. https://doi.org/10.2174/1570159X21666230727094431.
43. Hossain, K. R.; Alghalayini, A.; Valenzuela, S. M. Current Challenges and Opportunities for Improved Cannabidiol Solubility. *Int. J. Mol. Sci.* **2023**, *24* (19), 14514. https://doi.org/10.3390/ijms241914514.
44. Sales, A. J.; Crestani, C. C.; Guimarães, F. S.; Joca, S. R. L. Antidepressant-like Effect Induced by Cannabidiol Is Dependent on Brain Serotonin Levels. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **2018**, *86*, 255–261. https://doi.org/10.1016/j.pnpbp.2018.06.002.
45. Serretti, A. Anhedonia and Depressive Disorders. *Clin. Psychopharmacol. Neurosci.* **2023**, *21* (3), 401–409. https://doi.org/10.9758/cpn.23.1086.
46. Shoval, G.; Shbiro, L.; Hershkovitz, L.; Hazut, N.; Zalsman, G.; Mechoulam, R.; Weller, A. Prohedonic Effect of Cannabidiol in a Rat Model of Depression. *Neuropsychobiology* **2016**, *73* (2), 123–129. https://doi.org/10.1159/000443890.
47. Berger, M.; Li, E.; Rice, S.; Davey, C. G.; Ratheesh, A.; Adams, S.; Jackson, H.; Hetrick, S.; Parker, A.; Spelman, T.; Kevin, R.; McGregor, I. S.; McGorry, P.; Amminger, G. P. Cannabidiol for Treatment-Resistant Anxiety Disorders in Young People: An Open-Label Trial. *J. Clin. Psychiatry* **2022**, *83* (5). https://doi.org/10.4088/JCP.21m14130.
48. Pinto, J. V.; Crippa, J. A. S.; Ceresér, K. M.; Vianna-Sulzbach, M. F.; Silveira Júnior, É. D. M.; Santana Da Rosa, G.; Testa Da Silva, M. G.; Hizo, G. H.; Simão Medeiros, L.; Santana De Oliveira, C. E.; Bristot, G.; Campos, A. C.; Guimarães, F. S.; Hallak, J. E. C.; Zuardi, A. W.; Yatham, L. N.; Kapczinski, F.; Kauer-Sant'Anna, M. Cannabidiol as an Adjunctive Treatment for Acute Bipolar Depression: A Pilot Study: Le Cannabidiol Comme Traitement d'appoint de La Dépression Bipolaire Aiguë : Une Étude Pilote. *Can. J. Psychiatry* **2024**, *69* (4), 242–251. https://doi.org/10.1177/07067437231209650.
49. Ben-Shabat, S.; Fride, E.; Sheskin, T.; Tamiri, T.; Rhee, M.-H.; Vogel, Z.; Bisogno, T.; De Petrocellis, L.; Di Marzo, V.; Mechoulam, R. An Entourage Effect: Inactive Endogenous Fatty Acid Glycerol Esters Enhance 2-Arachidonoyl-Glycerol Cannabinoid Activity. *Eur. J. Pharmacol.* **1998**, *353* (1), 23–31. https://doi.org/10.1016/S0014-2999(98)00392-6.
50. Russo, E. B. The Case for the Entourage Effect and Conventional Breeding of Clinical Cannabis: No "Strain," No Gain. *Front. Plant Sci.* **2019**, *9*, 1969. https://doi.org/10.3389/fpls.2018.01969.
51. Ferber, S. G.; Namdar, D.; Hen-Shoval, D.; Eger, G.; Koltai, H.; Shoval, G.; Shbiro, L.; Weller, A. The "Entourage Effect": Terpenes Coupled with Cannabinoids for the Treatment of Mood Disorders and Anxiety Disorders. *Curr. Neuropharmacol.* **2020**, *18* (2), 87–96. https://doi.org/10.2174/1570159X17666190903103923.
52. Shin, J.; Choi, S.; Park, A. Y.; Ju, S.; Kweon, B.; Kim, D.-U.; Bae, G.-S.; Han, D.; Kwon, E.; Hong, J.; Kim, S. In Vitro and In Vivo Anti-Inflammatory and Antidepressant-like Effects of Cannabis Sativa L. Extracts. *Plants* **2024**, *13* (12), 1619. https://doi.org/10.3390/plants13121619.
53. Binkowska, A. A.; Jakubowska, N.; Redeł, A.; Laskowska, S.; Szlufik, S.; Brzezicka, A. Cannabidiol Usage, Efficacy, and Side Effects: Analyzing the Impact of Health Conditions, Medications, and Cannabis Use in a Cross-Sectional Online Pilot Study. *Front. Psychiatry* **2024**, *15*, 1356009. https://doi.org/10.3389/fpsyt.2024.1356009.
54. Bonn-Miller, M. O.; Loflin, M. J. E.; Thomas, B. F.; Marcu, J. P.; Hyke, T.; Vandrey, R. Labeling Accuracy of Cannabidiol Extracts Sold Online. *JAMA* **2017**, *318* (17), 1708. https://doi.org/10.1001/jama.2017.11909.
55. Li, L.; Fan, B.; Zhang, Y.; Zhao, M.; Kong, Z.; Wang, F.; Li, M. Cannabidiol Exposure during Embryonic Period Caused Serious Malformation in Embryos and Inhibited the Development of Reproductive System in Adult Zebrafish. *Sci. Total Environ.* **2024**, *950*, 175315. https://doi.org/10.1016/j.scitotenv.2024.175315.

Author

Juwon Kim is a senior at Seoul Academy in Seoul, Republic of Korea. She is interested in CBD and its applications for the future of human health. She hopes to pursue a career in scientific research.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

The Effect of Organic Food Consumption on Longevity and Healthspan in *Drosophila melanogaster*

Ian Min

Calvin Manitoba International School, 12 Haesong-ro, Incheon, 22000, Korea; imminain@gmail.com

ABSTRACT: This study investigates the effects of consuming specific organic food items—bananas and sugar—on longevity and healthspan using *Drosophila melanogaster* as a model organism. Contrary to popular belief, fruit flies fed on organic banana pulp showed no significant difference in lifespan compared to those fed on conventional banana pulp. Surprisingly, flies fed conventional banana peels exhibited increased lifespan, possibly due to faster decay and enhanced yeast production. Similarly, experiments with organic sugar showed no lifespan differences. Healthspan metrics, including climbing ability and fecundity, showed no significant improvements with organic food consumption. When given a choice, fruit flies preferred conventional over organic food. These findings challenge assumptions about organic food benefits and emphasize the need for evidence-based evaluation of organic food consumption's impact on health and longevity.

KEYWORDS: Biomedical and Health Sciences, Nutrition and Natural Products, Organic food, Lifespan, *Drosophila*.

Introduction

Recently, as people's interest in health has increased, there has been much attention on organically grown foods. Organic food is defined as produce grown without or with minimal use of artificial chemicals. Non-organic food, on the other hand, is food grown using conventional farming methods that include the use of pesticides, chemical fertilizers, and antibiotics. The debate over the potential benefits of organic food encompasses a variety of issues, including dietary choices, environmental conservation, and food safety.

For starters, several meta-analyses and individual studies have reported that organic foods contain more antioxidants.[1-3] One large meta-analysis published in 2014, for example, found that organic produce had, on average, 20-40% higher levels of antioxidants than non-organic produce.[1] The study synthesized data from 343 existing studies and found that the concentration of antioxidant compounds, particularly polyphenols, tended to be higher in organic foods. This difference may stem from the nature of organic farming methods. Organic farming limits the use of chemical fertilizers and pesticides, and plants can produce more antioxidants by activating their natural defense mechanisms. For example, in non-organic farming, pests and diseases are suppressed with chemicals, whereas in organic farming, these stressors may contribute to the plants' increased production of antioxidants. However, there is controversy over whether differences in antioxidant content translate into actual health benefits, and more research is needed to determine whether higher antioxidant content can cause significant physiological changes in the human body or what the health effects are. So, while the research is consistent that organic foods have higher antioxidant content, more research is still needed to determine the actual health benefits.

In contrast, several studies have shown that there is no significant difference in nutrient content between organic and non-organic foods.[4,5] For example, a recent meta-analysis of 237 articles published over the past 45 years found that while organic foods may contain a slightly higher percentage of antioxidants, as mentioned before, they do not differ significantly in the content of essential nutrients such as vitamins and minerals.[4] These results have been consistently reported in subsequent studies, and the practical impact of nutritional differences between organic and non-organic foods on health appears to be limited.[6,7]

Because organic foods limit the use of chemical fertilizers and synthetic pesticides, they generally have lower levels of pesticide residues than non-organic foods.[8] Some studies have reported that children who consume organic food show significantly lower urinary concentrations of organophosphorus pesticides.[9] However, some argue that pesticides and insecticides have become more regulated and less harmful, suggesting that even pesticide-treated crops are acceptable as long as they are properly washed.[10,11] Research on the long-term health effects of pesticide residue levels is still lacking and needs further investigation.

Potential health benefits of eating organic foods include reduced allergic reactions, increased antioxidant content, and a lower risk of certain diseases, but these are based primarily on observational studies.[4] These benefits have not been consistently shown in rigorous studies, such as randomized controlled trials.[12,13] In addition, the positive health effects of consuming organic foods may be related to improved overall dietary habits or lifestyle changes.

Research on the effects of consuming organic foods on aging is very limited. A recent study reported that fruit flies fed an organic diet lived longer than those on a conventional diet; however, one limitation of this study was that some of the foods provided to the flies, such as potatoes and soybeans, were not part of their natural diet. Although Chhabra *et al*. also included bananas and employed a standardized preparation method by blending and autoclaving the produce with agarose,

DOI: 10.36838/v8i3.23

which effectively minimized confounding effects from decay or fermentation, this study instead used fresh banana pulp and peel to better reflect the flies' natural feeding conditions. Thus, the differences in methodology should be taken into account when comparing the two studies.[14] For example, fruit flies fed conventional soybeans had an average lifespan of only eight days. Therefore, in this study, the effects of organic banana and sugar consumption on longevity and healthspan were investigated in Drosophila to provide a model-based perspective on how specific organic foods influence aging. Fruit flies are an excellent animal model for aging studies, sharing over 70% of the proteins found in humans and exhibiting similar aging mechanisms, but have a shorter lifespan of 60 to 70 days.[15]

Methods

1. Drosophila Stock and Husbandry:

All experiments utilized wild-type Canton-S *Drosophila melanogaster*. Flies were maintained at 25°C and 65% relative humidity, following a 12-hour light and 12-hour dark cycle. To prevent overcrowding during larval development, approximately 150 eggs were seeded into 250 cm^3 fly bottles containing 25–30 mL of culture medium.[16] The standard larval diet consisted of cornmeal-sugar-yeast-agar medium, prepared with 5.2% cornmeal (Hansol Tech, Korea), 11% sugar (Hansol Tech, Korea), 2.5% yeast (Saf-instant, Lesaffre, France), 0.5% propionic acid (Junsei Chemical Co., Ltd., Japan), 0.04% methyl-4-hydroxybenzoate (Yakuri Pure Chemicals Co., Ltd., Japan), and 0.8% agar (Milyang Agar Co., Ltd., Korea).

2. Lifespan Assay:

Lifespan measurements were conducted using three independent cages per condition, each containing 100 flies separated by sex. Organic bananas were sourced from local retailers selling both organic and conventional produce. Bananas were stored in a refrigerator without washing until use, and the pulp and peel were separated with sterile instruments immediately before experimentation to minimize additional contamination. The flies' diet was refreshed with new medium every 2–3 days, during which dead flies were removed and recorded. Organic sugar (CJ, Korea) and conventional sugar (CJ, Korea) were obtained from the same manufacturer to minimize batch variation. The control condition was defined as flies maintained on the standard cornmeal–sugar–yeast–agar diet prepared with conventional sugar, while the experimental condition replaced the sugar component with certified organic sugar. Survival analyses were performed with JMP software (SAS Institute), using both the log-rank test and the Wilcoxon test to assess differences between groups.

3. Bacterial culture:

Microbial samples were harvested from the surfaces of banana peels by sterile swabbing. The bananas used for microbial sampling were the same specimens shown in Figure 3 and were also part of the batches used in the lifespan assays. Cultures were incubated at 37°C under constant shaking at 180 rpm to ensure aerobic growth and uniform bacterial expansion.

4. Measurement of climbing ability:

Males were fed either organic or conventional bananas for 7 days before assessing locomotor function through the rapid iterative negative geotaxis (RING) assay. Ten flies of the same sex were introduced into a climbing device and gently tapped down three times to trigger negative geotaxis. Their positions were recorded 4 seconds after the start of climbing, and the number of flies that surpassed a 4 cm mark was counted. Each experiment included 20 replicates and four independent trials per group, with statistical comparisons made using Student's t-test.

5. Measurement of fecundity:

To assess reproductive output, newly eclosed males and virgin females were collected separately every 3 hours. Virginity was confirmed by the absence of larvae after 24 hours in isolation. On the second day, mating pairs (2 males: 1 virgin female) were established in vials. Every 24 hours, flies were transferred to new vials containing fresh organic or conventional banana food, and the number of eggs laid per female was counted daily for 10 days. Each treatment included 20 replicate vials, and data were analyzed via t-test.

6. Preference test:

A choice assay was conducted by positioning two vials containing either organic or conventional food on opposite sides of a cage. Funnels made of filter paper were inserted into the vials to restrict fly movement to one-way entry. After a 4-hour starvation period, 100 flies were released into the cage, and the number of flies that entered each vial was counted after 30 minutes.

7. Statistical analysis:

Survival analyses were conducted using the standard survival model in JMP software, employing log-rank tests for survival data. Fecundity and locomotion data were analyzed through one-way ANOVA. Statistical significance was denoted in figures by asterisks ($^*p < 0.05$, $^{**}p < 0.001$) when compared to controls.

Results

1. Lifespan changes with organic food consumption:

In the first experiment, the pulp and peel of bananas were separated and provided to the fruit flies. It was deemed meaningful to separate the pulp from the peel, as the pulp is less affected by pesticides and is the part typically consumed by humans. When comparing the lifespan of fruit flies fed organic bananas to those fed conventional bananas, there was no significant difference between the two groups for both females and males. In males, the average lifespan of fruit flies fed organic bananas was 26.23±0.91 days, while that of those fed conventional bananas was 26.45±0.76 days (Figure 1, χ^2=0.0474, p=0.8277). For females, the average lifespan of fruit flies that consumed organic bananas was 37.01±1.54 days, compared to 36.24±1.61 days for those consuming conventional bananas (Figure 1, χ^2=0.0147, p=0.9036).

Unexpectedly, for the peel, the lifespan was longer in the conventional banana group for both males and females. In males, the average lifespan of fruit flies fed organic banana peels was 3.12±0.04 days, while those fed conventional banana peels lived an average of 14.19±0.57 days (Figure 2, χ^2=333.65, p<0.001). In females, the average lifespan of flies fed organic banana peels was 6.9±0.27 days, whereas those fed conventional banana peels lived 24.23±0.51 days (Figure 2, χ^2=272.18, p<0.001), representing a more than threefold difference in lifespan for both males and females.

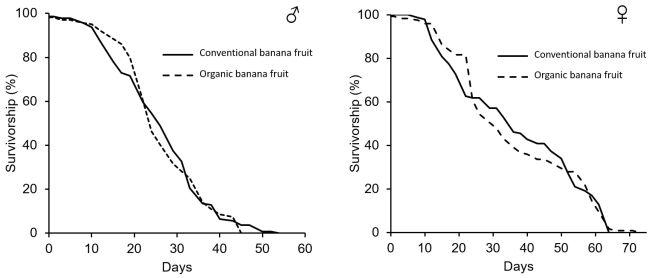


Figure 1: Lifespan of male (A) and female (B) fruit flies fed pulp from organic or conventional bananas. No significant differences were observed between groups in either sex.

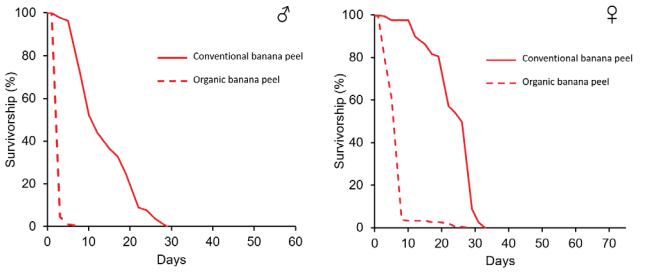


Figure 2: Lifespan of male (A) and female (B) fruit flies fed peel from organic or conventional bananas. The lifespan of flies fed conventional banana peel was longer than that of those fed organic banana peel in both sexes.

2. Difference in decay rate between organic and conventional bananas:

Based on the results above, it can be concluded that the pulp is nutritionally superior to the peel. A decrease in the lifespan of flies fed the peel from the conventional group was anticipated due to the presence of pesticides; however, the opposite result was observed. To investigate these unexpected outcomes, the decay rates of organic and conventional bananas were compared. The results indicated that the peel of conventional bananas decayed relatively faster than that of organic bananas (Figure 3), and the types and amounts of microorganisms cultured from the two groups were significantly different, confirming the nutritional disparity (Figure 4).

Figure 3: Representative images showing decay of organic (upper) and conventional (lower) bananas after 7 days at room temperature.

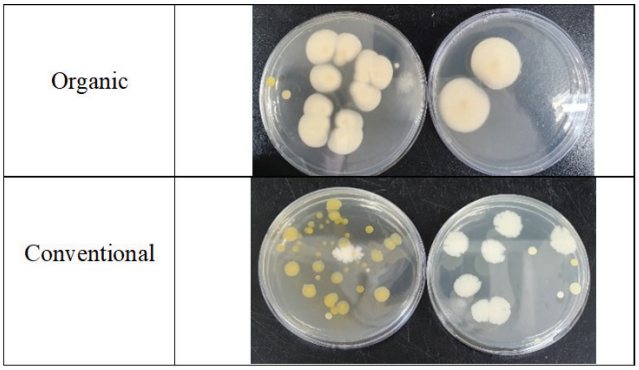


Figure 4: Microbial cultures isolated from the peels of organic (upper) and conventional (lower) bananas.

Since the primary energy source for fruit flies is the yeast produced during decay, conventional bananas likely provide more yeast than organic bananas. To eliminate the variable of nutritional differences between the two groups, the experiment was repeated using organic and conventional sugars—the Drosophila synthetic diet contains both yeast and sugar. Similar to the banana pulp experiment, no lifespan difference was found between the two groups (Figure 5, Male median lifespan, conventional 36.9±0.85 vs organic 37.4±0.91, p<0.5653; Female median lifespan, conventional 37.8±0.95 vs organic 37.2±1.02, p=0.4987). Compared to bananas, which may not provide the same nutrients due to various variables, sugar provides the same nutrients, and experiments using organic sugar yield more reliable results for confirming the hypothesis.

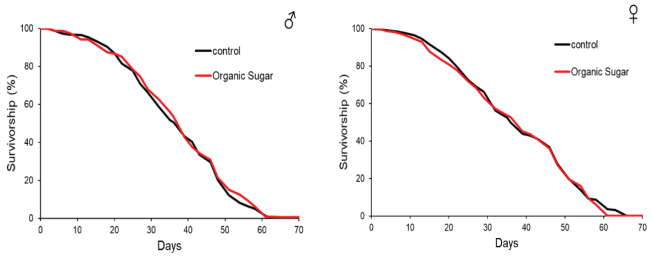


Figure 5: Lifespan of male (A) and female (B) fruit flies fed organic or conventional sugar diets. No significant difference was observed between groups.

DOI: 10.36838/v8i3.23

3. Healthspan changes with organic food consumption:

Fruit flies exhibit negative geotaxis behavior, which is an upward movement against gravity. This behavior decreases with age and is widely used as an indicator of healthy aging in fruit flies.[6,8] It is well known that many interventions that increase longevity also increase climbing ability.[6] Whether fruit fly motility changed with organic food consumption was tested, and no significant differences were found between the two groups (Figure 6).

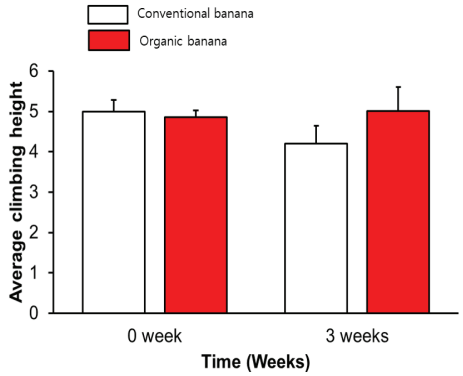


Figure 6: Climbing ability of male fruit flies fed organic or conventional banana pulp, measured at 1 week or after 3 weeks. No significant differences were observed between groups.

Fecundity in fruit flies decreases with age (reproductive senescence), and changes in this metric are commonly used as a marker of healthy aging.[6] When it was examined whether the fertility of fruit flies was affected by organic food consumption, a trend toward a decrease was observed in the group that consumed organic bananas, although this was not statistically significant (Figure 7).

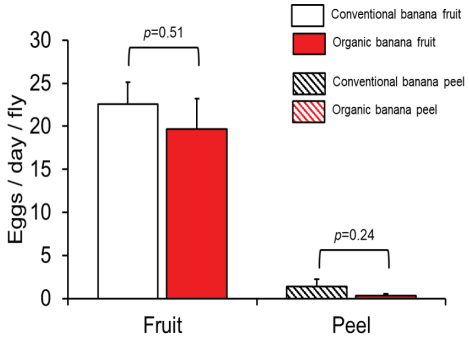


Figure 7: Fecundity of female flies fed organic or conventional banana pulp/peel. Egg production did not differ significantly between groups across 10 days.

4. Preference between organic and conventional bananas:

Human consumers choose food based on a combination of appearance, freshness, nutritional value, and taste. In contrast, animals select what and how much to eat based on a variety of interacting factors, such as nutritional requirements, satiety, reproductive status, food micronutrient composition, and the presence of defensive or toxic chemicals.[17] When fruit flies are given a choice between organic and conventional food, which do they prefer? Do they opt for conventional food due to its faster decay and higher nutritional value (Figure 3), or do they favor organic food because it has fewer preservatives and pesticide residues?

When fruit flies were presented with a choice between organic and conventional food, the majority chose conventional food (Figure 8). It can be concluded that fruit flies are more attracted to the yeast produced by the decay of conventional food than they are deterred by the preservatives or pesticides present in it. To the best of my knowledge, this is the first study to demonstrate that fruit flies prefer conventional bananas over organic ones, highlighting the balance they strike between toxicity and nutritional value in their food choices.

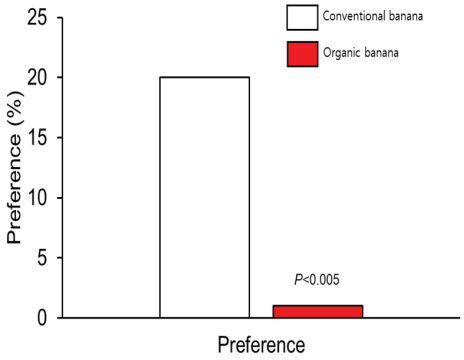


Figure 8: Food preference assay comparing organic and conventional bananas. A greater proportion of fruit flies entered vials containing conventional bananas, indicating a stronger attraction to conventional food sources.

Discussion

Using fruit flies as a model organism, this study aimed to investigate the effects of consuming specific organic foods—bananas and sugar—on lifespan and healthspan. The results provide insights into the impact of organic versus conventional food on the health and lifespan of fruit flies, challenging common assumptions about the benefits of organic food.

Contrary to popular belief, the study found no significant difference in lifespan between fruit flies fed organic banana pulp and those fed conventional banana pulp, suggesting that, at least for fruit flies, consuming organic food does not necessarily lead to increased lifespan. Unexpectedly, flies fed conventional banana peels exhibited a statistically significant increase in lifespan compared to those fed organic banana peels. This counterintuitive finding highlights the complexity of factors influencing lifespan and underscores the need for caution when making broad claims about the health benefits of organic foods.

The difference in lifespan observed in the peel experiment may be attributed to the faster decay of conventional banana peels, which enhances yeast production, a key food source for fruit flies. These findings emphasize the importance of considering both the nutritional content and microbial ecology of foods, rather than solely focusing on the presence or absence of pesticides, when assessing their health impacts. Differences in postharvest quality and decay in organic and conventional fruits may be linked to variations in the microbial community during storage, as demonstrated in a recent study.[18]

While the faster decay of conventional banana peel and subsequent yeast production likely explain the observed lifespan

extension, an alternative explanation should also be considered. Because *Drosophila* are routinely maintained on media containing preservatives such as nipagin (methyl-4-hydroxybenzoate) and propionic acid, flies may have developed long-term adaptation to such exogenous chemicals. This adaptation could potentially mitigate any harmful effects of pesticide residues present in conventional peel, thereby contributing to the improved survival observed in this group. Additionally, it is important to emphasize that banana peels are neither a typical component of the human diet nor a nutritionally adequate food source for fruit flies. The extremely short lifespan observed in the organic peel group (3–6 days) suggests that malnutrition or possible toxic effects, rather than intrinsic differences in peel quality, may have influenced the outcome. Thus, the peel feeding experiment should be interpreted with caution, and further studies using food sources that more closely reflect the natural diet of fruit flies are warranted.

Interestingly, when given a choice, fruit flies preferred conventional food over organic food. This preference may stem from the higher nutritional value of conventional food due to increased yeast production resulting from its faster decay. However, it should be emphasized that the preference assay provides only observational evidence, and the precise cause behind this behavior remains unclear. Because *Drosophila* is highly sensitive to volatile compounds such as ethanol and other fermentation byproducts, factors including scent profiles, microbial composition, and nutritional content may also have influenced the flies' choices. Future studies incorporating chemical analyses and detailed behavioral assays will be necessary to clarify the mechanisms driving this preference.

It is interesting to note that our results differ from those of Chhabra *et al*.[7] The previous study used food purchased from a Whole Foods Market and did not label the origin of the food. On the other hand, this study used bananas imported from the Philippines, and bananas imported into South Korea have strict pesticide standards, so there may be a difference in the amount of pesticides used in the conventional foods in the two studies. Bananas are one of the most consumed and traded fruits in the world, with 114 million tons produced in 2017, and are a highly pesticide-intensive crop. Many countries set maximum residue limits (MRLs) as a way to regulate the use of pesticides, which are typically expressed as the maximum concentration of a pesticide that can remain in food, expressed in mg/kg. These MRL values can vary from country to country, and in some cases can be quite different. For example, the MRL for the pesticide malathion in apples is 0.02 mg/kg in the United States and 8.0 mg/kg in the European Union, about 400 times higher.[19]

While this study provides valuable insights, there are several limitations. First, although fruit flies are a well-established model for studying aging, and many of the findings from fruit flies also apply to mammals[8] remains unclear to what extent these findings can be generalized to humans. Future research should explore similar issues in mammalian models. Second, this study primarily used bananas as the main food source and organic sugar as a secondary source. Expanding the range of foods to include organic yeast could provide a more comprehensive understanding of the impact of consuming organic foods; however, yeast was excluded from this study due to variations in quality and nutritional value among different food companies.

Conclusion

In conclusion, this study emphasizes the importance of an evidence-based approach when evaluating the health benefits of organic foods. While organic farming may offer environmental benefits and reduce pesticide use, its direct impact on longevity and healthspan may be less straightforward than commonly believed. Future research should aim to bridge the gap between these findings in fruit flies and their potential implications for human health.

Acknowledgments

I want to acknowledge Prof. Da-Yeon Shin (Department of Food and Nutrition, Inha University) for her guidance and support.

References

1. Barański, M., Srednicka-Tober, D., Volakakis, N., Seal, C., Sanderson, R., Stewart, G. B., Benbrook, C., Biavati, B., Markellou, E., Giotis, C., Gromadzka-Ostrowska, J., Rembiałkowska, E., Skwarło-Sońta, K., Tahvonen, R., Janovská, D., Niggli, U., Nicot, P., Leifert, C. Higher antioxidant and lower cadmium concentrations and lower incidence of pesticide residues in organically grown crops: a systematic literature review and meta-analyses. *British Journal of Nutrition*. **2014**, *112*(5), 794-811.
2. Brandt, K., Leifert, C., Sanderson, R., Seal, C.J. Agroecosystem management and nutritional quality of plant foods: The case of organic fruits and vegetables. *Critical Reviews in Plant Sciences*. **2011**, *30*(1–2), 177-197.
3. Średnicka-Tober, D., Barański, M., Seal, C.J., Sanderson, R., Benbrook, C., Steinshamn, H., Gromadzka-Ostrowska, J., Rembiałkowska, E., Skwarło-Sońta, K., Eyre, M., Cozzi, G., Krogh Larsen, M., Jordon, T., Niggli, U., Sakowski, T., Calder, P.C., Burdge, G.C., Sotiraki, S., Stefanakis, A., Stergiadis, S., Yolcu, H., Chatzidimitriou, E., Butler, G., Stewart, G., Leifert, C. Higher PUFA and n-3 PUFA, conjugated linoleic acid, α-tocopherol and iron, but lower iodine and selenium concentrations in organic milk: A systematic literature review and meta- and redundancy analyses. *British Journal of Nutrition*. **2016**, *115*(6), 1043-1060.
4. Smith-Spangler, C., Brandeau, M.L., Hunter, G.E., Bavinger, J.C., Pearson, M., Eschbach, P.J., Sundaram, V., Liu, H., Schirmer, P., Stave, C., Olkin, I., Bravata, D.M. Are organic foods safer or healthier than conventional alternatives? A systematic review. *Annals of Internal Medicine*. **2012**, *157*, 348-366.
5. Hunter, D., Foster, M., McArthur, J.O., Ojha, R., Petocz, P., Samman, S. Evaluation of the micronutrient composition of plant foods produced by organic and conventional agricultural methods. *Critical Reviews in Food Science and Nutrition*. **2011**, *51*(6), 571-582.
6. Bourn, D., Prescott, J. A comparison of the nutritional value, sensory qualities, and food safety of organically and conventionally produced foods. *Critical Reviews in Food Science and Nutrition*. **2002**, *42*(1), 1-34.
7. Forman, J., Silverstein, J. Organic foods: Health and environmental advantages and disadvantages. *Pediatrics*. **2012**, *130*(5), e1406-e1415.

DOI: 10.36838/v8i3.23

8. Gomiero, T. Food quality assessment in organic vs. conventional agricultural produce: findings and issues. *Applied Soil Ecology*. **2018**, *123*, 714-728.
9. Lu, C., Toepel, K., Irish, R., Fenske, R.A., Barr, D.B., Bravo, R. Organic diets significantly lower children's dietary exposure to organophosphorus pesticides. *Environmental health perspectives*. **2006**, *114*, 260-263.
10. Reeves, W.R., Bajwa, U., Hayes, W.J. Assessing the safety of pesticides in food: How current regulatory framework limits exposure. *Food and Chemical Toxicology*. **2019**, *134*, 110800.
11. Pretty, J., Sutherland, W.J., Ashby, J., Auburn, J., Baulcombe, D., Bell, M., Bentley, J., Bickersteth, S., Brown, K., Burke, J., Campbell, H., Chen, K., Crowley, E., Crute, I., Dobbelaere, D., Edwards-Jones, G., Funes-Monzote, F., Godfray, H.C.J., Griffon, M., Gypmantisiri, P., Haddad, L., Halavatau, S., Herren, H., Holderness, M., Izac, A.M., Jones, M., Koohafkan, P., Lal, R., Lang, T., McNeely, J., Mueller, A., Nisbett, N., Noble, A., Pingali, P., Pinto, Y., Rabbinge, R., Ravindranath, N.H., Rola, A., Roling, N., Sage, C., Settle, W., Sha, J.M., Shiming, L., Simons, T., Smith, P., Strzepek, K., Swaine, H., Terry, E., Tomich, T.P., Toulmin, C., Trigo, E., Twomlow, S., Vis, J.K., Wilson, J., Pilgrim, S. The top 100 questions of importance to the future of global agriculture. *International Journal of Agricultural Sustainability*. **2010**, *8*(4), 219-236.
12. Dangour, A. D., Dodhia, S. K., Hayter, A., Allen, E., Lock, K., Uauy, R. Nutritional quality of organic foods: a systematic review. *The American Journal of Clinical Nutrition*. **2009**, *90*(3), 680-685.
13. Baranski, M., Rempelos, L., Iversen, P.O., Leifert, C. Effects of organic food consumption on human health; the jury is still out!. *Food and Nutrition Research*. **2017**, *61*, 1287333.
14. Chhabra, R., Kolli, S., Bauer, J. H. Organically grown food provides health benefits to Drosophila melanogaster. *PLoS ONE*. **2013**, *8*, e52988.
15. Piper, M. D. W., Partridge, L. Drosophila as a model for aging. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*. **2018**, 1864, 2707-2717.
16. Klepsatel, P., Procházka, E., Gáliková, M. Crowding of Drosophila larvae affects lifespan and other life-history traits via reduced availability of dietary yeast. *Experimental Gerontology*. **2018**, 110, 298-308.
17. Carvalho, M.J.A., Mirth, C.K. Coordinating morphology with behavior during development: An integrative approach from a fly perspective. *Frontiers in Ecology and Evolution*. **2015**, 3.
18. Gao, Q., Zhang, Y., Gao, C., Li, H., Cheng, Y., Qian, X., Zhang, L., Liu, J., Ogunyemi, S.O., Guan, J. The Microbial Diversity in Relation to Postharvest Quality and Decay: Organic vs. Conventional Pear Fruit. *Foods*. **2023**, 12, 1980.
19. Handford, C.E., Elliott, C.T., Campbell, K. A review of the global pesticide legislation and the scale of challenge in reaching the global harmonization of food safety standards. *Integrative Environmental Assessment and Management*. **2015**, 11, 525-536.

Authors

Ian Min is a high school student with an extraordinary passion for biology. He attends Calvin Manitoba International School in South Korea. Currently, he is working to expand his interests by planning to double major in biology and kinesiology in college as a pathway to achieving his goal of becoming a researcher.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Cracking the Seizure Code: Leveraging Bi-LSTM Models for Neonatal EEG Interpretation and Seizure Classification

Sonali Santhosh

Hanford High School, 450 Hanford St, Richland, Washington, 99354, USA; sonalisanthosh08s@gmail.com

ABSTRACT: This research investigates how Bidirectional Long Short-Term Memory (Bi-LSTM) networks can be used to identify neonatal seizures from EEG data, with an emphasis on mirroring expert neurologist annotations rather than independently predicting seizures themselves. Timely detection of seizures in infants is crucial in Neonatal Intensive Care Units (NICUs); however, traditional methods often delay critical intervention and treatment. Machine learning offers a promising alternative solution. For this study, EEG data from 79 neonates within a publicly available dataset were cleaned, filtered, split into appropriate segments, and processed so the model could learn most effectively. The Bi-LSTM model, trained over eight epochs, achieved an overall accuracy of 83%. A key limitation of this study is the absence of "ground truth," as even expert annotations are subjective, which introduces variation within the training data. Additionally, the overlap between seizure and normal brain activity in the EEG signals contributes to lower model accuracy in certain patient cases. Nevertheless, this study highlights the potential of Bi-LSTM models in enhancing neonatal seizure detection and improving long-term outcomes for vulnerable neonates, offering an essential step toward faster, more effective diagnosis in NICUs.

KEYWORDS: Computational Biology and Bioinformatics, Computational Neuroscience, EEG Pattern Recognition, Neonatal Seizure Detection, and Machine Learning in Neurology.

Introduction

The neonatal period is a critical stage in which the infant's brain is particularly vulnerable. During this period, the brain undergoes rapid changes, characterizing it as a time of high risk for various neurological conditions. One of the most common neurological conditions is seizures, described by Huff & Murr as changes in the degree of "consciousness, behavior, memory, or feelings." Seizures are often caused by atypical, unconstrained electrical activity in the brain.[1] The prevalence of neonatal seizures in the general population is roughly 1.5%, with the overall occurrence being three per every 1000 live births. However, given the chance, a newborn is born prematurely, prevalence increases to around 57–132 per 1000 live births.[2] The most common cause of neonatal seizures is hypoxic-ischemic encephalopathy (HIE), a type of brain damage resulting from the lack of oxygen to the brain soon after delivery. Although they can also be caused by a variety of other factors, such as stroke, cranial blood clots, and other brain defects.[3] Not all seizures tend to have long-lasting effects since many are brief and only momentary. Regardless, prolonged neonatal seizures can lead to permanent brain damage, especially if they are not detected early enough.[4]

Currently, expert neurologists diagnose seizures in neonates by analyzing the patients' electroencephalogram (EEG) data, which measures the general electrical activity in various regions of the brain. Most often, however, they require specialized experts like pediatric neurologists to be able to diagnose neonates with higher efficacy. This results from the unique pathophysiology and electrographic findings of infants' EEG data, causing it to be more difficult to identify their seizures specifically. Most frequently, newborns do not show similar visual signs of seizures to adult patients, such as full-body convulsions. Their "symptoms" often appear to look like normal baby behavior, for instance, thrashing legs or random eye movements, along with a sucking tongue. Additionally, newborns may even have symptoms, such as jitteriness or sleep myoclonus, that tend to mimic seizure activity, even in the absence of genuine seizures.[5] Due to the lack of clear observable clinical manifestations, physicians must rely on more advanced brain monitoring techniques to accurately diagnose the presence of seizure activity in newborns. The most preferred method is continuous video EEG monitoring, allowing clinicians to determine if unusual features are seizures.[6] Once a seizure is detected or simply suspected, physicians may perform laboratory testing and other imaging modalities such as head ultrasonography and magnetic resonance imaging (MRI) to confirm potential causes before treatment. If continuous video EEG monitoring is inaccessible, practitioners may turn to amplitude-integrated EEG (aEEG). Amplitude-integrated EEG, described as being able to "present time-compressed and filtered EEG data," offers an alternative method for areas lacking the support that continuous video EEG monitoring requires.[5] However, aEEG does have a low sensitivity, which can lead to false negatives in the diagnosis, potentially allowing seizure activity to go undetected. Unfortunately, the entire process of traditional detection is costly, time-consuming, and does not guarantee full accuracy as annotations vary from professional to professional. Moreover, if a neonate is not diagnosed and treated in a timely fashion, the damage to the developing brain could cause "cognitive disorders, developmental delay, epilepsy, or cerebral palsy," displaying the crucial need for rapid diagnosis.[7] A faster, reliable

DOI: 10.36838/v8i3.29

diagnostic method would significantly benefit the medical specialty of neonatology.

A rising consideration in medicine is the use of Machine Learning (ML). Machine Learning allows for diagnosis beyond what the naked eye can detect and is anticipated to improve many areas of medicine, including accuracy, prediction methods, and quality of patient care.[8] At the moment, ML positions itself in a more supportive role in healthcare, easing the workload of physicians. In large facilities, machine learning techniques have been implemented in a multitude of ways, such as in record organization, medical imaging, and robot-assisted surgeries.[8] Countless studies have worked with EEG for seizure detection, leading to the possibility of earlier and more accurate seizure detection. However, implementing those computational methods on patients in real clinical settings will require further development as ML continues to evolve. Nevertheless, once it does, patients with neurological conditions, such as epilepsy, will receive significantly improved outcomes.

Neural networks, a subset of ML, have been well-studied for their use in the medical field.[9] Recently, with the increasing availability of public EEG data, many types of Neural Networks have been used, particularly to detect seizures. Recurrent Neural Networks (RNNs) deal with sequential, time-evolving data, which makes them well-suited for analyzing time-series data such as EEG signals, in which the order of information is essential. However, there is a lack of research on investigating RNNs' effectiveness in neonatal seizure detection. Additionally, this research solely relies on the key assumption that by treating annotations from human expert neurologists as the "ground truth", we are accurately detecting seizures within all of these infants. Building on the strengths of RNNs, the model essentially functions as a universal classifier, designed to generalize across multiple patients rather than being specialized for a specific neonate. This approach allows the model to learn and understand patterns that apply to a wider range of EEG data, making it useful for diverse clinical settings. This universal classification approach ensures that the model is accurate across varying cases and not limited to a single patient's data.

In this study, Recurrent Neural Networks (RNNs) were proposed for their ability to process sequential data and capture temporal dependencies, such as those of EEG data, specifically Bidirectional Long Short-Term Memory Networks (Bi-LSTMs), a certain type of RNN. Although traditional RNNs function as a valuable model in ML, they frequently experience a certain challenge known as the vanishing gradient problem. This occurs during backpropagation when the gradient becomes increasingly small until it ultimately "vanishes." Normal LSTMs overcome this issue by their additive update mechanism, with the use of gates. Specifically, the LSTM architecture is constructed of the following parts: a cell, an input gate, a forget gate, and an output gate. The forget gate is what allows the model to reset its state.[10] And the Bidirectional portion allows the LSTM to view data from both forward and backward time standpoints. Thus, this study intends to address the following research question: To what degree can a Bi-LSTM model accurately capture and interpret complex associations between neonatal EEG data and clinician-labeled annotations, and how accurately can it classify seizure events across diverse patient data?

Methods

Dataset:

The dataset chosen required high-quality data, corresponding annotations, and a large diversity. This was found in a publicly available dataset labeled "A Dataset of Neonatal EEG Recordings with Seizure Annotations."[11] The dataset has EDF files available for multi-channel EEG recordings for seventy-nine term neonates, a MAT file for the visual interpretations of the data, and a CSV file with the corresponding clinical information for each patient. The data was recorded from the Neonatal Intensive Care Unit at the Helsinki University Hospital, Finland, one of Europe's largest healthcare providers. Out of the seventy-nine neonates, seizure consensus variability was visible with 39 infants labeled as seizure-prone and 22 labeled as seizure-free. The interquartile range for the EEG recordings was broad, from 64 to 96 minutes, with a median recording duration of 74 minutes, and the signals themselves were sampled at 256 Hz. The EEG recordings' visual interpretations were annotated independently by three expert neurologists and stored in a MAT file for further use. The specialists defined seizure activity as visible on the EEG recordings when there was an "emergence of abnormal discharges in bursts, termed ictal epileptiform discharges… Escalate in frequency, evolving into rapid, continuous spikes and waves, and ultimately peak with numerous spikes accompanied by buried waves."[12] Annotations from the three experts, however, are also accessible in individual CSV files, categorized as 'A', 'B', and 'C'.

Data Preprocessing:

Several preprocessing steps were carried out to prepare data for input into the model, including bad signal removal, band-pass filtering, segmentation in fixed-length epochs, feature selection, and standardization. The EEG recording dataset has 21 total channels, with the bipolar montage displayed in Figure 1; however, two of those channels are Electrocardiogram (ECG) and Respiratory Effort Channels. Unnecessary noise, spike waves, and artifacts caused by the extra channels may corrupt training, so to prevent contamination of the pure EEG data, removal of the ECG and Resp-Effort Channels was required. The raw continuous EEG data is then band-pass filtered from 0.1 to 15 Hz due to the range capturing the relevant brainwave frequencies associated with neonatal seizures while minimizing noise. Afterward, the neonatal EEG data is segmented into epochs of one second each, corresponding with the annotations per patient.

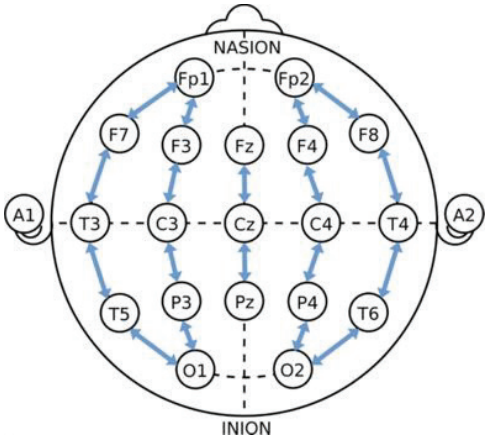


Figure 1: The bipolar montage represents the 21 EEG electrodes used in collecting the neonatal data used in the model. This montage follows the standard 10-20 system used globally for electrode placement, certifying comprehensive data collection. The recorded data was later processed and analyzed to train the Bi-LSTM efficiently for neonatal seizure detection.

$$Mean: E(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$Variance: Var(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - E(x))^2$$

$$Skewness: S(x) = E\left[\left(\frac{x - E(x)}{\sqrt{Var(x)}}\right)\right]^3$$

$$Kurtosis: K(x) = E\left[\left(\frac{x - E(x)}{\sqrt{Var(x)}}\right)\right]^4$$

Figure 2: The statistical features breakdown displays the mean, E(x), which refers to the average of all n data points collected, and the variance, Var(x), describing the average squared distance from the mean, or in simpler terms, the standard deviation of the data, squared. Skewness, S(x), indicates the measure of the asymmetry of the expected value (E) in the classic distribution of the data, and kurtosis, K(x), measures the "tailedness" or the sharpness of the peak in the data distribution, indicating how much of the variance in the data is due to extreme values.[13]

Statistical features regarding mean, variance, skewness, and kurtosis were extracted from the data (Figure 2). In addition to statistical features, time-domain features (TDF) and frequency-domain features (FDF) were also extracted. Within the TDF, zero crossing and peak-to-peak features were combined into a vector for each epoch. Zero Crossing calculates the number of times the EEG signal waves' amplitude values cross the zero-amplitude level.[14] Peak-to-peak measures the difference between the maximum and minimum value of the EEG signals, showcasing the range or variation within the signals. Referring to the frequency domain, Power Spectral Density (PSD) was calculated. PSD measures the "signal's power content versus frequency," applied through Welch's method.[14] The following three feature vectors, statistical features, time-domain features, and frequency-domain features, were combined, leading to one large feature matrix, used as the input for the model. The data was then split at 70:30; 70% of the data was used for training (74,276 training samples), while the other 30% was used for testing and validation (31,833 testing samples), as seen in the input shapes for the training and testing datasets. Lastly, the complete feature set data was standardized post-split for further accuracy. Standardization is where each data value is scaled to be centered around the mean with a standard deviation of one. This allows for increased model reliability by allowing it to learn necessary patterns in the data more effectively.

Since the EEG data was annotated by three separate neurologists, to improve results, the 3 CSV files were first imported as arrays and then combined into a matrix based on majority voting. Hence, a value is only added as '1' if two or more annotators marked that epoch as seizure activity. Otherwise, the value is added as '0' or normal brain activity. This allows for more consistency by considering consensus between multiple neurologists. Neonates that did not contain any positive '1' values were then removed from the training set.

Model Architecture:

The model was built using the Keras Sequential Model, which is beneficial for efficiently stacking multiple layers.[15] The architecture of this model was adapted from a study by Zeedan et al., with significant adaptations to better suit the feature selection requirements of this research.[16] Two Bidirectional Long Short-Term Memory (Bi-LSTM) layers were included with 64 units each to capture the necessary temporal dependencies. A Bi-LSTM creates an additional layer of the reverse structure to produce more efficient information than traditional LSTM-based models or even basic RNNs.[17] The Bi-LSTM layers used activation functions of hyperbolic tangent or TanH, which are commonly implemented for RNNs. Afterwards, two Dense layers were added with 32 and one unit(s) respectively for binary classification, each with Rectified Linear or ReLu activation functions. Several regularization techniques were used to prevent overfitting, including Dropout, implemented at a rate of 0.1, allowing the network to be readjusted with an alternate group of neurons discarded for every training sample. Aside from Dropout, Batch Normalization, and L2 Regularization were also implemented to prevent overfitting (Figure 3).

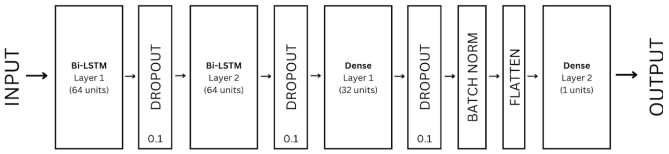


Figure 3: The following visual represents the architecture that the model utilized. The first layer depicts a Bi-LSTM layer of 64 units, in which the neural network receives its input matrix of (74276, 1, 2489), applies its transformations, and releases its output to the next layer, which in this case is a Dropout Layer. This layer effectively releases 10% of its neurons to increase variability within the model's training, before inputting it into another Bi-LSTM layer, where this cycle repeats once more. The data then enters a Dense layer with 32 units for further data processing before reaching the respective Dropout Layer. In the final stretch, Batch Normalization increases the stability of the activation functions by normalizing the data, and the following Flatten Layer converts the multi-dimensional output into a single vector. Lastly, the data is input into a Dense Layer with one sole unit, allowing the model to binarily classify the output as seizure or non-seizure data.

DOI: 10.36838/v8i3.29

The ADAM (Adaptive Moment Estimation) optimizer was used because of its fast convergence and is widely accepted as the best choice for optimizers in similar problems. The ADAM optimizer is set with a learning rate of 0.01. Most often, ADAM is used with a learning rate of either 0.01 or 0.001; however, a learning rate of 0.001 is better suited for image classification tasks rather than temporal features in our case.[18] The model was then trained with 8 epochs using the Binary Cross-Entropy loss function, most widely accepted for binary classification tasks. To handle the clear class imbalance present in the data, with most epochs having an output of 0, or normal brain activity, in all the neonates present in the data, the Synthetic Minority Oversampling Technique (SMOTE) was implemented to create synthetic data points for the minority class. In this case, positive '1' values indicating seizure activity were boosted through SMOTE to help balance the distribution of the data across the training sets, allowing the model to learn the patterns of both classes equally.

Evaluation Metrics:

The model's overall performance was evaluated using a variety of metrics. A confusion matrix calculates the False Negative, True Negative, False Positive, and False Negative values, along with a traditional classification report with the performance measurement tools displayed for the F1-Score, Recall, Precision, and Accuracy. Lastly, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) was calculated to understand the correlation between the True Positive Rate and False Positive Rate.

Results and Discussion

The resulting Bi-LSTM model was able to achieve accuracy rates of individual patients in a range from as low as 56% to as high as 93%. Overall, the model obtained an accuracy of 83% across the testing dataset of all 79 individuals. Although it was able to accurately classify 91% of neonates with a lack of seizure activity as negative, it was only able to diagnose 53% of neonates having a seizure as positive. These results are shown in the Confusion Matrix presented in Figure 4. As indicated, the Bi-LSTM tended to perform better with negative cases. The model had a Precision rate of 88%, a Recall of 90%, and an F1-score of 89% for predicting lack of seizure activity. However, the model only had a Precision rate of 58%, a Recall of 53%, and an F1-score of 55% for predicting the presence of seizure activity in neonates. This explains that the model has a high specificity but a low sensitivity.

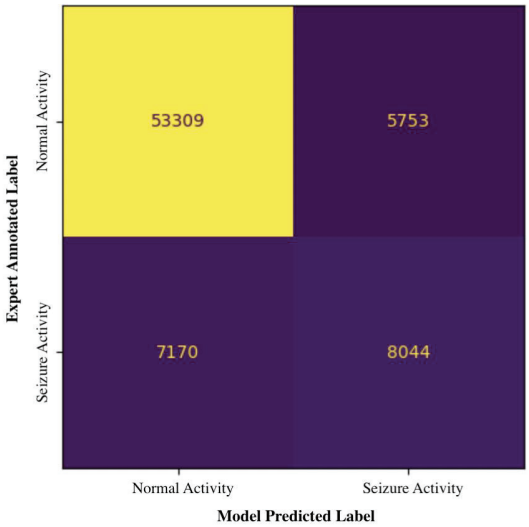


Figure 4: Each quarter of the confusion matrix displays a different outcome. The bottom left corner shows the false negative rate, or when the model misinterprets seizure activity as normal brain activity. The bottom right corner shows the true positive rate when the model accurately predicts seizure activity. The top left corner shows the true negative rate when the model accurately predicts normal brain activity. Lastly, the top right corner shows the false positive rate, or when the model misinterprets normal brain activity as seizure activity.

During the model's training process, the loss and accuracy graphs (Figures 5 and 6) suggest strong overfitting. In Figure 5, the validation loss is shown not to decrease significantly after running through the epochs, while the training loss continues to drop. Figure 6 supports this by showing oscillations in validation accuracy while training accuracy is increasing steadily throughout the entire training process. This gap between training and testing performance suggests that the Bi-LSTM was becoming overly specialized to the training data, even with techniques set in place attempting to reduce overfitting, such as Dropout, Batch Normalization, and L2 regularization. Thus, this reduced its ability to generalize to new, unseen cases. These observations align with the model's habit of performing well on negative cases while struggling with positive cases.

Figure 5: Displays the training vs. validation loss across multiple epochs. Training loss is depicted as consistently decreasing toward zero. Validation loss, on the other hand, plateaus around 0.42 to 0.44, indicating clear overfitting within the Bi-LSTM model.

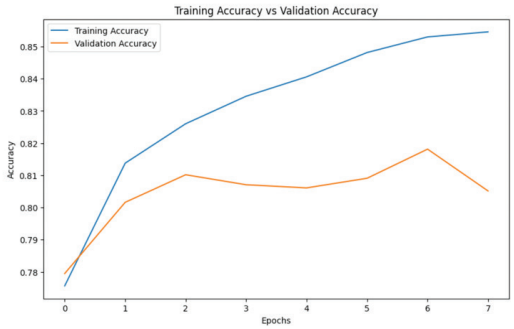


Figure 6: Displays the training vs. validation accuracy across multiple epochs. Training loss is depicted as consistently increasing toward one, while validation accuracy fluctuates near 0.81, indicating the model is becoming overly specialized to the training data itself.

Lastly, both the Confusion Matrix (Figure 4) and ROC curve (Figure 7) highlight the model's ability to reduce False Negatives while emphasizing the need for clearer detection of True Positive events. The Area Under the Curve (AUC) score is 0.808, which, although it is a significant jump from random prediction at 0.5 AUC, still requires large model improvement necessary in determining differences between seizure versus non-seizure EEG data.

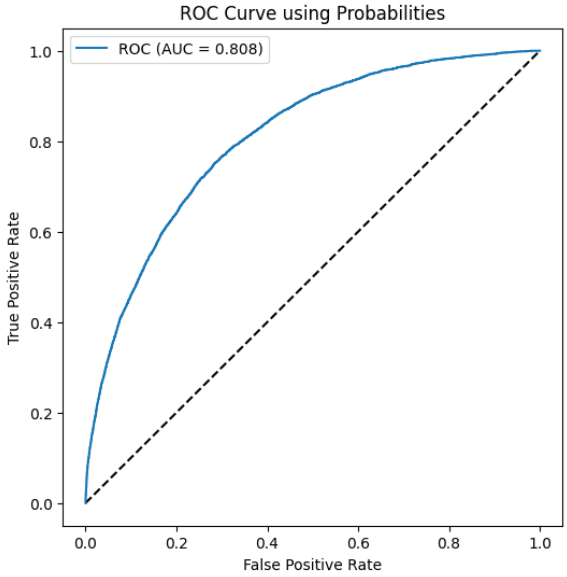


Figure 7: Represents the correlation between the false positive and true positive rate of the Bi-LSTM model through the receiver operating characteristic (ROC) curve.

Seizure vs. Non-Seizure Epochs:

By comparing EEG data examples from both groups, we can better see how the model decides to make its choices. Figures 7 and 8 display a side-by-side comparison of correctly classified seizure and non-seizure events, revealing the brain wave features that the Bi-LSTM was able to accurately understand.

In Figure 7, the non-seizure epochs display consistent brain wave patterns with limited spikes or sudden shifts. On the other hand, Figure 8, which shows seizure epochs, has chaotic activity, sharp peaks, and large amplitude changes that aren't seen often in non-seizure states. These differences likely influenced how the model made its predictions. However, as observed, the distinction between these two classes isn't as clear throughout all the infants, explaining why the accuracy varied considerably. Some cases had overlapping patterns, which could have led to the model missing certain seizures. To improve detection, further improvements to the model are needed, especially for less clear seizure cases.

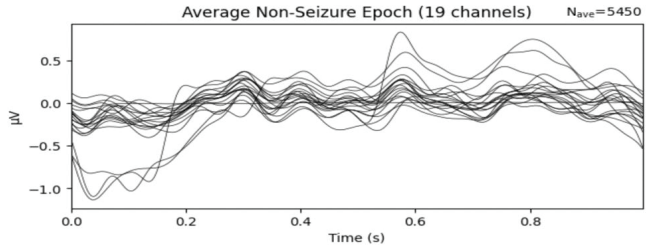


Figure 8: Illustrates the average EEG epoch across nineteen channels that were identified as normal brain activity for a single patient. Brain waves are characterized by minimal spikes or abrupt shifts, indicating the absence of seizures within the segment of data.

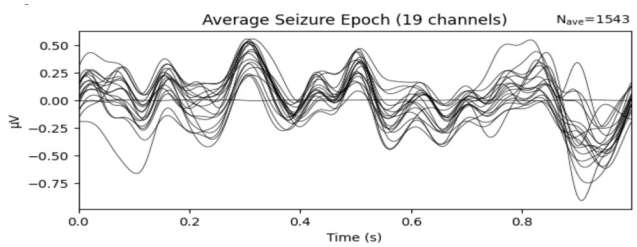


Figure 9 : Showcases the average EEG epoch across nineteen channels that were identified as seizure activity for a single patient. Visually depicts key features that distinguish the data from a normal EEG segment to both the Bi-LSTM model and neurologists.

Discussion:

The Bi-LSTM achieved an overall accuracy of 83%, demonstrating the model's ability to distinguish between normal and seizure epochs. The model's sensitivity refers to the number of correctly diagnosed True Positive events, or accurately diagnosed seizure events, over the total amount of seizure events in the EEG data. The specificity of this model refers to the number of correctly diagnosed True Negative events, or accurately classified non-seizure events, over the total amount of regular events in the EEG data. One of the key strengths of this model was its high specificity. The model was able to accurately diagnose 90% of non-seizure epochs, demonstrating the model's reliability in classifying normal cases without the presence of any seizure activity. However, the model did lack sensitivity. The Bi-LSTM model was only able to accurately diagnose 53% of seizure epochs, as reflected in the confusion matrix (Figure 4). This distinction indicates that, while the model performs well in detecting normal brain activity in neonatal patients, it may overlook more subtle seizure patterns, resulting in the lower sensitivity rates presented. The resulting false negatives are particularly concerning with infants, however, since missed seizures could result in delayed treatment options. This issue emphasizes the need for increased sensitivity, even if it comes at the cost of a lower specificity.

DOI: 10.36838/v8i3.29

Limitations:

These complications can be caused by the various limitations present in the data. First of all, even the medical professionals who annotated the dataset struggled with determining a positive case. The three annotators all had varying annotations with each patient within the dataset. Some would consider one epoch a sign of key seizure activity, while others would label the same epoch as a negative case. Attempts to combat this issue were made by combining the three neurologists' annotations and using the majority rule to determine the information used in the model. However, this did not guarantee that the model was tested on true information, as although the majority rule relies upon consensus, one cannot state that the consensus was 100% valid. The accuracy of the model depends on the accuracy of the expert's annotations, which will always remain uncertain. Some results that were analyzed as False Positives could indeed mean the patient has symptoms of seizure activity, and the experts had just overlooked it in the analysis. Nonetheless, there is no certain way to fully know whether these cases were truly accurate or not. Thus, this model does not predict the onset of neonatal seizures but showcases how close artificial intelligence can come to predicting the methods behind human neurologists' annotation of neonatal seizures through the visual analysis of EEG data. Each percentage of accuracy indicates how similar the model came to predicting seizures in the way human experts can.

Second, visual comparisons of the classified positive and negative epochs (Figures 7 and 8) reveal key characteristics that may have guided the model's predictions. Commonly, non-seizure epochs exhibited regular-appearing EEG activity, with no apparent erratic spikes and fluctuations seen in the seizure epochs.

However, the overlap between normal and seizure-like activity in certain patient cases very likely contributed to the Bi-LSTM's lower sensitivity. This is shown in the performance of the model, which varies significantly across individuals, with accuracy having an overall range of 56% to 93% across all 79 patients. Differences in EEG signal quality could have also caused this variation, although the distinctiveness of seizure patterns across patients in the lower range of accuracy certainly impacted the model's apparent performance. Neonates with clearer seizure patterns were more easily classified, while those with larger, ambiguous activity led to a lower sensitivity of the model.

Third, analysis of the training process revealed strong signs of overfitting, as indicated by the loss graph in Figure 5. The validation loss oscillated frequently, never truly declining the way the training loss was able to. This could have been averted by running a larger number of epochs; however, the size of the data led to numerous epochs becoming computationally prohibitive. This caused the model to be overly specialized to the training data and struggle with unseen cases in the testing data.

Future Work:

Significant improvements are required to improve the model's sensitivity to neonatal seizure data. To enhance the model's performance and increase its clinical applicability, several key areas for future study can be explored. First of all, using advanced feature extraction techniques may provide more detailed signals for seizure detection. Additionally, optimizing the model architecture, possibly by using a more complex hybrid-based (Conv-LSTM) model, may provide more accurate results. Lastly, during preprocessing, it may be beneficial to analyze patient data separately and feed it through the model individually, rather than combining multiple patients' data as done in this study. These improvements could allow for increased model accuracy and more personalized neonatal seizure classification, resulting in better clinical outcomes.

Conclusion

A Bidirectional Long Short-Term Memory (Bi-LSTM) model was utilized in this study to analyze its usefulness in predicting neonatal seizures relative to human experts. The Bi-LSTM model was selected because of its effectiveness in working with time series data, capturing both past and future information simultaneously, without running into challenges often seen with basic Recurrent Neural Networks. Three feature sets were extracted from the Helsinki dataset, involving EEG data from 79 term infants, including statistical features, time-domain features, and frequency-domain features. The corresponding seizure annotations from three expert neurologists were combined based on majority rule and fed into the model. The Bi-LSTM model achieved an overall accuracy as a universal classifier for all 79 neonates of 83%.

Although the model was unable to predict the majority of seizure-activity instances, this research still represents a valuable step toward improving neonatal care worldwide. Enhancing early seizure detection, even with its valid limitations, could facilitate earlier interventions in many critical cases, considering the model does indeed have a higher True Positive rate than a False Positive Rate. Even now, it is difficult for less-equipped healthcare facilities to attain the proper assistance and tools required for suitable infant care, having to travel large distances to larger hospitals during critical times. With the use of AI, these hospitals may have an opportunity for rapid detection before they can receive appropriate care from medical professionals. That, in turn, has the potential to reduce the severity and long-term neurological effects commonly associated with neonatal seizures. Further improvements in this model could enhance its sensitivity, offering even greater benefits to the developmental outcomes and quality of life of the affected newborns.

Thus, this study contributes to the ever-growing body of knowledge on neonatal care and seizure detection by leveraging advanced recurrent neural network techniques. These findings demonstrate the potential to enhance the timeliness and accuracy of neonatal seizure detection. With further refinement and validation, this research approach holds promise for future clinical applications, potentially improving outcomes for newborns at risk of seizures. Combining artificial intelligence's rapid computational power with clinicians' empathy and observational expertise projects a revolution in the future of neonatal patient care.

■ Acknowledgments

I would like to express my sincere gratitude to David DiStefano for their invaluable guidance and support throughout this research. Their insights and expertise greatly contributed to the development of this project, and their encouragement was instrumental in bringing this work to completion.

■ References

1. Huff, J. S.; Murr, N. I. Seizure. In *StatPearls [Internet]*; StatPearls Publishing: Treasure Island (FL), 2024. https://www.ncbi.nlm.nih.gov/books/NBK430765/.
2. Panayiotopoulos, C. P. *The Epilepsies: Seizures, Syndromes and Management*; Bladon Medical Publishing: Oxfordshire, U.K., 2005; Chapter 5, Neonatal Seizures and Neonatal Syndromes. Available at https://www.ncbi.nlm.nih.gov/books/NBK2599/.
3. Krawiec, C.; Muzio, M. R. Neonatal Seizure. In *StatPearls [Internet]*; StatPearls Publishing: Treasure Island, FL, 2024; Updated 2023, Jan 2. Available at https://www.ncbi.nlm.nih.gov/books/NBK554535/.
4. Neonatal (Newborn) Seizures: Conditions: UCSF Benioff Children's Hospitals. https://www.ucsfbenioffchildrens.org/conditions/neonatal-seizures#:~:text=Some%20neonatal%20seizures%20are%20mild,and%20excessive%20brain%20cell%20activity (accessed 2024-07-12).
5. Stieren, E. S.; Rottkamp, C. A.; Brooks-Kayal, A. R. Neonatal Seizures. *NeoReviews* **2024**, *25* (6). DOI:10.1542/neo.25-6-e338.
6. Kanner, A. M.; Wirrell, E. Video EEG Test: Diagnosing Seizures. https://www.epilepsy.com/diagnosis/eeg/video-eeg (accessed 2024-07-24).
7. Kaminiów, K.; Kozak, S.; Paprocka, J. Neonatal Seizures Revisited. *Children* **2021**, *8* (2), 155. DOI:10.3390/children8020155.
8. Habehh, H.; Gohel, S. Machine Learning in Healthcare. *Current Genomics* **2021**, *22* (4), 291–300. DOI:10.2174/1389202922666210705124359.
9. Shahid, N.; Rappon, T.; Berta, W. Applications of Artificial Neural Networks in Health Care Organizational Decision-Making: A Scoping Review. *PLOS ONE* **2019**, *14* (2). DOI:10.1371/journal.pone.0212356.
10. Van Houdt, G.; Mosquera, C.; Nápoles, G. A Review on the Long Short-Term Memory Model. *Artificial Intelligence Review* **2020**, *53* (8), 5929–5955. DOI:10.1007/s10462-020-09838-1.
11. Stevenson, N.; Tapani, K.; Lauronen, L.; Vanhatalo, S. A Dataset of Neonatal EEG Recordings with Seizures Annotations [Data Set]; Zenodo, 2018. DOI:10.5281/zenodo.4940267.
12. Ramakrishnan, S.; Asuncion, R. M. D.; Rayi, A. Localization-Related Epilepsies on EEG. In *StatPearls [Internet]*; StatPearls Publishing: Treasure Island (FL), 2024. Available at https://www.ncbi.nlm.nih.gov/books/NBK557645/ (accessed 2024-04-30).
13. Savadkoohi, M.; Oladunni, T.; Thompson, L. A Machine Learning Approach to Epileptic Seizure Prediction Using Electroencephalogram (EEG) Signal. *Biocybernetics and Biomedical Engineering* **2020**, *40* (3), 1328–1341. DOI:10.1016/j.bbe.2020.07.004.
14. Altın, C.; Er, O. Comparison of Different Time and Frequency Domain Feature Extraction Methods on Elbow Gesture's EMG. *European Journal of Interdisciplinary Studies* **2016**, *5* (1), 35. DOI:10.26417/ejis.v5i1.p35-44.
15. Chollet, F. Keras; GitHub, 2015. [Online] https://github.com/fchollet/keras (accessed 2024-07-07).
16. Zeedan, A.; Al-Fakhroo, K.; Barakeh, A. *EEG-based seizure detection using feed-forward and LSTM neural networks based on a neonates dataset* **2022**. DOI:10.36227/techrxiv.20728411.v1.
17. Yang, M.; Wang, J. Adaptability of Financial Time Series Prediction Based on Bilstm. *Procedia Computer Science* **2022**, *199*, 18–25. DOI:10.1016/j.procs.2022.01.003.
18. Abbasi, M. U.; Rashad, A.; Basalamah, A.; Tariq, M. Detection of Epilepsy Seizures in Neo-Natal EEG Using LSTM Architecture. *IEEE Access* **2019**, *7*, 179074–179085. DOI:10.1109/access.2019.2959234.

■ Author

Sonali Santhosh is a senior at Hanford High School. She plans to major in neuroscience and pursue a career in medicine, particularly surgery. Sonali hopes to expand access to healthcare for underserved populations in the future through innovative solutions such as AI-driven medical tools, combining her passion for medicine and technology to create a lasting impact

DOI: 10.36838/v8i3.29

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Reproductive Risk in India: Lab Evaluation of Prenatal Screening And IVF-Linked Diagnostics

Mannat Oberoi

The Shri Ram School Moulsari, V-37, Moulsari Ave, DLF Phase 3, Sector 24, Gurugram, Haryana 122002; mannatkoberoi@gmail.com
Mentor: Dr. Ankita Dave

ABSTRACT: Infertility is a rising public health issue in India, yet diagnostic practices remain unevenly standardized. This study presents a laboratory-based evaluation of five case studies involving four key diagnostic tools in fertility and prenatal care: Antinuclear Antibody (ANA) testing, Double Marker, Quad Marker, and Non-Invasive Prenatal Testing (NIPT). Using original patient data from an assisted reproductive technology (ART) clinic, the study assessed procedures, interpreted results, and evaluated clinical utility. ANA testing suggested a link between autoimmunity and IVF failure, highlighting the role of immunological factors in implantation. The Double and Quad Marker tests proved to be cost-effective options for early aneuploidy screening but showed moderate specificity. In contrast, NIPT demonstrated high sensitivity and specificity for detecting trisomy, especially when fetal fraction exceeded 4%, making it a reliable tool for prenatal risk assessment. Despite their clinical value, these diagnostics are underutilized due to high costs, limited awareness, and inconsistent regulations. The findings underscore disparities in access and the need for standardized protocols and policy support to improve reproductive healthcare in India. In this case-series, ANA positivity was generally linked with IVF failure, while Double and Quad Marker results remained consistently low-risk. NIPT findings confirmed high reliability when fetal fraction exceeded 4%, demonstrating how these diagnostics function in real-world Indian IVF settings.

KEYWORDS: Translational Medical Sciences, Disease Detection and Diagnosis, Biomarkers, IVF, NIPT.

Introduction

According to the World Health Organization (WHO), reproductive health means that people of all ages should be able to have safe and satisfying sexual lives and make informed choices about if, when, and how they want to have children.[1] Reproductive health is a cornerstone of public health, and in India, improving access to accurate, timely, and culturally sensitive diagnostic services for women remains a pressing challenge.[2]

While maternal health indicators have improved significantly over the past few decades, reproductive health remains a critical concern, particularly for women in India. Among the various issues, infertility has emerged as a growing public health challenge, both globally and nationally. The condition is clinically defined as the inability to conceive after engaging in regular, unprotected sexual intercourse for 12 months or more. Infertility is broadly categorized into two types: primary infertility, wherein an individual has never conceived, and secondary infertility, which is characterized by difficulties in conception following a previous successful pregnancy.[3]

Cultural and Societal Factors Affecting Women's Health:

The experience of infertility is psychologically and emotionally distressing.[4] In many Southeast Asian countries, there is an additional layer of social burden due to deeply rooted societal beliefs and expectations that prioritize reproduction and childbearing.[5] Within such contexts, the inability to conceive may disrupt traditionally ascribed roles and identities, especially for women, leading to further marginalization.

Despite its prevalence, infertility is surrounded by stigma and taboo, which discourages open discussion and limits access to essential support systems. The perception of infertility as a deviation from normative familial and gendered expectations amplifies the psychological toll, often resulting in social exclusion, marital strain, and personal distress.[6]

Widge and Cleland conducted a postal survey with 6,000 gynecologists across India to gain insight into the health services provided and challenges faced in the public sector concerning infertility management. The study revealed that the public sector played a limited role in infertility care. Inadequate infrastructure, lack of information and training, absence of clear protocols at all levels, and private practice by public health doctors were the key problems mentioned.[3]

Government Policies to Address Infertility Issues in India:

With rising infertility rates, the Indian Parliament enacted the Assisted Reproductive Technology (Regulation) Act, 2021.[7] It protects the rights of infertile couples and surrogate mothers, ensuring financial security for surrogates. The Act regulates ART clinics and banks, mandating the ethical and medically supervised handling of gonadal tissues, embryos, and gametes for both clinical and research use. The Indian Council of Medical Research (ICMR) also issued guidelines to ensure ethical and medical compliance.

As ART services expand, *in vitro* fertilization (IVF) has become central to infertility management. The IVF cycle typically involves several stages: controlled ovarian stimulation using hormonal injections, egg retrieval, fertilization in a lab-

DOI: 10.36838/v8i3.36

oratory setting, embryo culture, and embryo transfer into the uterus, followed by hormonal support and monitoring to assess implantation and pregnancy outcomes.[8,9]

Accurate and timely diagnostic testing is critical in the IVF process. Tests such as antinuclear antibody (ANA) screening, double and quad marker tests, and non-invasive prenatal testing (NIPT) are increasingly integrated into IVF and prenatal care.

The ANA test detects antibodies that target cell nuclei and can provide insight into autoimmune contributors to infertility. **The Double Marker Test (First Trimester Maternal Serum Screening)**, performed between 9–13 weeks of pregnancy, measures Pregnancy-Associated Plasma Protein A (PAPP-A) and free β-hCG to assess the risk of chromosomal abnormalities. When combined with a nuchal translucency (NT) scan, this test improves early risk assessment for genetic disorders.

The Quad Marker Test (Second Trimester Maternal Serum Screening), performed between 15 and 22 weeks, evaluates four biochemical markers to estimate the risk of chromosomal abnormalities and neural tube defects.

With the increasing availability of such diagnostic tools, there is a growing need to assess their clinical impact, accessibility, and interpretation in real-world practice. These modalities not only provide early risk assessments for genetic and immunological disorders but also guide decision-making during pregnancy and fertility treatments.

IVF-related prenatal screening includes NIPT, ultrasound imaging, and invasive diagnostic procedures. NIPT analyzes circulating fetal DNA fragments in maternal blood to detect chromosomal abnormalities such as Down syndrome (trisomy 21), Edwards syndrome (trisomy 18), Patau syndrome (trisomy 13), and sex chromosome anomalies. Ultrasound examinations provide structural and anatomical assessments, while invasive methods like chorionic villus sampling can directly diagnose genetic conditions.

Despite technological progress, barriers such as inconsistent guidelines, lack of awareness, limited standardization, and socio-economic disparities hinder the effective implementation of these tests in routine reproductive healthcare.

This study seeks to address these gaps by conducting a laboratory-driven investigation rooted in practical diagnostic experiences and supported by original data analysis. The aim is to assess the performance, utility, and limitations of NIPT, IVF-related immunodiagnostics (Figure 1), and prenatal screening tests in Indian diagnostic settings, while also considering broader implications for healthcare providers and policymakers seeking to improve reproductive health outcomes.

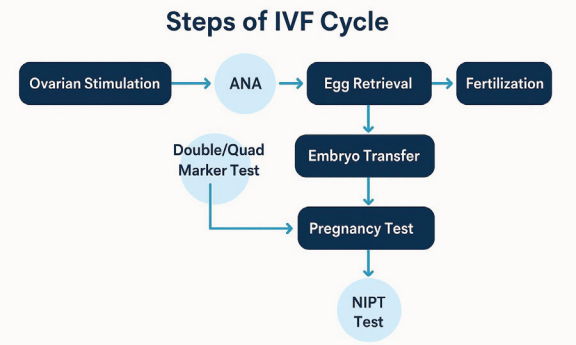


Figure 1: Diagrammatic representation of the stages of the IVF process at which ANA, Double Marker, Quad Marker, and NIPT tests are typically performed, aligned with clinical decision points.

Objectives of the Research:

To learn, evaluate, and analyze key diagnostic methods used in fertility and prenatal care through hands-on laboratory testing and interpretation of real patient data.

Why This Study is Different:

Unlike literature-based reviews, this study integrates personal lab-based experimentation with original clinical data, offering firsthand insights into diagnostic accuracy and real-world applicability.

Methodology

Study Design and Participants:

This was a descriptive case-series analysis of samples from patients (identification was kept anonymous) and laboratory reports obtained from a single diagnostic laboratory. Reports of five patients were selected for each of the following investigations: antinuclear antibody (ANA), first-trimester double-marker screening, second-trimester quadruple-marker screening, and non-invasive prenatal testing (NIPT). Patient identifiers were removed, and each report was assigned an anonymized code (P001–P005). The dataset contained maternal age, gestational age at the time of testing, laboratory values (raw concentrations and multiples of median, MoM), ANA index values, NIPT results, fetal fraction percentages, and IVF outcomes when available.

Sample Collection and Laboratory Analysis:

1. ANA Testing:

Peripheral blood samples were analyzed using an enzyme-linked immunosorbent assay (ELISA) for ANA. The assay was performed according to the manufacturer's protocol (ANAscreen ELISA Assay Kit®, Eagle Bioscience). Optical density was measured at 450 nm on a spectrophotometer. ANA index values were calculated as the sample absorbance/negative control absorbance (Table 1).

Table 1: Parameters for ELISA assay for ANA test. The key analytical parameters used in the ELISA assay for ANA detection, including standard cut-off values.

Parameter	Values / Description
Expected Values	Negative: < 0.9 Borderline: 0.9 – 1.1 Positive: > 1.1
Cut-off Index	1.1 (normal range)

2. Double-Marker Screening:

Maternal serum samples collected between 11–14 weeks of gestation were tested for free β-hCG and pregnancy-associated plasma protein A (PAPP-A). Measurements were performed on Elecsys free β-hCG and PAPP-A assays on the Roche cobas e / Elecsys platform. Raw concentrations were converted to MoM values using laboratory gestational-age medians, with adjustment for maternal weight where applicable.

a. MoM β-hCG > 2.0: high risk for Trisomy 21

b. MoM PAPP-A < 0.5: high risk for Trisomy 21/Trisomy 18

3. Quadruple-Marker Screening:

Maternal serum samples collected between 15–22 weeks of gestation were analyzed for α-fetoprotein (AFP), unconjugated estriol (uE3), free β-hCG, and inhibin-A. Analytes were measured on Roche Elecsys / cobas e (electrochemiluminescence immunoassay), and MoMs were calculated using the laboratory's internal medians. Composite risk scores for trisomies were generated by the lab software, using cut-offs of MoM > 2.0 for β-hCG and inhibin-A, and MoM < 0.5 for PAPP-A and AFP (Table 2).

Table 2: Interpretation of Quad Marker Screening Results using the four biochemical markers measured in the quad test, with clinical implications of high and low levels for each marker in the context of prenatal screening.

Marker	High-Level	Low-Level
AFP (Alpha-fetoprotein)	May indicate neural tube defects (e.g., spina bifida)	May suggest Down syndrome (Trisomy 21)
hCG (Human Chorionic Gonadotropin) uE3 (Unconjugated Estriol)	Linked to Down syndrome (Trisomy 21) —	May indicate Edwards syndrome (Trisomy 18) Can be seen in both Trisomy 21 and Trisomy 18
Inhibin A	Associated with Down syndrome (Trisomy 21)	—

4. Non-Invasive Prenatal Testing (NIPT):

NIPT was performed between 11–13+4 weeks of gestation. The procedure was performed as described by LeFevre & Sundermeyer (2020).[10] cfDNA was extracted from 10 mL maternal plasma collected into cfDNA stabilization tubes, using Qiagen QIAamp™ Circulating Nucleic Acid Kit. Library preparation and sequencing were conducted on the Illumina NextSeq 500/550 sequencing platform. Reads were aligned and analyzed with the laboratory's VeriSeq™ NIPT Analysis Software to detect aneuploidy for chromosomes 13, 18, and 21. Fetal fraction was estimated by the SNP-based method/fragment size distribution method. Samples with fetal fraction < 4% were considered unreliable.

Data Extraction and Analysis:

From each report, the following variables were extracted: patient code, maternal age, gestational age at test, raw analyte concentrations, MoM values, ANA index, NIPT result (positive/negative for trisomy 13/18/21), fetal fraction, and IVF outcome. Data were compiled into summary tables. Given the small sample size (n = 5 per test), results were presented descriptively as counts, ranges, and mean values where appropriate. No inferential statistics were applied.

■ Results

In this study, I have made an effort to bring together my experience of witnessing the diagnosis and interpretation of lab findings for the IVF procedure. I analyzed the reports of 5 patients for each test- ANA, double marker, Quadruple Marker test, and NIPT to understand the procedure involved, sensitivity, and the interpretation of the results obtained. To streamline presentation, all patient results are consolidated into Table 3.

1. ANA Testing: Two of five patients (40%) were ANA-positive (index values: 1.62 and 1.32). One ANA-positive patient experienced recurrent IVF failure, while all three ANA-negative patients had successful IVF outcomes.

2. Double Marker: All five patients screened low-risk between 11–14 weeks of gestation. Mean MoM values were 1.96 for free β-hCG (range: 1.28–3.31) and 1.52 for PAPP-A (range: 0.95–1.92). No abnormal risk scores were observed.

3. Quad Marker: All patients tested between 15–22 weeks were low-risk. Mean MoMs were AFP 1.18 (range: 0.66–1.76), uE3 1.28 (range: 1.14–1.68), β-hCG 1.32 (range: 0.71–2.23), and inhibin-A 1.51 (range: 0.99–2.23). No results exceeded risk thresholds.

4. NIPT: Four patients received negative results for trisomies 13/18/21, while one patient (20%) was flagged as high-risk for Trisomy 21. Fetal fraction values (not shown) ranged from 8.4% to 13.4%, all above the 4% reliability threshold.

Table 3: De-identified laboratory and outcome data from five patients. MoM values are multiples of the median adjusted for gestational age. IVF outcomes reported where available.

Patient ID	ANA Result	Double Marker β-hCG (MoM)	Double Marker PAPP-A (MoM)	Quad AFP (MoM)	Quad uE3 (MoM)	Quad β-hCG (MoM)	Quad Inhibin A (MoM)	IVF Outcome	NIPT Result
P001	Positive	3.31	0.95	2.02	1.65	0.99	1.62	Success	Negative
P002	Negative	1.28	0.66	1.68	1.25	1.84	1.84	Success	Negative
P003	Negative	1.92	1.39	1.14	0.71	1.60	1.60	Success	Positive (T21)
P004	Positive	1.70	1.76	1.25	0.76	1.74	1.74	Recurrent failure	Not applicable
P005	Negative	1.58	1.15	1.06	2.23	1.40	2.23	Success	Negative

■ Discussion

The integration of advanced diagnostics for infertility management in India faces systemic challenges: limited awareness, inconsistent clinical guidelines, socio-economic barriers, and variation in result interpretation. There is a pressing need to evaluate how these tools are being applied in real-world settings and their impact on clinical decision-making in IVF **(Figure 1)**. This study evaluates four diagnostic modalities—ANA testing, Double Marker, Quad Marker, and NIPT—in the context of IVF and prenatal care in India. The findings highlight both their clinical relevance and systemic challenges to implementation.

ANA Testing and IVF Outcomes:

According to the ANA result and respective IVF outcome, the ANA-negative patient successfully conceived, whereas one of the positive cases experienced failed outcomes (Table 3). This pattern aligns with multiple studies showing that ANA positivity is associated with lower oocyte maturation, reduced rate of high-quality embryos, and poorer IVF outcomes, including decreased pregnancy and implantation rates

DOI: 10.36838/v8i3.36

and increased miscarriage risk.[11] ANAs are thought to impair both embryo quality and endometrial receptivity, potentially via autoimmune-mediated inflammation.[12] In contrast, ANA-negative results help rule out immune-related causes, allowing clinicians to proceed with standard IVF protocols. For ANA-positive patients, immunomodulatory treatments such as low-dose prednisone, aspirin, or IVIG are sometimes used to reduce inflammation and improve implantation chances, although their benefits remain debated. Our results also showed that in one of the case studies, where the patient was ANA-positive, a successful pregnancy occurred.[11] This could be possible with the use of immunomodulatory treatments. ANA testing is particularly valuable for women with repeated IVF failure or unexplained infertility, as it can guide more personalized management. However, it is not yet part of standard infertility workups in many Indian clinical settings, despite being inexpensive, minimally invasive, and clinically informative.[11]

Double Marker Screening:

All patients screened were low-risk, consistent with the test's utility as a first-line, early, non-invasive screening method for chromosomal abnormalities.[13] When combined with a nuchal translucency ultrasound, the double marker improves risk estimation for Trisomy 21 and 18.[14] However, its moderate specificity increases false positives, necessitating confirmatory follow-up in high-risk cases. Despite these limitations, it remains widely accessible and cost-effective in India, making it a practical option compared with more advanced tests such as NIPT.[15]

Quad Marker Screening:

Second-trimester Quad Marker testing also yielded low-risk results across all patients. While less specific than NIPT, it plays a vital role in detecting neural tube defects—an area where NIPT has limited utility.[16] It is particularly useful in rural and low-resource settings, where first-trimester screening may be missed and NIPT is cost-prohibitive. Despite moderate specificity, its affordability and accessibility make it a staple in India's public and semi-urban healthcare systems.[16]

Non-Invasive Prenatal Testing (NIPT):

These results demonstrate a mix of high-risk and low-risk interpretations, providing a clear opportunity to assess both the negative predictive value (NPV) and the positive predictive value (PPV) of NIPT in a real-world cohort. The three negative cases are clinically reassuring, given NIPT's NPV exceeds 99.9% for common aneuploidies such as Trisomy 21, 18, and 13 (Table 3).[17] The elevated fetal fractions (not shown in the table) in all cases further support the reliability of these results, as fetal fraction is a key determinant of test accuracy. While postnatal outcomes were not available, the lack of reported complications or follow-up interventions suggests confidence in these negative findings.[18] The single high-risk result in this case series (P003) highlights the clinical utility of NIPT as a sensitive screening tool. A fetal fraction of 12.4% ensured adequate representation of cell-free DNA (cfDNA), and the identification of elevated risk for Trisomy 21 demonstrates the test's sensitivity. Existing research indicates that NIPT's sensitivity for Trisomy 21 exceeds 99%, with a specificity exceeding 99.9%.[19] However, the PPV of a positive result varies depending on maternal age, prior risk, and population prevalence. In high-risk groups, such as women of advanced maternal age or those undergoing IVF, the PPV is considerably higher than in the general population.[20] In all the given case studies, the fetal fractions greater than 8% minimize the risk of test failure or inconclusive results. Fetal fraction, which represents the proportion of fetal cfDNA in maternal plasma, is known to directly affect the reliability of NIPT. A threshold of greater than or equal to 4% is typically required for accurate interpretation. Low fetal fraction is associated with reduced sensitivity and test failure and is more common in patients with high BMI, early gestational age, or underlying medical conditions.[21] None of the cases here fell into that category, which strengthens the reliability of the interpretation. Despite its high performance, NIPT is not a diagnostic test. False positives can occur due to biological variables such as confined placental mosaicism, vanishing twin syndrome, or maternal chromosomal abnormalities. Accordingly, clinical guidelines consistently recommend that all positive NIPT results be confirmed via invasive testing, typically chorionic villus sampling (CVS) or amniocentesis, before any clinical decisions are made.[22]

More affordable alternatives to NIPT include first-trimester combined screening (serum markers plus nuchal translucency) and the second-trimester quadruple marker test. While these options are less expensive and widely available, their sensitivity and specificity are lower than NIPT.

While NIPT outperforms conventional serum-based tests, barriers such as high cost, lack of insurance coverage, and inequitable access limit widespread use in India.[23] Genetic counseling remains essential for appropriate patient interpretation of both positive and negative results.[24] Patients must understand that while a negative result is highly reassuring, it does not rule out all chromosomal or structural abnormalities. Conversely, a positive result indicates an elevated risk, not a diagnosis. Counseling plays a vital role in helping patients interpret their results, particularly when follow-up diagnostic procedures are indicated.

■ Limitations

This study has several limitations. The small sample size (n = 5 per test) and descriptive design limit the statistical strength and generalizability of the findings. All data were derived from a single diagnostic laboratory, which may not reflect broader clinical practice across India. The absence of detailed patient demographics, such as IVF cycle number or comorbidities, further reduces the depth of interpretation. In addition, the lack of follow-up outcomes restricts the ability to validate long-term clinical relevance. These constraints highlight the need for larger, multi-center studies with longitudinal data to better assess the role of these diagnostics in improving reproductive healthcare.

Conclusion and Future Perspective

This study highlights the value of integrating immunological and genetic diagnostics into reproductive healthcare in India. Unlike prior literature-based reviews, this work draws on original case data, showing that while ANA-positivity generally correlated with IVF failure, one ANA-positive patient achieved success—an anomaly that both supports and complicates existing findings. ANA testing can help uncover immune-related barriers to successful IVF, while serum-based marker tests remain widely accessible tools for early prenatal screening. NIPT offers superior accuracy for detecting common aneuploidies but remains constrained by cost and inequitable access.

Moving forward, expanding access to advanced diagnostics like NIPT will require supportive policies, wider insurance coverage, and integration into public health programs. At the same time, greater awareness of affordable immunological and serum-based tests can improve baseline infertility and prenatal care, especially in resource-limited settings. Strengthening genetic counselling services is also essential to ensure that patients and clinicians can interpret results appropriately and make informed decisions.

India's reproductive diagnostics landscape is at a pivotal stage: combining low-cost screening with advanced genomic tools, supported by standardized protocols and equitable access, has the potential to significantly improve IVF outcomes and maternal-foetal health in the coming decade. With the Indian women's reproductive health market projected to grow by 6.7% by 2033,[25] these diagnostic advances—if paired with policy reforms—can translate into more equitable and effective fertility care.

Ethical Statement

This study was conducted under the supervision of qualified professionals in a certified diagnostic laboratory. The student did not handle any clinical specimens directly. All sample processing followed institutional biosafety and ethical guidelines, and no personal or identifiable data were used.

Acknowledgments

I would like to sincerely thank AGILE (Advanced Genomics Institute of Laboratory mEdicine) for their invaluable support and cooperation throughout the course of this research. Their guidance during my training and their assistance in providing access to relevant data played a crucial role in shaping the direction and depth of this study.

References

1. United Nations. *Report of the International Conference on Population and Development, Cairo, 5–13 September*; Document A/Conf.171/13; United Nations: New York, 1994.
2. World Health Organization. *AC597B1 Indicator.* https://data.who.int/indicators/i/C071DCB/AC597B1 (accessed 2025-08-05).
3. Widge, A.; Cleland, J. The Public Sector's Role in Infertility Management in India. *Health Policy Plan.* **2009**, *24* (2), 108–115. https://doi.org/10.1093/heapol/czn053.
4. Lansakara, N.; Wickramasinghe, A. R.; Seneviratne, H. R. Feeling the Blues of Infertility in a South Asian Context: Psychological Well-Being and Associated Factors among Sri Lankan Women with Primary Infertility. *Women's Health* **2011**, *51* (4), 383–399. https://doi.org/10.1080/03630242.2011.574790.
5. Patel, A.; Sharma, P. S. V. N.; Kumar, P.; Binu, V. S. Sociocultural Determinants of Infertility Stress in Patients Undergoing Fertility Treatments. *J. Hum. Reprod. Sci.* **2018**, *11* (2), 172–179. https://doi.org/10.4103/jhrs.JHRS_134_17.
6. Xie, Y.; Ren, Y.; Niu, C.; Zheng, Y.; Yu, P.; Li, L. The Impact of Stigma on Mental Health and Quality of Life of Infertile Women: A Systematic Review. *Front. Psychol.* **2023**, *13*, 1093459. https://doi.org/10.3389/fpsyg.2022.1093459.
7. Government of India. *The Assisted Reproductive Technology (Regulation) Act, 2021 (Act No. 42 of 2021)*. https://www.indiacode.nic.in/bitstream/123456789/17031/1/A2021-42%20.pdf (accessed 2025-08-05).
8. Zhang, J. J.; Merhi, Z.; Yang, M.; Bodri, D.; Chavez-Badiola, A.; Repping, S.; van Wely, M. Minimal Stimulation IVF vs Conventional IVF: A Randomized Controlled Trial. *Am. J. Obstet. Gynecol.* **2016**, *214* (1), 96.e1–96.e968. https://doi.org/10.1016/j.ajog.2015.08.009.
9. Shrestha, D.; La, X.; Feng, H. L. Comparison of Different Stimulation Protocols Used in *In Vitro* Fertilization: A Review. *Ann. Transl. Med.* **2015**, *3* (10), 137. https://doi.org/10.3978/j.issn.2305-5839.2015.04.09.
10. LeFevre, N. M.; Sundermeyer, R. L. Fetal Aneuploidy: Screening and Diagnostic Testing. *Am. Fam. Physician* **2020**, *101* (8), 481–488. https://www.aafp.org/pubs/afp/issues/2020/0415/p481.html.
11. Zhu, Q.; Wu, L.; Xu, B.; Hu, M.-H.; Tong, X.-H.; Ji, J.-J.; Liu, Y.-S. A Retrospective Study on IVF/ICSI Outcome in Patients with Anti-Nuclear Antibodies: The Effects of Prednisone plus Low-Dose Aspirin Adjuvant Treatment. *Reprod. Biol. Endocrinol.* **2013**, *11* (1). https://doi.org/10.1186/1477-7827-11-98.
12. Zeng, M.; Wen, P.; Duan, J. Association of Antinuclear Antibody with Clinical Outcome of Patients Undergoing a Fertilization/Intracytoplasmic Sperm Injection Treatment: A Meta-Analysis. *Am. J. Reprod. Immunol.* **2019**, *82* (3). https://doi.org/10.1111/aji.13158.
13. Lakhi, N.; Govind, A.; Moretti, M.; Jones, J. Maternal Serum Analytes as Markers of Adverse Obstetric Outcome. *Obstet. Gynaecol.* **2012**, *14* (4), 267–273. https://doi.org/10.1111/j.1744-4667.2012.00132.x.
14. Carmichael, J. B.; Liu, H.-P.; Janik, D.; Hallahan, T. W.; Nicolaides, K. H.; Krantz, D. A. Expanded Conventional First Trimester Screening. *Prenat. Diagn.* **2017**, *37* (8), 802–807. https://doi.org/10.1002/pd.5090.
15. Suresh, S.; Cuckle, H. S.; Jagadeesh, S.; Ghosh, K.; Vemavarapu, G.; Taval, T.; Suresh, S. Down's Syndrome Screening in the First Trimester with Additional Serum Markers: Indian Parameters. *J. Obstet. Gynecol. India* **2019**, *70* (1), 12–17. https://doi.org/10.1007/s13224-018-1198-1.
16. Keller, N. A.; Rijshinghani, A. Advantages of the Quadruple Screen over Non-Invasive Prenatal Testing. *Clin. Case Rep.* **2016**, *4* (3), 244–246. https://doi.org/10.1002/ccr3.493.
17. La Verde, M.; De Falco, L.; Torella, A.; Savarese, G.; Savarese, P.; Ruggiero, R.; Conte, A.; Fico, V.; Torella, M.; Fico, A. Performance of Cell-Free DNA Sequencing-Based Non-Invasive Prenatal Testing: Experience on 36,456 Singleton and Multiple Pregnancies. *BMC Med. Genomics* **2021**, *14* (1). https://doi.org/10.1186/s12920-021-00941-y.
18. Caldwell, S.; Almasri, E.; Schmidt, L.; Xu, C.; Dyr, B.; Wardrop, J.; Cacheris, P. Not All Low Fetal Fraction Cell-Free DNA Screening

DOI: 10.36838/v8i3.36

Failures Are at Increased Risk for Aneuploidy. *Prenat. Diagn*. **2021**, *41* (11), 1372–1379. https://doi.org/10.1002/pd.5918.
19. Stokowski, R.; Wang, E.; White, K.; Batey, A.; Jacobsson, B.; Brar, H.; Balanarasimha, M.; Hollemon, D.; Sparks, A.; Nicolaides, K.; Musci, T. J. Clinical Performance of Non-Invasive Prenatal Testing (NIPT) Using Targeted Cell-Free DNA Analysis in Maternal Plasma with Microarrays or Next Generation Sequencing (NGS) Is Consistent Across Multiple Controlled Clinical Studies. *Prenat. Diagn*. **2015**, *35* (12), 1243–1246. https://doi.org/10.1002/pd.4686.
20. Taneja, P. A.; Snyder, H.; de Feo, E.; Kruglyak, K. M.; Halks-Miller, M.; Curnow, K. J.; Bhatt, S. Non-Invasive Prenatal Testing in the General Obstetric Population: Clinical Performance and Counseling Considerations in over 85,000 Cases. *Prenat. Diagn*. **2016**, *36* (3), 237–243. https://doi.org/10.1002/pd.4766.
21. Canick, J. A.; Palomaki, G. E.; Kloza, E. M.; Lambert-Messerlian, G. M.; Haddow, J. E. The Impact of Maternal Plasma DNA Fetal Fraction on Next-Generation Sequencing Tests for Common Fetal Aneuploidies. *Prenat. Diagn*. **2013**, *33* (7), 667–674. https://doi.org/10.1002/pd.4126.
22. Liehr, T. False-Positives and False-Negatives in Non-Invasive Prenatal Testing (NIPT): What Can We Learn from a Meta-Analysis on >750,000 Tests? *Mol. Cytogenet*. **2022**, *15* (1). https://doi.org/10.1186/s13039-022-00612-2.
23. Verma, I. C. Non-Invasive Prenatal Testing: The Indian Perspective. *J. Fetal Med*. **2014**, *1* (3), 113–118. https://doi.org/10.1007/s40556-014-0025-8.
24. Sachs, A.; Blanchard, L.; Buchanan, A.; Norwitz, E.; Bianchi, D. W. Recommended Pre-Test Counseling Points for Non-Invasive Prenatal Testing Using Cell-Free DNA: A 2015 Perspective. *Prenat. Diagn*. **2015**, *35* (10), 968–971. https://doi.org/10.1002/pd.4666.
25. IMARC Group. *India Women's Reproductive Health Market: Industry Trends, Share, Size, Growth, Opportunity, and Forecast 2025–2033*. https://www.imarcgroup.com/india-women-reproductive-health-market (accessed 2025-08-05)

Author

I'm Mannat Oberoi, a Grade 12 IBDP student at The Shri Ram School, Moulsari, passionate about biochemistry and biomedical sciences. This paper builds on my prior research in IVF diagnostics, finding that while all four tests aid fertility care, NIPT is most accurate yet often inaccessible in India.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Enhanced Neuroprotective Delivery of Ashwagandha via Chitosan Nanoparticles: Improving Blood-Brain Barrier Permeability for Accessible Mental Health Therapeutics

Mirae Do

Singapore American School, 40 Woodlands Street 41, 738547, Singapore; mirae07do@gmail.com
Mentor: Woo Rin Lee

ABSTRACT: Tailored approaches to mental health care are crucial in enhancing its accessibility and efficacy. Focusing on cultural sensitivity, Ashwagandha (ASH), a serum cortisol-reducing evergreen shrub, is one of many herbal remedies in non-Western traditional medicine. Since existing off-the-counter ASH supplements lack efficiency in permeating the blood-brain-barrier (BBB), this study aims to investigate methods to increase the efficient delivery of ASH to brain cells and simultaneously enhance protective effects on neuron cells. We used a transwell model to mimic the BBB and tested the protective effects of the ASH-chitosan complex on brain cells subjected to H_2O_2-induced oxidative stress. Using a transwell BBB model with human endothelial and neuronal (SH-SY5Y) cells, we evaluated neuroprotection against H_2O_2-induced oxidative stress (7% concentration, inducing ~65% cell death). ASH-CNP complexes demonstrated significantly enhanced cell viability (83.0 ± 2.1%) compared to H_2O_2 controls (63.0 ± 1.8%, $p = 0.0016$), while individual ASH or CNP treatments showed no significant improvement. This higher cell viability was more prominent in the ASH-chitosan complex as compared to ASH or chitosan, individually. Through this study, a low-cost and applicable method of creating a complex that has protective effects and permeates the BBB more effectively, demonstrating possible improvements that can be made to existing ASH supplements. These findings demonstrate at least approximately 32% improvement in neuroprotection, supporting the development of enhanced over-the-counter ASH supplements for underserved populations with limited access to conventional mental health care.

KEYWORDS: Neurobiology, Ashwagandha (ASH), Blood-Brain-Barrier (BBB), Cultural Competence, Accessible Mental Health Care.

Introduction

Withania somnifera, an evergreen shrub commonly known as Indian ginseng or Ashwagandha (ASH), is a medicinal herb usually found in South Asia, Northern Africa, and the Middle East.[1] It is known for its effects on lowering stress and anxiety, which can be attributed to its ability to reduce serum cortisol levels. In addition, studies have shown that it can heighten sleep quality and improve cognitive function. The use of this Indian traditional medicine can be traced back approximately 3000 years, demonstrating the prominence of the shrub within the Indian culture of Ayurveda medicine.[2]

Similarly, traditional medicine such as Chinese Traditional Medicine (TCM), Unani Medicine, and Siddha Medicine are widely practiced in various regions of the world with their roots deeply ingrained within the respective cultures.[3] Involving acupuncture, herbal medicine, and spiritual therapies, traditional practices are preferred for some due to their perceived capabilities, affordability, and cultural significance. Globally, the TCM market size in 2024 was approximated to be 247.22 billion United States dollars, with projections for it to reach around 450 billion United States dollars by 2033.

Still, Modern Western Medicine (MWM) remains the dominant medical system globally. In contrast to traditional medicine, it employs a method in which medical conditions are treated through evidence-based approaches involving pharmaceutical drugs, radiation, and surgery.[4] In the United States alone, from the years 2000 to 2018, the pharmaceutical industry invested an estimated 879.3 million dollars into novel drug development, including failure and capital costs.[5]

Despite the proven effectiveness of MWM, the incorporation of traditional methods, also known as Complementary and Alternative Medicine (CAM), can increase the care accessibility of marginalized communities. This is executed by addressing cultural needs, reducing health gaps related to race and ethnicity, and providing a more comprehensive approach to health.[6] The World Health Organization encourages the evidence-based integration of CAM into conventional medical care to increase the availability of healthcare.

Specifically within the domain of mental health and psychiatric care, consideration of one's culture is especially important.[7] Studies have found that racial minorities have a greater skepticism toward prescription drugs compared to their white counterparts. This calls for more research to deepen the understanding of cultural attitudes toward MWM and strengthen the effectiveness in the integration of CAM and culturally competent practices that respond to the needs and limitations of various groups.[4]

In the case of Korean Traditional Medicine, acupuncture is often used to alleviate symptoms of depression and anxiety. Other methods, such as Uwhangchungsimwon, a herbal med-

DOI: 10.36838/v8i3.42

icine, are also used to treat mood disorders.[8] Alternatively, in Unani Medicine, which originates from Greece, hydrotherapy in the form of baths and compresses, along with yoga, is used to manage mental health.[9] Currently, despite their long-standing histories in the respective cultures, there is still a need for further research to support their treatment efficacy and cultural significance. Therefore, culturally competent integrative practices of medicine within the realm of mental health and psychiatric care can allow for decreased stigma and a greater receptivity to treatment for non-Western communities.[10]

This study aims to address the current limitations of CAM, specifically over-the-counter ASH supplements, by investigating the method of infusion in chitosan nanoparticles (CNPs). Though ASH has been extensively studied, it still has a restricted delivery capacity to human brain cells. Previous studies have shown that ASH is unable to effectively penetrate the blood-brain barrier (BBB).

To minimize this limitation, our study aims to develop a new delivery system using nanoparticles, specifically CNP. This type of nanoparticle is known to be effective in delivering specific particles through the BBB while maintaining low costs and being biodegradable. By examining how to improve the efficacy of over-the-counter ASH supplements, understanding appropriate dosage, and maintaining an appropriate price range, this study targets those with limited access to quality mental health care. Including cultural and ethnic minorities, communities of a lower socioeconomic status, and those in rural areas, we have aimed to address gaps in vital care.

Methods

Cell culture and materials:

RPMI1640 cell culture media was used to culture the human neuronal SH-SY5Y cell line (initially, the cell passage number was 10). The cell culture media were exchanged every three days to maintain the cells' health with nutrition. The cells were incubated in the CO2 incubator at 37 degrees Celsius. The cells were detached using Trypsin-EDTA dissociation buffer. Then, the experimental cell groups were prepared in 24 24-well cell culture plate.

Preparation of Chitosan Nanoparticles (CNPs) from Chitin:

The chitosan nanoparticles were synthesized from chitin. The molecular weight and degree of deacetylation were not measured in this process. The chitin was incubated with 1M NaOH for 8 hours to remove all proteins. The resulting materials were incubated with HCl to remove excess alkali. Then, the purified chitosan was used to synthesize the nanoparticles with sodium tripolyphosphate with 1% (v/v) acetic acid solution.

Ashwagandha (ASH) infusion into CNPs:

The ASH extract solution was prepared, and only solubilized fractions were isolated to remove the insoluble debris. The ASH extract was added dropwise to the chitosan nanoparticle solution to allow the complex to be formulated.

Cell Viability Test Using Hydrogen Peroxide on Human Brain Cells:

To stimulate the high-stress environment in neurodegenerative conditions, SH-SY5Y cells were exposed to hydrogen peroxide concentrations (0%, 0.5%, 1%, 3%, 7%, 10%) for 24 hours. Then, a cell viability assay was conducted using an automated cell counter machine (LUNA-FL). The topical concentration that induced around 50-70% cell death (7% H_2O_2) was selected for the next experiments.

Designing the Blood-Brain Barrier (BBB) System with Transwell:

A two-chamber transwell (0.4 μm) was used to model the BBB. Human endothelial cells were added to the upper chamber, and the neuronal cells were cultured in the lower chamber. The transwell was maintained for seven days to ensure the barrier formation.

Brain Cell Viability test of ASH-infused CNPs:

Following barrier formation, the upper chamber was treated with the following groups: Untreated control, H_2O_2 only (7%) – positive control for oxidative stress, H_2O_2 + ASH extract, H_2O_2 + CNP, and H_2O_2 + ASH–CNP complex. Each treatment was applied to the upper chamber while hydrogen peroxide was added to the lower chamber. This setup allowed for the analysis of the nanoparticle complex transport through the simulated BBB and its protective effects on neuron cells.

Results and Discussion

Optimizing hydrogen peroxide concentration for stress test on brain cells:

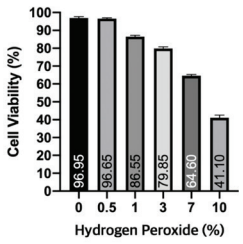


Figure 1: Increasing the hydrogen peroxide concentration decreased the viability of brain cells. The bar graph represents the mean and standard deviation of cell viability for each hydrogen peroxide concentration. This stress test provided more information that aided in the selection of hydrogen peroxide concentration based on its effect on cell viability in brain cell samples.

Table 1: Statistical analysis of cell viability under varying hydrogen peroxide concentrations. One-way ANOVA followed by Tukey's multiple comparisons test was performed to determine statistically significant differences in cell viability between untreated cells (0% H_2O_2) and each treatment group. Significant reductions in viability were observed from 1% H_2O_2 and higher concentrations compared to the control. The 7% H2O2 concentration sample was selected because it had a sufficient level of damage, allowing for neuroprotective effects to be observed most effectively.

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value
0 vs. 0.5	0.3	-3.139 to 3.739	No	ns	0.9989
0 vs. 1	10.4	6.961 to 13.84	Yes	***	0.0002
0 vs. 3	17.1	13.66 to 20.54	Yes	****	<0.0001
0 vs. 7	32.35	28.91 to 35.79	Yes	****	<0.0001
0 vs. 10	55.85	52.41 to 59.29	Yes	****	<0.0001

The purpose of testing the cell viability levels of brain cell samples with various hydrogen peroxide concentrations was to simulate a high-stress environment. Therefore, this experiment aimed to find the optimal concentration of hydrogen peroxide for inducing brain cell stress. This optimal concentration of hydrogen peroxide was then planned for the downstream ex-

periment to investigate the impact of ASH-infused CNPs on brain cell protection.

Figure 1 shows six samples of brain cells with the respective hydrogen peroxide concentration (%) of 0, 0.5, 1, 3, 7, and 10, and their cell viability. The sample with 0% hydrogen peroxide yielded the highest cell viability of 96.95%; the 0.5% hydrogen peroxide sample yielded cell viability of 96.65%; the 1% hydrogen peroxide sample yielded cell viability of 86.55%; the 3% hydrogen peroxide sample yielded cell viability of 79.85%; the 7% hydrogen peroxide sample yielded cell viability of 64.60%; and the 10% hydrogen peroxide sample yielded the cell viability of 41.10%.

Compared to the 0% sample, the 0.5% sample showed no significant difference in cell viability ($p = 0.9989$) (Table 1). However, compared to the 0% sample, the 1% sample indicated a significant difference in cell viability ($p = 0.0002$). Between the 0% sample and the 3% sample, results showed a significant difference in cell viability ($p < 0.0001$). Similarly, compared to the 0% sample, both the 7% and 10% samples showed a significant difference in cell viability ($p < 0.0001$).

We aimed to use a sample with cell viability within the 50% range. This range was most desirable as cell viability of less than 50% would be too damaged to test the impacts of ASH-infused CNPs. On the other extreme, samples with greater than 70% cell viability would not be the most desirable for observing the protection effects of the ASH-infused CNPs, as they are already in relatively healthy condition compared to the 50% cell viability sample. As the 7% H2O2 concentration sample was within this range with a cell viability of 64.60%, it was selected for the following experiments.

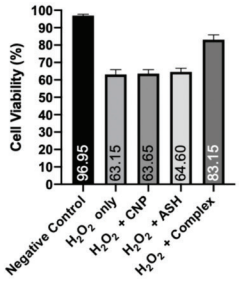


Figure 2: Ashgawanda with chitosan nanoparticle complex showed efficient transport through human blood vessel cells to human brain cells and showed a protective effect against hydrogen peroxide cellular stress. CNP, ASH, and CNP + ASH (Complex) were treated on top of the transwell for seven days. This result demonstrated a synergistic positive relationship between the CNP-ASH complex and cell viability.

Table 2: Statistical comparison of treatment groups under hydrogen peroxide-induced stress. One-way ANOVA followed by Tukey's multiple comparisons test was used to assess the significance of each treatment's effect on neuronal viability. The ASH–CNP Complex significantly improved cell viability compared to H_2O_2-only treatment, while CNP or ASH alone showed no statistically significant effect. The combination of CNP and ASH showed enhanced neuroprotective effects ($p=0.0016$) as compared to CNP and ASH, respectively ($p>0.05$).

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant	Summary	Adjusted P Value
H2O2 only vs. H2O2 + CNP	-0.5	-9.491 to 8.491	No	ns	0.9992
H2O2 only vs. H2O2 + ASH	-1.45	-10.44 to 7.541	No	ns	0.96
H2O2 only vs. H2O2 + Complex	-20	-28.99 to -11.01	Yes	**	0.0016

The experiment shown in Figure 2 was performed to test the impact of CNP, ASH, and the Complex on the cell viability of neurons treated with hydrogen peroxide. This was conducted by inducing hydrogen peroxide stress to the bottom chamber of the transwell, where the neurons lie, and by treating CNP, ASH, or the Complex on the upper chamber of the transwell, where the endothelial blood vessel cells lie. We hypothesized that the combination of CNP and ASH, the Complex, would increase the BBB permeability of ASH. Simultaneously, we tested the protective effects of ASH on H_2O_2 stress-induced environments.

The first sample was a negative control sample with no H_2O_2, CNP, or ASH. Such samples without any treatment maintained a high viability of approximately 97%. It serves as the reference for normal, healthy cell conditions. The second sample was exposed to hydrogen peroxide, inducing a stress environment. Treating cells with H_2O_2 alone significantly reduced cell viability to approximately 63%, confirming the known cytotoxic effect of H_2O_2 via oxidative stress. The third sample was exposed to H_2O_2 and CNP. This did not yield a statistically significant improvement in cell viability compared to sample 2, as adding CNP resulted in a viability of about 63% (Table 2). This result suggests that CNP by itself, under these conditions, provides minimal to no protective effects. The fourth sample was the addition of H_2O_2 and ASH. Treating ASH alone showed a slight improvement in viability of around 64%, but it was not statistically significant (Table 2). This indicates some protective trend, but not enough to fully counteract oxidative stress. Finally, the fifth sample of H_2O_2 and the ASH–Chitosan Complex resulted in a marked increase in cell viability of around 83% (Table 2). This improvement, compared with sample 2 of only H_2O_2, was statistically significant ($p = 0.0016$), indicating a synergistic or enhanced protective effect against H_2O_2-induced damage.

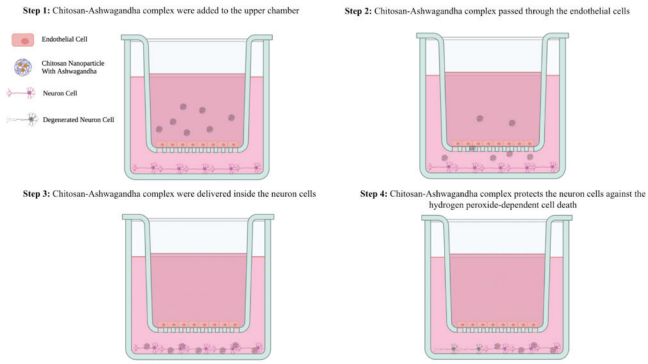


Figure 3: The overall summary of how the chitosan-ASH complex protects human neuron cells against hydrogen peroxide-dependent cell death. The results highlight the complex's potential to improve BBB-like delivery and offer a more accessible, safe alternative to conventional neuroprotective treatments. While limitations exist in model complexity, this study supports further optimization of CNP+ASH formulations and positions them as promising candidates for over-the-counter mental wellness support, particularly in underserved populations.

The transwell method was adopted, in which an upper and lower chamber are present with endothelial cells in the upper chamber and neuron cells in the lower chamber (Figure 3). This method was employed to study the particle migration of

DOI: 10.36838/v8i3.42

the Chitosan-ASH complex through the endothelial cells to the neuron cells in the lower chamber. The transwell with a pore size of 0.4μm was selected as it was most suitable for the relatively small size of the nanoparticles as compared to cells. This process can be divided into four distinct steps. First, the Chitosan-Ashwangandha complex was added to the upper chamber. Second, the Chitosan-ASH complex passed through the endothelial cells and then through the pores of the transwell. Third, the complex was delivered inside the neuron cells after entering the lower chamber. Fourth, the Chitosan-ASH complex was observed to protect neuronal cells against hydrogen peroxide-induced cell death, as cell viability was higher than that of the sample with H_2O_2 but without the Chitosan-ASH complex.

ASH is known for its antioxidant and adaptogenic properties. ASH alone showed a slight improvement in viability. Though not statistically significant, it suggests some protective potential, but insufficient improvement to counteract the oxidative stress induced in the setup.

The Chitosan-ASH complex, however, demonstrated a significant increase in cell viability. This indicates the potential enhanced delivery of ASH's bioactive compounds via CNP. As such, conjugation with chitosan can improve cellular uptake, stability, and sustained release of active molecules, enhancing the protective effect of ASH. These results demonstrate the synergistic protective effects of combining CNP and ASH for relieving H_2O_2 stress-induced cellular environments (Figure 3).

This enhanced neuroprotective effect of the ASH + CNP complex can be explained by the intracellular antioxidant defense mechanisms and improved BBB permeability. Due to their cationic nature, CNP interacts with endothelial cell membranes (negatively charged), to move across the BBB and thereby allowing a greater amount of ASH to reach the neurons. Then, the compound exerts adaptogenic and antioxidant effects by finding reactive oxygen species, reducing lipid peroxidation, and regulating stress-responsive signaling pathways. These include the hypothalamic-pituitary-adrenal (HPA) stress pathway. With the demonstrated enhanced neuronal survival, it is suggested that the CNP-mediated delivery simultaneously improved bioavailability while prolonging the retention and stability of ASH in neuronal environments.

The limitations of this study are that the representative models used, such as the transwell system and cell type, do not perfectly represent the characteristics of the BBB and the complexity of neuronal cells. The BBB is composed of multiple cell types, while there is only one cell type (SH-SY5Y) represented in the experiments. To better mimic the BBB, more cell types could be included in future studies to increase similarity to the multicellular nature of the barrier. Alternatively, mouse animal models can be used to test the efficacy of CNP + ASH (Complex). This mouse model may represent a structure more similar to the human BBB.

Though this study supports the idea that the infusion of ASH in CNP enhances its delivery and protection qualities, the optimization of the ratio between ASH and CNP is yet to be investigated. By studying this further, the overall stability of the ASH + CNP complex may reveal the understanding of its real-life applications. To further determine the approximate increase in cost-efficiency, additional research without the presence of the H_2O-induced environment must be conducted.

One of the most common ways in which the public consumes ASH is through over-the-counter supplements. Taking this into account, analyzing its overall stability and optimal ratio is crucial in understanding the viability of over-the-counter ASH-infused CNP supplements. The main factors in determining viability include shelf life, efficient mass production, and human metabolism rate.

This study increases the accessibility of non-Western medicine, specifically in the context of over-the-counter supplements. Targeting specifically off-the-counter ASH supplements due to their ties with improved mental health, availability, and affordability, this study aims to deliver enhanced results for communities that may experience augmented benefits. Some of these groups that experience difficulty accessing mental health care include youth, low-income communities, rural communities, and non-Western regions with hesitation toward psychiatric prescription drugs. This is not to say that the studied Chitosan-ASH complex can serve as a replacement for psychiatric prescription drugs, but rather a more accessible, alternative option to prescription drugs. This enhanced BBB delivery efficiency of the ASH-infused CNPs can result in lower dosages of ASH required to produce similar results, thereby reducing possible side effects associated with higher dosages of the evergreen shrub. In addition, this complex is safe for human consumption, as those with allergies—even shellfish allergies—can safely consume chitosan. The Chitosan-ASH complex is easily digestible by the human body, as it fully decomposes, and is non-toxic. As chitosan is positively charged, it can successfully be attracted to other negatively charged substances through electrostatic interaction; this poses an alternative research direction for chitosan complexes with other substances than ASH. Through these benefits of the complex investigated in this study, reservations surrounding the efficacy of non-Western traditional medicine can be challenged, further encouraging subsequent research of alternative medicine that may better cater to the needs of less-represented groups.

■ Conclusion

Previous studies have highlighted ASH's ability to relieve stress and improve cognitive functions. However, ingestion of over-the-counter ASH supplements poses limited benefits due to their restricted ability to penetrate the BBB. Therefore, this study leverages CNPs to formulate a complex with ASH to minimize this limitation. A transwell BBB system was developed to observe the impact of the Chitosan-ASH complex's BBB permeability and protective abilities against hydrogen peroxide stress. Our main result indicates that this infusion of ASH in CNPs augments its benefits through statistically significant increases in neuron cell viability. This result provides a new avenue for more effective over-the-counter ASH supplements, increasing the accessibility and maintaining costs of CAM for those with limited access to quality mental health

care, including cultural and ethnic minorities, communities of lower socioeconomic status, and those in rural areas.

Acknowledgments

I want to thank my research mentor, Prof. Lee, for his support and guidance throughout my research journey. His dedication to ensuring that I receive the most useful feedback and critiques has truly allowed me to engage thoroughly and meaningfully in this research.

I would also like to express my gratitude to my family and friends who have engaged in entertaining and intellectually stimulating discussions about my research, for fueling my interest and perseverance throughout the process.

References

1. Kulkarni, S. K.; Dhir, A. Withania Somnifera: An Indian Ginseng. *Prog Neuropsychopharmacol Biol Psychiatry* **2008**, *32* (5), 1093–1105. DOI: 10.1016/j.pnpbp.2007.09.011.
2. Salve, J.; Pate, S.; Debnath, K.; Langade, D. Adaptogenic and Anxiolytic Effects of Ashwagandha Root Extract in Healthy Adults: A Double-Blind, Randomized, Placebo-Controlled Clinical Study. *Cureus* **2019**, *11* (12), e6466. DOI: 10.7759/cureus.6466.ü
3. Patwardhan, B.; Warude, D.; Pushpangadan, P.; Bhatt, N. Ayurveda and Traditional Chinese Medicine: A Comparative Overview. *Evid. Based Complement. Alternat. Med.* **2005**, *2* (4), 465–473. DOI: 10.1093/ecam/neh140.
4. Mortada, E. M. Evidence-Based Complementary and Alternative Medicine in Current Medical Practice. *Cureus* **2024**, *16* (1), e52041. DOI: 10.7759/cureus.52041.
5. Fang, Y.; Tang, M. X.; Liu, X. Traditional Chinese Medicine Trade among RCEP Countries: Structural Characteristics and Determinants. *Front. Public Health* **2024**, *12*, 1508839. DOI: 10.3389/fpubh.2024.1508839.
6. Tucker, C. M.; Marsiske, M.; Rice, K. G.; Nielson, J. J.; Herman, K. Patient-Centered Culturally Sensitive Health Care: Model Testing and Refinement. *Health Psychol.* **2011**, *30* (3), 342–350. DOI: 10.1037/a0022967.
7. Adams, C.; Chatterjee, A.; Harder, B. M.; Mathias, L. H. Beyond Unequal Access: Acculturation, Race, and Resistance to Pharmaceuticalization in the United States. *SSM Popul. Health* **2018**, *4*, 350–357. DOI: 10.1016/j.ssmph.2018.04.003.
8. Oh, H.-M.; Lee, J.-S.; Kim, S.-W.; Oh, Y.-T.; Kim, W.-Y.; Lee, S.-B.; Cho, Y.-R.; Jeon, Y.-J.; Cho, J.-H.; Son, C.-G. Uwhangchungsimwon, A Standardized Herbal Drug, Exerts an Anti-Depressive Effect in a Social Isolation Stress-Induced Mouse Model. *Front. Pharmacol.* **2019**, *10*, 1674. DOI: 10.3389/fphar.2019.01674.
9. Elendu, C. The Evolution of Ancient Healing Practices: From Shamanism to Hippocratic Medicine: A Review. *Medicine (Baltimore)* **2024**, *103* (28), e39005. DOI: 10.1097/MD.0000000000039005.
10. Ahad, A. A.; Sanchez-Gonzalez, M.; Junquera, P. Understanding and Addressing Mental Health Stigma across Cultures for Improving Psychiatric Care: A Narrative Review. *Cureus* **2023**, *15* (5), e39549. DOI: 10.7759/cureus.39549.

Author

Mirae Do is an interdisciplinary researcher with a particular interest in combining her advocacy for mental health literacy and minority populations with neuroscience and psychology. By exploring ways to increase accessibility and inclusivity within mental healthcare through scientific research, she aims to foster a system that is considerate of one's unique needs.

DOI: 10.36838/v8i3.42

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Riverbank Filtration and Nitrate Reduction in South Korea: Subsidence in Regions Prone to Groundwater Extraction

William J. Hong[1], Christian J. Han[2]

1) Chadwick International, 45 Art Center-daero 97beon-gil, Yeonsu-gu, Incheon, 22002, South Korea; jw2026h@gmail.com
2) The Canadian International School of Guangzhou, No.122 Dongyi Road, Panyu District, Guangzhou, Guangdong, 511492, China

ABSTRACT: For its large-scale accessibility, groundwater is the principal water source worldwide. In some countries, however, overreliance on groundwater has resulted in ground subsidence. Riverbank filtration (RBF) is a promising solution wherein river water is extracted and filtered through the river bed's subsurface layers through physical and biochemical processes. However, little research has been done to assess the use of RBF in rural South Korea. This study aims to address this research gap by analyzing nitrate levels, which pose a danger to human consumption, in water sources near rural agricultural areas. The Tancheon River, which fits this description, will be our sample area to do so. The analyzed nitrate concentrations of our samples (ground and river water) ranged widely from 1.6 to 139.2 mg/L NO_3 (mean = 42.6 mg/L), with 40% of the examined samples containing nitrate concentrations exceeding the Korean Drinking Water Standard (44 mg/L NO_3). The increase in nitrate pollution in the Tancheon River suggests that agricultural practices contaminate our study area. This paper emphasizes that the future application of RBF in rural areas should be highly considered in areas of high nitrate pollution.

KEYWORDS: Earth and Environmental Science, Water Science, Riverbank Filtration (RBF), Nitrate Pollution, South Korea.

Introduction

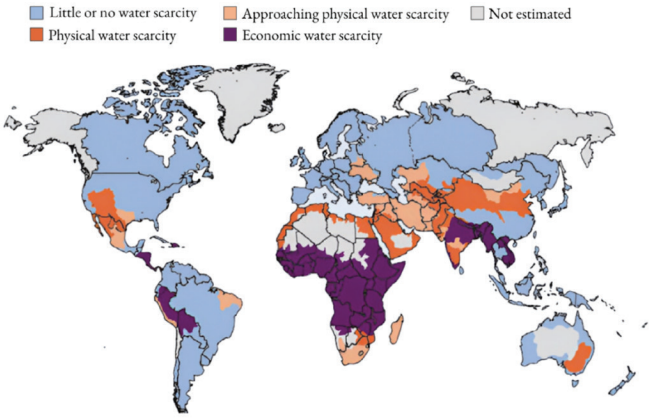


Figure 1: Global distribution of water scarcity types, distinguishing between physical scarcity due to limited natural availability, economic scarcity arising from lack of access to infrastructure, and transitional zones approaching unsustainable water use.

To better understand the global distribution of water scarcity, the map below categorizes countries based on the type and severity of water stress they experience. The color-coded scheme highlights that a significant number of countries across North Africa, the Middle East, South Asia, and parts of North America face physical water scarcity (orange). Economic water scarcity (purple), where infrastructure or governance prevents access to water despite availability, is concentrated in Sub-Saharan Africa. Meanwhile, countries in Southern and Central Asia, as well as parts of South America, are shown to be approaching physical water scarcity (peach), indicating worsening conditions. Currently, over four billion people have no access to public supplies of clean freshwater (Figure 1), and more will live in water-scarce countries in the coming decades.[1] By 2025, the Food and Agriculture Organization states that 1.8 billion people will likely face absolute water scarcity, and two-thirds of the global population could be under water-stressed conditions.[2]

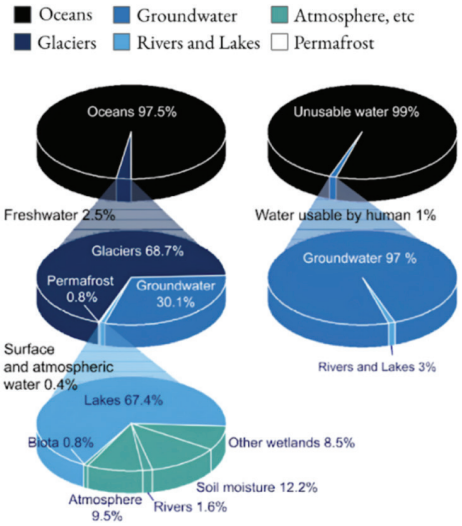


Figure 2: Global distribution of water resources by type and availability, highlighting oceans (black), glaciers (dark blue), groundwater (blue), rivers and lakes (light blue), atmosphere and biota (teal), and permafrost (white).

Figure 2 illustrates the distribution of global water resources, highlighting that only a tiny fraction—just 1%—is usable by humans, with groundwater making up 97% of this accessible freshwater and surface waters such as rivers and lakes accounting for only 3%. Ultimately, environmental pollution and ecological destruction—consequences of climate change, natural disasters, population growth, poverty, warfare, global-

DOI: 10.36838/v8i3.47

ization, urbanization, and disease—significantly impact the water sector and can lead to a global drinking water crisis.[3] As potable freshwater grows scarce, its cost synchronously increases, resulting in a greater demand for groundwater among freshwater reservoirs. As it is, groundwaters hold about 97 percent of the world's freshwater, while surface waters such as rivers and lakes are only about 3 percent (Figure 2). In Jakarta, nitrate pollution from agricultural fertilizers has proliferated, contaminating surface water sources and prompting the population to rely immensely on groundwater.[4] Consequently, the use of the water source has weakened the ground structure, ultimately causing ground subsidence.[5] Likewise, in China, due to groundwater over-extraction and the sheer weight of urban buildings and infrastructure, its major cities are facing similar repercussions of subsidence.[6] Portions of heavy-weight cities such as Beijing have been sinking by centimeters every year.[7]

These international trends have also been reflected in South Korea, where cases of land subsidence have been increasingly linked to excessive groundwater extraction. A notable example is the 2014 Lotte sinkhole incident in Seoul's Songpa District, where uncontrolled subsurface groundwater flow during nearby construction contributed to ground instability and collapse. In South Korea, as climate change threatens to shrink the nation's water resource supplies, alternative water sources are being sought. After investigating numerous options, South Korea has shown an increasing interest in the Riverbank Filtration system (RBF), an economical and simple purification technology capable of producing water of potable quality and ample quantity. South Korea has implemented the RBF in Gyeong Sang Province, a part of the nation that faces a water shortage and is efficiently supplying water to the area today.[8]

The RBF is a natural water purification system that improves water quality by directing river water into alluvial formations through vertical or horizontal collection wells in aquifers near rivers or lakes. As the water passes through the soil, contaminants are naturally removed before the filtered water is pumped to the surface.[9,10] The produced high-quality filtrate of the RBF is the result of natural processes in the riverbed and bank, such as sorption, biodegradation, and groundwater mixing, which occur as raw water passes through river or lake beds into aquifers. A portion of Dissolved Organic Matter, also known as DOM, is eliminated by biofilm layers in the upper filtration zone, further improving water quality.[11] This technology has been implemented in numerous developed regions and has been used for many years to improve drinking water quality, even mitigating damages from hydrogeochemical accidents such as chemical spills.[12]

In rural regions, where agriculture dominates the landscape, the interaction between river water and groundwater significantly influences water quality, introducing various pollutants such as nitrate from fertilizers and animal waste. These areas are distinctly vulnerable to nitrate pollution and in need of a solution. Moreover, elevated nitrate levels in alluvial groundwater pose serious health risks, including methemoglobinemia, a blood disorder that affects how red blood cells carry oxygen, and contribute to environmental issues like eutrophication, which can create algal blooms that degrade aquatic ecosystems.[13]

Therefore, a major aspect of this research is to identify the mixing effects between these water sources and to assess how natural filtration processes can attenuate pollutant concentrations. By comparing riverbank filtered water—a mixture of river water and groundwater—with agricultural groundwater, we can determine the effectiveness of riverbank filtration in reducing nitrate levels.

In considering the mixing effects between water sources, the hydrologic cycle becomes a crucial element to this study as it facilitates, dilutes, and spreads nitrates across bodies of water, including in the study area. Understanding and incorporating this cycle will aid our study to track the processes of contaminant movement, furthering our comprehension of the effects of nitrate pollution spread by hydrological means.

While past literature exists regarding the utility of riverbank filtration—such as projects that tested iron and manganese,[14] providing details on the potential implementation of RBFs,[15] and ones that concerned the process of groundwater pollution attenuation by riverbank filtration[16]—little research has been done in examining how the technology is applicable in rural areas such as the Tancheon river, given local agricultural practices and the expected nitrate pollution.

This study aims to bring such knowledge to a rural context by examining the hydrochemical characteristics and nitrate contamination of shallow groundwater in the Tancheon area of South Korea. The major purposes of this study are: (1) to analyze the chemical composition of main waters related to the hydrologic cycle; (2) to understand the effects of agricultural practices on alluvial groundwater quality, focusing on nitrates; (3) to assess the hydrogeochemical processes, merging river and groundwater in RBF water. The study results will provide insights into the effectiveness of natural filtration processes and inform strategies to safeguard groundwater resources, ensuring their sustainability for future generations in both rural and urban settings.

Methods

1. Study area:

The study focuses on the Tancheon River in Seongnam City, where a number of riverbank filtration facilities are located. Sampling and analysis conducted in July 2024 revealed variations in water composition within the study area.[17] The Tancheon River serves as a typical model for river systems in South Korea because it passes through an extremely urbanized area with intense anthropogenic effects, including residential, industrial, and agricultural land uses. This makes it an ideal model for examining nitrate pollution problems that are normally encountered along South Korea's rivers. Moreover, the overexploitation of alluvial groundwater in these regions close to the Tancheon River contributes to increased risks of land subsidence and sinkholes, with multiple cases having been reported across the field. Therefore, understanding how RBF interacts with such hydrogeological conditions is necessary.

By an examination of the effectiveness of RBF herein, implications could be drawn and applied to larger contexts in

South Korea as a whole, thereby providing information that can guide the nationwide sustainable management of water and control of pollution throughout the country.

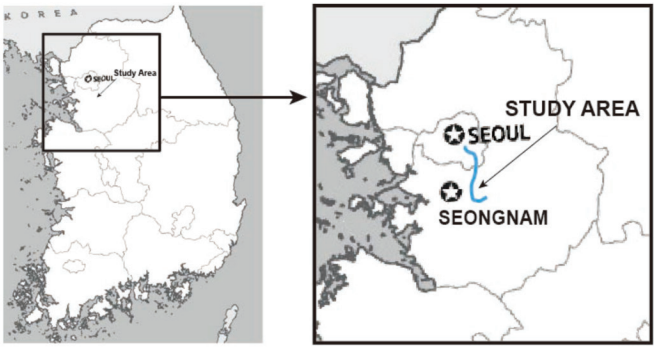


Figure 3: Location map of the study area for riverbank filtration facilities, showing the regional context within South Korea (left), and a magnified view of the study area along the Tancheon River between Seoul and Seongnam (right), where sampling and analysis were conducted.

The Tancheon River is a 36 km-long stream in South Korea, originating from Yongin City in Gyeonggi Province and flowing through Seongnam City before merging with the Han River in Seoul (Figure 3). It has a catchment area totaling 302 km^2, and traverses a variety of urban, industrial, and agricultural landscapes. These diverse land uses introduce pollutants such as domestic sewage, industrial effluents, and agricultural runoff, posing significant challenges to water quality and treatment efforts.

The designated study area lies in Seongnam City, roughly halfway along the Tancheon River. The region receives an average annual rainfall of 1,349 millimeters, with 72% concentrated between May and August, and 40% falling in July alone (as per Seongnam City statistics).[17]

Figure 4: (1) A riverbank filtration (RBF) facility was built along the riverbed to allow for the extraction of subsurface water through infiltration and gravel-based filtration. (2) Water is being pumped from the RBF system to a nearby wetland site to improve the ecology and hydrology.

In this area, an RBF facility plays a critical role in water treatment and ecosystem management. It filters and supplies approximately 5,300 m^3 of water per day. This filtered water is then pumped into nearby artificial wetlands 5 to 10 times daily, contributing to both local water treatment processes and ecological conservation (Figure 4).

Despite the negative impact of population growth and urbanization on the river's water quality, the river in recent years has seen improvements due to development restrictions and efforts to mitigate potential pollution sources. A Korea Institute of Geoscience and Mineral Resources (KIGAM) Report in 2016 shared that Seongnam City measured the biochemical oxygen demand (BOD) of 10 sites of Tancheon River in 2023 and found that the water quality was 1.8 mg/l on average, which is considered a first-class level.[17]

It is important to note that the land use of the surrounding area is largely characterized by agricultural activities, with vegetables grown in house facilities year-round. Agricultural water supply functions on shallow groundwater. Large numbers of irrigation wells in the farmlands of the alluviums—from which shallow alluvial groundwaters are tapped for supplying irrigation waters—are sprayed on the agricultural fields via sprinkler.

Figure 5: Compost piles used for intensive agricultural activities in the study area.

Figure 5 presents the compost piles stored for agricultural use, illustrating the intensity of fertilizer application in the study area. Significant amounts of synthetic nitrogen fertilizers such as urea ($CO(NH_2)_2$), $(NH_4)_2SO_4$, and composite fertilizers are applied mainly in spring and fall seasons. In winter, fertilizers are also applied for crop cultivation in the conservatory. Similarly, the application of compost on agricultural lands is heavily used in the study areas (Figure 5). Vegetables growing in agricultural house facilities in the study area are mainly supplied for school meals. The need for organic products has the continued use of natural-based compost.

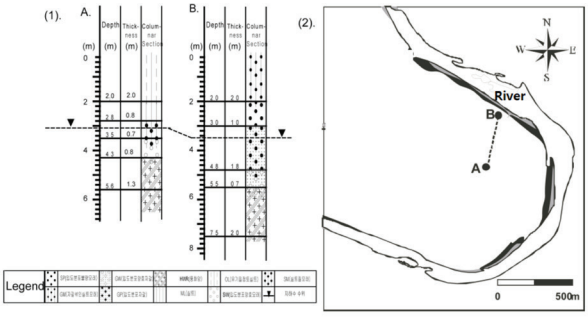


Figure 6: (1) The depth profiles of drilled boreholes (A) and (B) from 5 km downstream of the study area. (2) shows the location of (1) of A and B.

The wide alluvium our study investigates has been formed alongside the Tancheon River. From the river head toward the mouth, the areal extent and depth of alluviums generally increase, whereas the grain size of alluvial aquifer materials tends to be finer. According to the drilling data from a site 5 km down from the study area, inner area A (Figure 6) is composed of silt (2.0 m thick), silty sand (0.7 m thick), coarse-grained sand (0.8 m thick), weathered rock (1.3 m thick), and the low-

DOI: 10.36838/v8i3.47

est bedrock, granitic gneiss. The geology of the outer land, Area B (Figure 6), is composed of silty sand (2.0 m thick), cohesive sand (1.0 m thick), fine-grained sand (1.8 m thick), gravel (0.7 m thick), weathered rock (2.0 m thick), and the bedrock is the same granitic gneiss as in Area A.[17]

As a result, the proportion of clay and silt layers decreases moving towards the river, where the alluvium is primarily composed of sand and gravel (Figure 6/(1)). Alluvial groundwater has a higher permeability and hydraulic conductivity than silt or clay layers within porous media such as gravel and sand. These properties are favorable for groundwater replenishment conditions, but their connection to surface water makes them vulnerable to anthropogenic contamination, such as fertilizers and pesticides, from the surface.

2. Sampling and Analysis:

Figure 7: Spatial distribution of sampling and monitoring points in the study area, including agricultural groundwater wells (SNG1, SNG2), river channel (SNR2), residential river branch (SNR1), subway station groundwater (SNG3), and riverbank filtration facility (SNF1), all situated in proximity to the main and tributary streams.

Water samples were collected from four distinct sites: the main river channel (SNR2), a connected tributary branch (SNR1), agricultural groundwater sources (SNG1 and SNG2), and seepage water originating from a nearby subway station (SNG3). Over the course of three consecutive sampling sessions in July 2024, a total of 11 samples were obtained. These comprised one sample from SNR1, three from SNR2, one from SNG1, two from SNG2, and three from SNG3, in addition to a single rainwater sample (Figure 7).

Figure 8: Part of a field sampling and *in-situ* measurement (pH, Temp., DO, Eh, EC, and alkalinity).

As shown in Figure 8, the sampling procedures included the following steps: 1) purging wells (two to three volumes) for groundwater, 2) sample collection, 3) sample pretreatment, and 4) preservation. Temperature, pH, Eh (redox potential), EC (electrical conductivity), and DO (dissolved oxygen) of water were measured in the field. Bicarbonate alkalinity was also analyzed *in situ* by acid titration with 0.05N HNO_3. Samples of cation and anion analyses were filtered using 0.45 μm cellulose membranes (Figure 8). Cation samples were acidified to a pH of 2 using concentrated HNO_3 to prevent oxidation reactions and precipitation. All samples for chemical analysis were chilled to 4°C in an icebox and refrigerator until analyzed. Major cations were analyzed by ICPAES (Perkin Elmer ELAN 3000) and anions by IC (Dionex 120). We utilized the APHA (American Public Health Association)'s standard method for the examination of water and wastewater. Ensuring the integrity of groundwater samples is crucial for accurate chemical analysis. Strict protocols were followed to minimize contamination, oxidation, and alteration of water chemistry during sampling, storage, and analysis. The use of 0.45 μm cellulose membrane filters helped remove suspended particles that could interfere with ion analysis. Acidification of cation samples to a pH of 2 using concentrated nitric acid prevented precipitation and oxidation reactions that might alter metal concentrations. Furthermore, maintaining all collected samples at 4°C slowed down biological activity and chemical changes, preserving the original composition of the water.

■ Results and Discussion

To understand the effectiveness of the RBF, we examine 1) the hydrochemistry of different water types, 2) nitrate contamination in groundwater, and 3) the hydrogeochemical approach to evaluating the mixing dynamics between river water and groundwater within aquifers affected by RBF. Nitrate pollution, particularly in agricultural groundwater areas, poses significant risks to both human health and aquatic life. This issue is often exacerbated in areas where RBF facilitates the mixing of surface water and groundwater, potentially accelerating the spread of harmful chemical pollutants. Furthermore, a comprehensive hydrogeochemical assessment of this mixing process is critical, as it provides insight into the interactions between river water and groundwater and elucidates the pathways through which contaminants such as nitrates infiltrate these systems.

1. Hydrochemistry of Different Water Types:

Table 1 presents physicochemical data that characterize rainwater, river water, and groundwater within the study area's hydrologic cycle. This data will be used to analyze the riverbank filtration mixing process, highlighting the interactions between river and groundwater.

Table 1: Physicochemical data for hydrologic classification. SNR1 exhibits low levels of Dissolved Oxygen.

Hydrological Cycle	Sampling Site	Site Description	Sampling Date	Temperature (oC)	pH	Eh (mV)	DO (mg/L)	EC (µS/cm)	Dissolved concentration (mg/L) Na+	K+	Ca2+	Mg2+	Cl-	SO4 2-	NO3-	HCO3 -	SiO2
Rainwater		Near SNG2 (groundwater)	18/07/24	24.5	6.27	322	8.1	58.7	3.5	1.1	4.3	1.5	4.4	0.9	5.4	7.6	ND
River water	SNR1	A branch of the river	09/07/24	26.0	7.25	218	0.4	530.5	46.1	21.9	59.1	14.6	50.5	92.4	78.3	175.4	6.4
	SNR2	River channel	09/07/24	24.4	7.66	430	7.3	327.6	33.6	10.3	35.9	6.7	52.6	41.5	12.2	76.3	5.3
	SNR2	River channel	17/07/24	25.0	7.63	198	7.3	281.6	35.4	8.4	33.9	7.5	46.1	38.2	12.1	83.9	4.8
	SNR2	River channel	18/07/24	23.7	7.54	303	7.6	282.6	41.2	9.2	29.1	5.5	51.1	47.1	15.7	53.4	5.1
Groundwater	SNG1	Agricultural Groundwater	09/07/24	24.8	6.65	418	2.7	638.8	27.5	11.7	102.5	13.7	37.5	53.2	139.2	175.4	15.4
	SNG2	Agricultural Groundwater	09/07/24	23.1	6.82	462	8.2	640.1	29.6	12.9	101.3	3.5	22.5	76.5	88.1	152.4	18.6
	SNG2	Agricultural Groundwater	18/07/24	22.9	7.22	322	9.4	580.1	21.8	7.6	87.6	2.9	15.5	55.3	76.3	122.0	17.7
	SNG3	Seepage Water	09/07/24	24.7	7.63	674	8.2	214.8	5.7	1.5	24.3	2.5	3.4	33.4	1.6	45.8	7/5
	SNG3	Seepage Water	18/07/24	25.1	7.71	329	8.2	250.3	7.2	2.2	30.1	4.5	4.4	32.2	2.2	76.3	7/2
Riverbank Filtration Water	SNF1	Near SNR2 (river water)	17/07/24	24.9	7.68	393	7.6	323.4	35.8	5.3	51.3	9.3	48.4	55.6	5.4	91.5	5.3

Rainwater typically exists in high purity due to the processes of the hydrologic cycle. During evaporation, water vapor rises into the atmosphere with minimal incorporation of dissolved solids, gases, or particulate matter. As condensation occurs within clouds, rainwater has limited contact with natural components such as mineral dust or sea salts, resulting in relatively low solute concentrations. However, atmospheric pollutants can influence rainwater composition, especially in urban or industrial regions. During episodes of intense or prolonged precipitation, the scavenging effect of rainfall tends to reduce atmospheric pollutant levels, thereby minimizing their incorporation into the rainwater.

SNR1—a branch originating from residential areas and replete with agricultural runoffs and sewage— exhibited significant contamination, evidenced by its greenish color and strong odor, and a much lower water flow compared to the main channel. Dissolved Oxygen (DO) was recorded at 0.4 mg/l, indicating an anoxic condition unsuitable for aquatic life. Therefore, SNR1 will be excluded from the interpretation of water type because SNR1 exhibited a different appearance and was a non-representative sample compared with SNR2.

Our studies categorize water types by considering the hydrologic cycle and the dissolved composition. Table 2 summarizes the physicochemical properties of four water types in the study area: rainwater, river water, polluted groundwater, and unpolluted groundwater. Major ion compositions, particularly nitrate (NO_3^- : 82.2 vs. 13.3 mg/L) and electrical conductivity (EC: 610.1 vs. 232.5 µS/cm), show distinct differences between polluted and unpolluted groundwater.

Table 2: A summary of the physicochemical properties of four water types in the study area, with each value representing the average for its respective category. Polluted groundwater exhibits the highest basicity and electrical conductivity.

Water type*	Sampling Site	Number of Sample	Temperature (oC)	pH	Eh (mV)	DO (mg/L)	EC (µS/cm)	Dissolved concentration (mg/L) Na+	K+	Ca2+	Mg2+	Cl-	SO42-	NO3-	HCO3-
Rainwater (Ca, Cl-NO3)*	near SNR2	1	24.5	6.27	322	8.1	58.7	3.5	1.1	4.3	1.5	4.4	0.9	5.4	7.6
River water (Ca,Cl-NO3)	SNR2	3	24.4	7.60	310	7.4	297.3	36.7	9.3	33.0	6.6	49.9	42.3	13.3	71.2
Polluted Groundwater (Ca,Cl-NO3)	SNG1&2	3	23.9	7.02	392	8.8	610.1	25.7	10.3	94.5	3.2	19.0	65.9	82.2	137.2
Unpolluted Groundwater (Ca-HCO3)	SNG3	2	24.9	7.67	501	6.2	232.5	6.5	1.9	27.2	3.5	3.9	32.8	1.9	61.1

* Water type is classified as hydrologic cycle and the dissolved composition.

As indicated in Table 2, the pH becomes progressively more basic in the following sequence: rainwater, unpolluted groundwater, river water, and polluted groundwater. EC levels increase in the order of rainwater (58.7 µS/cm), unpolluted groundwater (232.5 µS/cm), river water (297.3 µS/cm), and polluted groundwater (610.1 µS/cm), with polluted groundwater exhibiting the highest conductivity. When substituting EC for TDS, the dissolved components in polluted groundwater are 10 times higher than in rainwater and 2.6 times higher than in unpolluted groundwater.

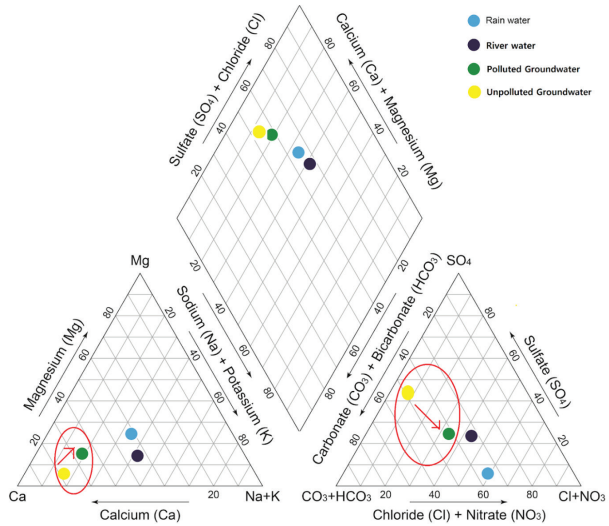


Figure 9: Piper diagram showing the chemical composition of four water types.

The Piper Diagram provides an efficient graphical method for distinguishing relevant analytical data, aiding in the understanding of the sources of dissolved constituents in water and the hydrogeochemical evolution of groundwater. As shown in Figure 9, rainwater, river water, and polluted groundwater are characterized by Ca–Cl–NO_3 facies, indicating high concentrations of dissolved chemical components, whereas unpolluted groundwater shows a Ca–HCO_3 facies. The contrast between polluted (Ca–Cl–NO_3) and unpolluted groundwater (Ca–HCO_3) is further supported by nitrate concentrations in Table 2 and relative anion ratios in Figure 9. Unpolluted groundwater is classified as Ca–HCO_3 type, typically reflecting background water quality influenced by water–rock interactions. The polluted groundwaters are the Ca-NO3-(Cl)-HCO3 type, suggesting local contamination from agricultural activities (fertilizers and/or manure). The shift in major anions from bicarbonate to nitrate and/or chloride is likely due to groundwater pollution from agricultural activities.[15] The hydrogeochemical evolution of groundwater is primarily influenced by natural processes, such as water–rock interactions, and anthropogenic impacts, including agricultural practices and, to a lesser extent, domestic effluents (Figure 9). Notably, the horizontal groundwater flow direction toward the Tancheon River aligns with the shift in water type from Ca–HCO_3 (unpolluted groundwater) to Ca–NO_3(–Cl) (polluted groundwater). Therefore, we speculate that river water quality from polluted groundwater by the influence of intensive agri-

DOI: 10.36838/v8i3.47

cultural activities may be exacerbated along the groundwater flow direction in the study area.

2. Nitrate Pollution in Groundwater:

In our four sampling sites—SNR2, SNG1, SNG2, and SNG3—nitrate levels ranged between 1.6 and 139.2 mg/L. According to the Korean Drinking Water Standards (KDWS), nitrate levels exceeding 44 mg/L are classified as non-potable.

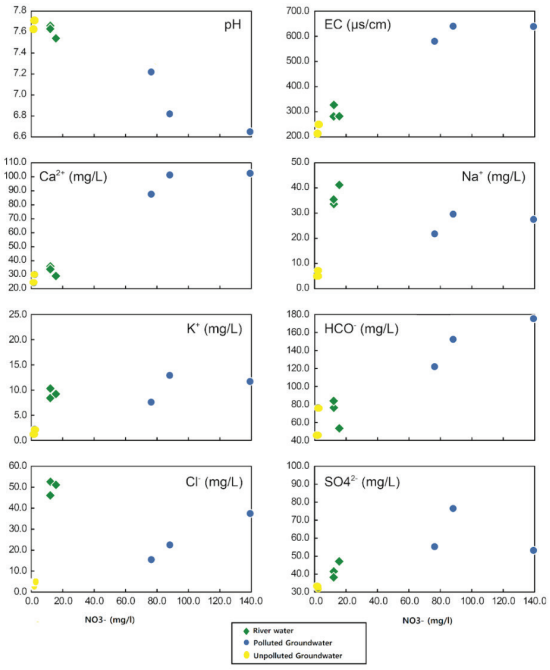


Figure 10: Correlation between NO_3^- concentration and selected physicochemical parameters (pH, EC, Ca^{2+}, Na^+, K^+, Mg^{2+}, HCO_3^-, Cl^-, SO_4^{2-}) in river water, and in polluted and unpolluted groundwater from the study areas; rainwater data excluded by default. In polluted groundwater, EC, Ca^{2+}, HCO_3^-, and Cl^- concentrations generally increase with NO_3^- levels, while pH shows an inverse correlation. River water exhibits a positive correlation between EC, Ca^{2+}, K^+, SO_4^{2-}, and NO_3^-, with pH negatively correlated.

The strong correlations observed between NO_3^- and EC also suggest that anthropogenic influence plays a significant role in EC increase, nitrate being a major contributing factor. The general increase of Ca^{2+} and Mg^{2+} to NO_3^- in river waters and polluted groundwaters indicates their likely origin from lime ($CaCO_3$) applied to farmlands with nitrogen fertilizers. On the contrary, the higher concentrations of Na^+ and K^+ in river waters and groundwaters likely have their origins in manure and sewage leakage from residential districts. The chloride ions most plausibly originate from either the addition of NaCl, $CaCl_2$ (snowmelt), or fertilizers such as K_2MgSO_4, KCl, and NH_4SO_4, as it is abundantly used in the area.

NO_3^- originates from the oxidation (nitrification) of dissolved ammonium ions (NH_4^+).[16, 17] Additional sources of ammonium ions include nitrogen-based agricultural fertilizers,[18,19] domestic wastewater ($CO(NH_2)_2$) from residential areas,[20] and the oxidation of organic matter (or organic carbon), which also contributes to the rising NO_3^- concentration in groundwater. These processes can be illustrated through the following series of reactions.

$$CO(NH_2)_2 \Rightarrow NH_3 \Rightarrow NH_4^+ \quad (1)$$

$$NH_4^+ + 2O_2 \Rightarrow NO_3^- + H_2O + 2H^+ \quad (2)$$

$$C_{106}H_{263}O_{110}N_{16}P + 138O_2 \Rightarrow 106CO_2 + 16NO_3^- + HPO_4^{2+} + 122H_2O + 18H^+ \quad (3)$$

These processes deviate from the typical trend of pH elevation observed during water–rock interactions, instead indicating a tendency for pH to decrease in groundwater due to increasing hydrogen ion (H^+) concentrations. This acidification significantly reduces the concentration of dissolved bicarbonate (HCO_3^-) ions, reflecting a shift in carbonate equilibrium dynamics within the aquifer system.

It is important to note that some hydrochemical parameters that display negative correlations with NO_3^- in polluted groundwater types are generally associated with the denitrification process (see equation 4), which produces enhanced alkalinity as well as increased concentrations of reduced species such as Fe, Mn, and NH_4^+.[21,22] However, manure and/or sewage are likely sources of sulfate in groundwater as sulfate concentrations rise alongside increasing nitrate levels.

$$5FeS_2 + 14NO_3^- + 4H^+ \rightarrow 7N_2 + 10SO_4^{2-} + 5Fe^{2+} + 2H_2O \quad (4)$$

In the case of SO_4 and NO_3, denitrification is likely insignificant due to the highly sandy composition of the aquifer materials, which maintains aerobic conditions. As a result, nitrate concentrations exhibit a positive correlation (Figure 6).[23]

Our analysis of nitrate correlation with the selected physicochemical parameters indicated both positive and negative trends, depending on the variable. Furthermore, we examined the origins of the prevalence of nitrates in the water bodies.

The sources of nitrate in polluted groundwaters are largely classified as non-point (chemical fertilizers) and point sources (septic tanks, sewage systems, and manure piles).[10] In agricultural areas, these origins are mainly associated with agricultural and residential activities. Nitrogen isotope study with geochemical data will offer a critical means for source identification of nitrate in groundwater and is used as an effective indicator of the hydrochemical process (i.e., denitrification) in aquifers. [15, 24-28]

3. The Hydrogeochemical Assessment of River Water and Groundwater Mixing Through the Riverbank Filtration System in Aquifers:

The riverbank filtration water (SNF1) in the study area is analyzed in Table 1. Given that SNF1 combines river water and groundwater, whose characteristics are represented by hydrogeochemical factors measured at each site, a binary mixing model can be applied to estimate the relative contributions of river water and groundwater in RBF.[29, 30] This method is used to evaluate the mixing ratio that produces RBF water.

The mixing ratio was calculated using the following binary mixing model. Where fa represents the 19 mixing ratio of end member "a," f_b represents the mixing ratio of end member "b," and for each factor, C_m represents the measured concentration of the sample. C_a represents the measured value of end member "a," and C_b represents the measured concentration of end member "b."

$$f_a + f_b = 1\ ,\ f_b = 1 - f_a$$

$$C_m = f_a C_a + f_b C_b = f_a C_a + (1 - f_a) C_b$$

$$f_a = \frac{C_m - C_b}{C_a - C_b}$$

In this study, the 3 conditions are determined as follows. (1) Tracer decision: Tracer (ion: Cl, SiO_2). (2) River water: end-member decision SNR2 (mean concentration of 3 samples). (3) Groundwater: end-member decision SNG1 and SNG2 (mean concentration of 3 samples). SNG3 (unpolluted groundwater) has been excluded from the investigation due to the limitation of being too far from the RBF. Table 3 summarizes Cl and SiO_2 concentrations in two end-members (the mean of river water and groundwater) and riverbank filtration water (SNF1).

Table 3: Chloride (Cl^-) and silica (SiO_2) concentrations show that SNF1 values closely match river water (SNR2) but differ substantially from groundwater (SNG1 and SNG2), indicating river water as the dominant source of recharge.

Classification	Riverbank Filtration Water (SNF1)	River Water (SNR2)		Groundwater (SNG1 and SNG2)	
	Mean	Number of Samples	Mean	Number of Samples	Mean
Cl (mg/l)	48.4	3	49.9	3	25.2
SiO_2 (mg/l)	5.3	3	5.1	3	17.2

Table 4 represents the results of calculating the contribution of river water to riverbank filtration water (SNF1). Based on Cl and SiO_2 concentrations, the river water contribution rates to SNF1 are 93.9% and 98.1%, respectively. Factors that affect the mixing ratio of SNF1 include the distance between river water and groundwater, the direction from which the water is collected, and the water level. It can be seen that the influence of river water is dominant due to the location of the SNF1 in the study area (Figure 7). Moreover, it is also presumed that there was a reflection of the flood season.

Table 4: Results of calculating the contribution of river water to riverbank filtration water (SNF1). River water contributed the majority of recharge to SNF1, with mixing ratios of 93.9% (Cl^-) and 98.1% (SiO_2), confirming that river water is the dominant source of riverbank filtration water in the study area.

Tracer	Mixing Ratio of River Water to Riverbank Filtration Water (SNF1)
Cl	93.9%
SiO_2	98.1%

As the calculations of the mixing ratio display, in SNF1, the estimated nitrate concentration is 14.6 mg/l, based on SiO_2, and 17.5 mg/l, based on CI. However, our samples presented a 5.4 mg/l nitrate concentration, a level well below the KDWS of 44 mg/l. Since the nitrate concentration in SNF1 cannot be solely attributed to the mixing of river and groundwater, further investigation is required to elucidate the mechanisms responsible for nitrate reduction.

One of the primary limitations of this study was the selection of tracers used in hydrochemical mixing methods, as we relied on chemical tracers, such as major ions and specific solutes, instead of isotopic tracers. Various tracers can be employed in hydrochemical mixing approaches to estimate the proportion of river water and groundwater. Common tracers include major ions (e.g., Cl^-, SO_4^{2-}), stable isotopes (e.g., δD), and environmental tracers (e.g., dissolved organic carbon or specific conductance). The choice of tracer depends on the characteristics of the water sources and the specific objectives of the study. While chemical tracers provide valuable insights into the mixing process, they also present certain limitations. Unlike isotopes, which often offer a more precise distinction between water sources due to their unique signatures, chemical tracers can be influenced by geochemical reactions, dilution effects, and temporal variability, complicating the interpretation of results. The absence of isotopic data was a key limitation in this study, as it may have reduced the accuracy of quantifying the proportion of river water and groundwater within the mixed system.

Another limitation of this study was the exclusion of temperature as a factor in the mixing ratio calculations. Temperature is considered a highly conservative tracer, similar to oxygen isotopes or Cl^- ions, and can be useful in assessing mixing dynamics. However, its influence was less pronounced during the study period compared to winter, when the greater contrast between river water and groundwater temperatures would enhance its applicability as a tracer. During the season in which this study was conducted, temperature variations were relatively minor, reducing its effectiveness in accurately determining mixing ratios.

The mixing ratios of river water and groundwater contributing to riverbank filtration water (SNF1) were estimated based on variations in chemical composition. However, the temporal and spatial aspects of this variability were not comprehensively analyzed, and the findings should therefore be considered preliminary. A reliable assessment of riverbank filtration mixing ratios requires detailed characterization of the temporal and spatial concentrations of key environmental tracers, including stable isotopes, in both river water and groundwater within the study area.

River water sampling should be conducted at least monthly to capture seasonal and hydrological fluctuations, while groundwater sampling should occur semi-annually, particularly during the flood and dry seasons, to account for significant hydrological changes. Spatially, groundwater should be sampled at representative locations across the study area, considering variations in depth, to ensure that the heterogeneity of the groundwater system is adequately captured. Additionally, re-

DOI: 10.36838/v8i3.47

gional differences in aquifer properties, flow patterns, and recharge rates should be taken into account to improve the spatial resolution of the analysis. On the other hand, the most frequent problem in the long-term operation of riverbank filtration water supply facilities is that complaints arise due to ground subsidence or decreased groundwater volume due to the falling groundwater level in the surrounding area, and it is expected to continue to be a social issue, especially when there is a shortage of water sources due to drought. Subsequently, the hydrogeochemical mixing methods of calculating the contribution of groundwater to riverbank filtration water can be used as a basis for calculating the appropriate pumping amount of RBF and preventing ground subsidence.

■ Conclusion

This study aimed to identify the level of impact of agricultural activities on the quality of alluvial groundwater, delving into the various hydrogeochemical alterations the sampled river areas underwent from the farmland chemicals nearby. The primary determinant of water quality in this study was the presence of impurities, which were narrowed to three origins: 1) synthetic nitrate-containing fertilizers, 2) organic composts, and 3) manures and domestic sewage. The study sampled four sites along the Tancheon River, each showing varying degrees of contamination from these sources. The river channel (SNR2) contained moderate concentrations of calcium (33.0 mg/L), bicarbonate (71.2 mg/L), and a mild level of sulfate (42.3 mg/L), silica (5.1 mg/L). No signs of immense pollution were found. In contrast, the branch of the channel (SNR1) near residential areas showed noticeable differences. Characterized by a greenish color, strong odor, and a much lower water flow compared to the main channel, SNR1 was impacted by both residential waste and agricultural runoffs. A substantial number of chemicals existed in unpotably high concentrations, namely, potassium (21.9 mg/L). Among all the sites, the polluted groundwater (SNG1 and SNG2) contained the most impurities (EC 610.1 μS/cm). These sites were heavily impacted by adjacent agricultural activities and agricultural runoff, contributing to their high concentration of contaminants, particularly nitrates, which exceeded the Korean Drinking Water Standards (KDWS).

Thus, our data indicates that agricultural activities in the alluviums, especially the over-use of chemical fertilizers and composts, threaten the water supply from alluvial aquifers. Geologically, such an impact is more serious in more permeable aquifers, as contaminants can more easily travel through porous geological formations. For future research, we believe that a systematic hydrogeochemical study is essential to understand the alluvial groundwater system and to prepare an appropriate measure for water quality protection.

Furthermore, to build upon the findings of this study, further investigation into the phenomenon of ground subsidence is required. This paper emphasizes the role of RBF in nitrate reduction; however, subsidence caused by excessive groundwater extraction remains a critical concern, particularly in regions with highly permeable alluvial formations in South Korea. Incorporating geotechnical monitoring and long-term groundwater level analysis in future studies would help quantify the extent of land deformation and better illustrate the physical impacts of unsustainable groundwater use. Future research can provide a more comprehensive evaluation of RBF not only as a water purification method but also as a strategy to reduce groundwater dependency.

■ Acknowledgments

We would like to share our gratitude towards the Environmental Geochemistry Laboratory at Korea University for their invaluable support and expertise in conducting the Inductively Coupled Plasma (ICP) analysis. Their contribution played a crucial role in the accuracy of the sample data used in this study.

■ References

1. Ford, C. *More than 4 billion people don't have access to clean water at home*. Vox. https://www.vox.com/future-perfect/367224/clean-drinking-water-access-four-billion-science-study#.
2. *Food and Agriculture Organization of the United Nations. Water Scarcity | Land & Water | Food and Agriculture Organization of the United Nations | Land & Water | Food and Agriculture Organization of the United Nations. www.fao.org. https://www.fao.org/land-water/water/water-scarcity/en/.*
3. Scientific. The United Nations World Water Development Report 2019; United Nations, 2019. https://doi.org/10.18356/0d-8fe383-en.
4. Umezawa, Yu. "Sources of Nitrate and Ammonium Contamination in Groundwater under Developing Asian Megacities - PubMed." PubMed, 1 Jan. 2008, https://pubmed.ncbi.nlm.nih.gov/18533227/. 5. Hendarto, H., and J. R. Standing. Influence of Groundwater Extraction on Land Subsidence in Jakarta.2019, pp.1–8,https://www.ecsmge-2019.com/uploads/2/1/7/9/2179080/0511r-ecsmge-2019_hendarto.pdf
6. McGrath, M. Water Extraction and Weight of Buildings See Half of China's Cities Sink. www.bbc.com. April 18, 2024. https://www.bbc.com/news/science-environment-68844731.
7. Omolere, M.P. Why Are Chinese Cities Sinking? | Earth.org. https://earth.org/why-are-chinese-cities-sinking-a-comprehensive-analysis-of-causes-effects-and-solutions/.
8. Lee. 경북도, "비점오염원 집중관리로 하천 수질개선". 브레이크뉴스, 29 Nov. 2021, http://www.breaknews.com/851670.
9. Ray, C.; Melin, G.; Linsky, R. B. Riverbank Filtration; Springer Science & Business Media, 2006. 10. JARAMILLO, M. Riverbank Filtration: An Efficient and Economical Drinking-Water Treatment Technology. Dyna (Colombia) 2012, 79 (171), 148–157. 11. Noh, J.-H.; So, S.-H.; Park, J.-W.; Kim, S.-Y.; Song, K.-G.; Choi, J.; Kim, G.-B.; Son, H.; Kim, H.; Maeng, S.-K. An Assessment of the Effectiveness of Riverbank Filtration in a Sewage Plant Effluent-Impacted River Using a Full-Scale Horizontal Well. Water 2022, 14 (12), 1873. https://doi.org/10.3390/w14121873. 12. Glynn P.D., Plummer L.N., 2005. Geochemistry and the understanding of groundwater systems. Hydrogeol. J., 13, 263–287. 13. Min J.H., Yun S.T., Kim K., Chae G.T., Kim, H.S., 2002. A study on nitrate contamination of alluvial groundwater in the Nakdong River basin, Korea. Geosciences J., 6, 35-46. 14. Kreitler, C.W. and Jones, D.C., 1975, Natural soil nitrate: the cause of the nitrate contamination of groundwater in Runnels County, Texas. Ground Water, 13, 53–61. 15. Zilberbrand, M., Rosenthal, E., and Shachnai, E., 2001, Impact

of urbanization on hydrochemical evolution of groundwater and on unsaturated-zone gas composition in the coastal city of Tel Aviv, Israel. J. Contam. Hydrol. 50, 175–208. 16. Nkotagu, H., 1996. Origins of high nitrate in groundwater in Tanzania. J. African Earth Sciences 21, 471–478.
17. Korean Research Institute of Geoscience and Mineral Resources; GeoGreen21; Korea Environment Institute. 서울특별시 지하수관리 계획: 2015–2024; Seoul Metropolitan Government: Seoul, 2016. https://sejong.nl.go.kr/search/searchDetail.do?rec_key=SH1_KMO201747124&kwd=
18. Min J.H., Yun S.T., Kim K., Kim H.S., Kim D.J., 2003. Geologic controls on the chemical behavior of nitrate in riverside alluvial aquifers, Korea. Hydrol. Processes, 17, 1197-1211.19. Chae G.T., Kim K., Yun S.T., Kim K.H., Kim S.O., Choi B.Y., Kim H.S., Rhee C.W., 2004. Hydrogeochemistry of alluvial groundwaters in an agricultural area of Korea: an implication for groundwater contamination susceptibility. Chemosphere, 55, 369-378.
20. Teppei Fukada, Kevin M. Hiscock, Paul F. Dennis, 2004, A dual-isotope approach to the nitrogen hydrochemistry of an urban aquifer, Applied Geochemistry, Volume 19(5), 709-719.
21. Robertson, W.D., Russell, B.M., and Cherry, J.A., 1996, Attenuation of nitrate in aquitard sediments of southern Ontario. Journal of Hydrology, 180, 267–281.
22. Ptacek, C.J., 1998, Geochemistry of a septic-system plume in a coastal barrier bar, Point Pelee, Ontario, Canada. Journal of Contaminant Hydrology, 33, 293–312.
23. Exner, M.E. and Spalding, R.F., 1994, N-15 identification of non-point sources of nitrate contamination beneath cropland in the Nebraska Panhandle: two case studies. Applied Geochemistry, 9, 73-81.
24. Mariotti, A., Landreau, A. and Simon, B., 1988, 15N isotope biogeochemistry and natural denitrification process in groundwater: application to the chalk aquifer of northern France. Geochimica et Cosmochimica Acta, 52, 1869–1878.
25. Bottcher, J., Strebel, O., Voerkelius, S. and Schmidt, H.L., 1990, Using isotope fractionation of nitrate-nitrogen and nitrate-oxygen for evaluation of microbial denitrification in a sandy aquifer. Journal of Hydrology, 11, 413–424.
26. Aravena, R., Evans, M.L., and Cherry, J.A., 1993, Stable isotopes of oxygen and nitrogen in source identification of nitrate from septic systems. Ground Water, 31, 180–186.
27. Wassenaar, L., 1995, Evaluation of the origin and fate of nitrate in the Abbotsford aquifer using the isotopes of δ15N and δ18O in NO3- Applied Geochemistry, 10, 391–405.
28. Fogg, G.E., Rolston, D.E., Decker, D.L., Louie, D.T., and Grismer, M.E., 1998, Spatial variation in nitrogen isotope values beneath nitrate contamination sources. Ground Water, 36, 418–426.
29. Crandall, C. A., Katz, B. G., & Hirten, J. J., 1999, Hydrochemical evidence for mixing of river water and groundwater during high-flow conditions, lower Suwannee River basin, Florida, USA. Hydrogeology Journal, 7(5), 454-467
30. Kalbus, E., Reinstorf, F., & Schirmer, M., 2006, Measuring methods for groundwater-surface water interactions: a review. Hydrology and Earth System Sciences Discussions, 10(6), 873-887.

Authors

William J. Hong is an 11th grader currently situated in Seoul, Korea. He attends Chadwick International School and has an interest in geology, hydrogeology, and other topics.

Christian Jookyung Han is a 10th grader currently situated in Guangzhou, China. He attends the Canadian International School of Guangzhou and has interests in philosophy, apologetics, law, history, political science, and others.

DOI: 10.36838/v8i3.47

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Measurement Uncertainty – Based Assessment of Laboratory Tests for Ovarian Cancer-Related Tumor Markers

Youn Joo Lee[1], Woochang Lee[2]

1) Chadwick International School, 45, Art Center-daero 97 beon-gil, Yeonsu-gu, Incheon, 22002, South Korea; yjlee3731@gmail.com
2) Department of Laboratory Medicine, Asan Medical Center, University of Ulsan College of Medicine, 88, Olympic-ro 43-gil, Songpa-gu, Seoul, 05505, South Korea

ABSTRACT: Measurement uncertainty is an intrinsic parameter showing the dispersion of numerical values. Cancer antigen 125 (CA 125) and human epididymis protein 4 (HE4), along with the risk of ovarian malignancy algorithm (ROMA) value, are widely used tumor markers to determine the risk of ovarian cancer. Although predefined cutoffs interpret tumor markers, their intrinsic uncertainties produce an ambiguous range for risk assessment. This study assessed the measurement uncertainty of CA 125, HE4, and ROMA values to determine reliable reporting ranges for these markers, which suggest the optimal reporting strategy. Results indicated that the reliable reporting range was below 32.5 or above 37.9 U/mL for CA 125, below 64.5 or above 76.6 pmol/L for premenopausal HE4, below 128.9 or above 153.1 pmol/L for postmenopausal HE4, below 6.7 or above 8.4% for premenopausal ROMA value, and below 22.7 or above 28.6% for postmenopausal ROMA value. Analysis of real patient data revealed that approximately 4–6% of test results fell outside of the reliable reporting range, underscoring the importance of cautious interpretation. The findings suggested that incorporating measurement uncertainty into clinical practice could enhance the reliability of tumor marker results, potentially improving patient management and decision-making.

KEYWORDS: Biochemistry, Medical Biochemistry, Tumor Marker, Uncertainty Measurement, Reliable Reporting Range.

Introduction

Measurement uncertainty (MU) is a non-negative value that indicates the dispersion of the data and is a commonly used concept in various industries and engineering fields. Although most of the information provided by clinical laboratories has a quantitative character, the importance of uncertainty has been raised relatively recently. In practice, it represents the range within which the true value of the measurand is expected to fall, given a specified level of confidence. While the word "uncertainty" is often associated with doubt in everyday language, in the context of medical laboratories, it instead conveys greater assurance about the reliability of a reported measurement. Uncertainty estimation can be performed using two approaches: the top-down and bottom-up methods. The Guide to the Expression of Uncertainty in Measurement (GUM), published in 1996, states that the bottom-up approach is the standard method for estimating MU, which involves the identification of all sources of uncertainty in the measurement procedures, estimation of their magnitudes, and calculation of the combined uncertainty according to the law of error propagation.[1,2] Meanwhile, the MU guideline for laboratory medicine recommends that the top-down approach using imprecision data obtained from internal quality control (IQC) results is practical and particularly well-suited to closed measurement systems (instruments, calibrators, and reagents from the same vendor).[3]

Globally, ovarian cancer is the eighth most common cancer in women and the second most common gynecologic cancer in the United States, and ovarian cancer causes more deaths than any other cancer of the female reproductive system.[4] Its five-year survival rate is about 50%, and it is markedly influenced by stage.[5] That is, early detection is very important for the better management of patients. In this regard, the importance of useful biomarkers of ovarian cancer has been stressed for several decades. CA 125 (cancer antigen 125) is a well-known biomarker for ovarian cancer and has been widely used for monitoring and recurrence detection of ovarian cancer.[6,7] Also new ovarian cancer marker, HE4 (human epididymis protein 4), has been developed and introduced into clinical practice.[8-11]

The risk of ovarian malignancy algorithm (ROMA) value is another valuable diagnostic marker to predict the risk of epithelial ovarian cancer, with higher overall accuracy than HE4 and CA-125, and is characterized as having higher specificity and negative predictive value.[12] ROMA value is calculated from the predictive index (PI) derived from CA 125 and HE4 in premenopausal and postmenopausal status and shows the percent risk for epithelial ovarian cancer.[13,14]

Most of the tumor markers are interpreted based on cutoffs discriminating low risk vs. high risk, or reference intervals derived from the healthy population.[15] However, those cutoffs are a relative compromise with intrinsic uncertainty, rendering the interpretation of test results prone to errors. An appropriate understanding of the possible impact caused by MU could be helpful for the optimal use of tumor markers in clinical practice. In this regard, the estimation of MU and establishment of a reliable reporting range could have practical utility in the interpretation of tumor markers. As well, the establishment of MU within a narrow range is crucial in that this can contribute to increasing test reliability and ultimately reduce the proportion of test results falling outside of the reliability range as much as possible. The previous article reported a permissible

DOI: 10.36838/v8i3.56

MU limit of 15.97% for tumor markers.[16] The purpose of this study was to estimate the MU range of ovarian cancer-related tumor markers with a top-down approach, to establish their reliable reporting ranges using MU, and, in addition, to assess the proportion of test results of tumor markers not within the reliable reporting range when applied to real patients' data.

Methods

Reference intervals of tumor markers:

Tumor markers CA 125 and HE4 were measured with the Abbott Alinity i system using dedicated calibrators and reagents from Abbott (Abbott Laboratories, Abbott Park, IL, USA). The assay systems were based on chemiluminescent microparticle immunoassay.

A cut-off value is a specific threshold used to interpret laboratory test results. Test results above, below, or within a certain cut-off help categorize patients for clinical decisions such as diagnosis or risk assessment. IQC is a routine laboratory practice that uses control samples with known concentrations to monitor the consistency and reliability of analytical instruments and procedures.

Based on the package insert's date, 94.4% of healthy female subjects had CA 125 values at or below 35.0 U/mL, which was the reference interval for the CA 125 assay in the hospital where this study was conducted. As for the HE4 assay, 96 % of the healthy premenopausal subjects had an HE4 assay value at or below 70 pmol/L, and 95% of the healthy postmenopausal subjects had an HE4 assay value at or below 140 pmol/L. Based on this data, the reference intervals of the HE4 assay were at or below 70 pmol/L and at or below 140 pmol/L for premenopausal and postmenopausal female subjects, respectively. Eventually, CA 125 over 35.0 U/mL, HE4 over 70 pmol/L in premenopausal women, and HE4 over 140 pmol/L in postmenopausal women suggested a high risk of epithelial ovarian cancer.

ROMA value (%) was obtained using predictive index (PI) that was calculated with CA 125 and HE4, both in premenopausal and postmenopausal status.[13]

Premenopausal PI = -12.0 + 2.38LN[HE4] + 0.0626LN[-CA125]

Postmenopausal PI = -8.09 + 1.04LN[HE4] + 0.732LN[-CA125]

ROMA value (%) = $(e^{PI} / 1 + e^{PI}) \times 100$

The cut-off point of the ROMA value in the Abbott assays was 7.4% in premenopausal women and 25.3% in postmenopausal women. Consequently, ROMA values of 7.4% and over in premenopausal and 25.3% and over in postmenopausal women suggested a high risk of epithelial ovarian cancer, while the values of less than 7.4% and 25.3% suggested a low risk.

Estimation of MU for CA 125, HE4, and ROMA value:

IQC was performed using dedicated three-level QC material provided by the manufacturer Abbott (Alinity i; Abbott Diagnostics, Chicago, IL, USA), and IQC data were collected over one year, specifically from May 2023 to April 2024. We used a single immunoassay measurement system and two concentration levels of IQC materials (Alinity i CA125 control and Alinity i HE4 control), including multiple reagent lots during the study period. The coefficient of variation (CV) is the standard deviation expressed as a percentage of the mean.[17] And maximal CV was determined to be the standard MU of each assay. The calculation process of MU was performed as described in the previous study. That is, IQC values obtained from each reagent lot were collected separately, and standard uncertainties for each lot subgroup were combined to obtain the overall uncertainty.[18] For each assay, expanded uncertainty was determined by multiplying the standard MU by 1.65, the coverage factor, used for a one-sided 95% level of confidence (**Figure 1**).

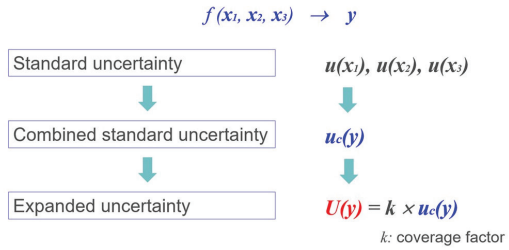


Figure 1: Overview of measurement uncertainty estimation. The scheme of measurement uncertainty from standard uncertainty to expanded uncertainty is depicted.

For the uncertainty estimation of ROMA value, standard measurement uncertainties of CA 125 and HE4 were error propagated to have combined uncertainty, and the expanded uncertainty was determined by multiplying the combined uncertainty by 1.65, the coverage factor used for a one-sided 95% level of confidence.

Calculation of the reliable reporting range for CA 125:

Reliable reporting ranges for CA 125 were assessed with the estimated expanded uncertainty of CA 125. If the measured CA 125 was smaller than or equal to 35.0 U/mL, the measured CA 125 plus expanded uncertainty should be smaller than 35.0 U/mL to be reliably at low risk. If the measured CA 125 was greater than 35.0 U/mL, the measured CA 125 minus expanded uncertainty should be greater than 35.0 U/mL to be reliably at high risk. With these calculations, a reliable reporting interval for CA 125 was determined.

Calculation of the reliable reporting range for HE4:

Reliable reporting ranges for HE4 were determined with the estimated expanded uncertainty of HE4 in premenopausal and postmenopausal women separately. For premenopausal women, if the measured HE4 was smaller than or equal to 70 pmol/L, the measured HE4 plus expanded uncertainty should be smaller than 70 pmol/L to be reliably at low risk. If the measured HE4 was greater than 70 pmol/L, the measured HE4 minus expanded uncertainty should be greater than 70 pmol/L to be reliably at high risk. For postmenopausal women, if the measured HE4 was smaller than or equal to 140 pmol/L, the measured HE4 plus expanded uncertainty should be smaller than 140 pmol/L to be reliably at low risk. If the measured HE4 was greater than 140 pmol/L, the measured

HE4 minus expanded uncertainty should be greater than 140 pmol/L to be reliably at high risk. With these calculations, a reliable reporting interval for HE4 in postmenopausal women was determined.

Calculation of the reliable reporting range for the ROMA value:

The reliable reporting ranges for the ROMA value were determined by the comparable method as HE4. For premenopausal women, the calculated ROMA value smaller than 7.4% plus expanded uncertainty should be less than 7.4 % to be reliably at low risk. While the calculated ROMA, larger than 7.4% minus expanded uncertainty, should be greater than 7.4% to be reliably at high risk. For postmenopausal women, adding expanded uncertainty to the calculated ROMA value, smaller than 25.3%, should be smaller than 25.3% to be reliably at low risk. On the other hand, extracting expanded uncertainty from the calculated ROMA value larger than 25.3% should be greater than 25.3% to be reliably at high risk.

Proportion of CA 125, HE4, and ROMA test results of real patients not within the reliable reporting range:

The tumor markers of real patients were extracted from data sources already established from September 2023 to August 2024 in the hospital where this study was conducted. As clinical laboratories are not knowledgeable about the menopausal status of each patient, HE4 and ROMA values were analyzed for both premenopausal and postmenopausal respectively.

The proportions of the test results that did not fall within the reliable reporting range were obtained by calculating the aspect of the above and below cut-off value of CA 125, HE4, and ROMA value, respectively.

Results and Discussion

Results:

Estimation of MU for CA 125, HE4, and ROMA value:

Based on the internal quality control (IQC) data for a year, standard measurement uncertainties of CA 125 and HE4 were 4.59 % and 5.21%, respectively. And expanded uncertainties of CA 125 and HE4 were 7.57% and 8.60%, respectively. Also, uncertainty estimation of the ROMA value was performed as follows. With the error propagation of standard measurement uncertainties of CA 125, 4.59%, and HE4, 5.21%, the combined uncertainty of the ROMA value was determined to be 6.94%, and the corresponding expanded uncertainty was 11.46%.

Reliable reporting range for CA 125:

The reliable reporting range for CA 125 was below 32.5 U/mL and above 37.9 U/mL. That is, CA 125 between 32.5 U/mL and 37.9 U/mL was deemed to be within the borderline range, so their risk assessment was not reliable by a single measurement (Figure 2).

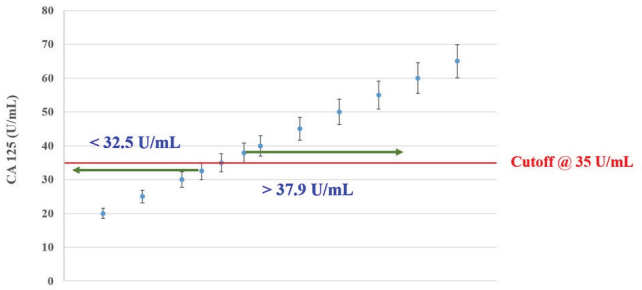


Figure 2: Measurement uncertainty-based estimation of the cutoff for CA 125. Each data point denotes representative CA 125 values such as 20, 25, 30, 35, 40, etc. U/mL with corresponding expanded uncertainties. For the CA 125 value of 32.5 U/mL, 'CA 125 + expanded uncertainty' is close to 35 U/mL, and for the CA 125 value of 37.9 U/mL, 'CA 125 – expanded uncertainty' is close to 35 U/mL. Therefore, the reliable reporting range for CA 125 is below 32.5 U/mL or above 37.9 U/mL.

Reliable reporting range for HE4:

Reliable reporting range for HE4 in premenopausal women was below 64.5 pmol/L or above 76.6 pmol/L. For postmenopausal women, the reliable reporting range of HE4 was below 128.9 pmol/L or above 153.1 pmol/L. Thus, HE4 between 64.5 pmol/L and 76.6 pmol/L in premenopausal women, as well as between 128.9 pmol/L and 153.1 pmol/L in postmenopausal women, was regarded to be within the borderline range (**Figure 3**).

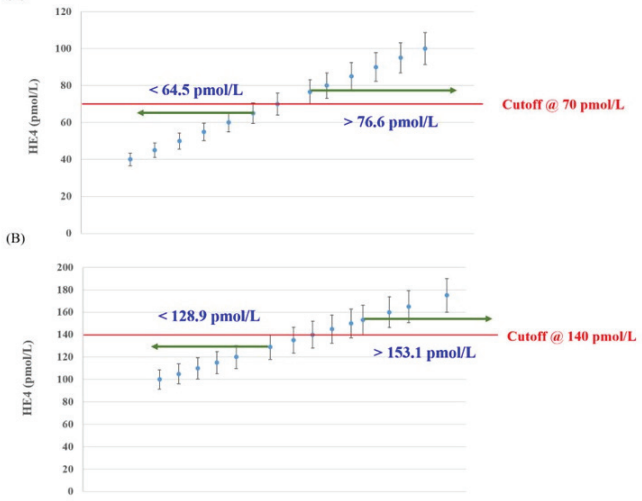


Figure 2: Measurement uncertainty-based estimation of cutoff for HE4 for (A) premenopausal and (B) postmenopausal status. (A) Each data point denotes representative HE4 values such as 40, 45, 50, 60, 70, 80, etc. pmol/L with corresponding expanded uncertainties under the assumption of premenopausal status. For the HE4 value of 64.5 pmol/L, 'HE4 + expanded uncertainty' is close to 70 pmol/L, and for the HE4 of 76.6 pmol/L, 'HE4 – expanded uncertainty' is close to 70 pmol/L. Therefore, the reliable reporting range for HE4 is below 64.5 pmol/L or above 76.6 pmol/L in premenopausal status. (B) Each data point denotes representative HE4 values such as 100, 120, 140, 150, 160, etc., pmol/L with corresponding expanded uncertainties under the assumption of postmenopausal status. For the HE4 value of 128.9 pmol/L, 'HE4 + expanded uncertainty' is close to 140 pmol/L, and for the HE4 of 153.1 pmol/L, 'HE4 – expanded uncertainty' is close to 140 pmol/L. Therefore, the reliable reporting range for HE4 is below 128.9 pmol/L or above 153.1 pmol/L in premenopausal status.

Reliable reporting range for the ROMA value:

The reliable reporting range for the ROMA value in premenopausal women was below 6.7% or above 8.4%. For

DOI: 10.36838/v8i3.56

postmenopausal women, the reliable reporting range of the ROMA value was below 22.7% or above 28.6 %. Thus, ROMA values between 6.7% and 8.4 % in premenopausal women, as well as between 22.7% and 28.6%, were considered within the borderline range (**Figure 4**).

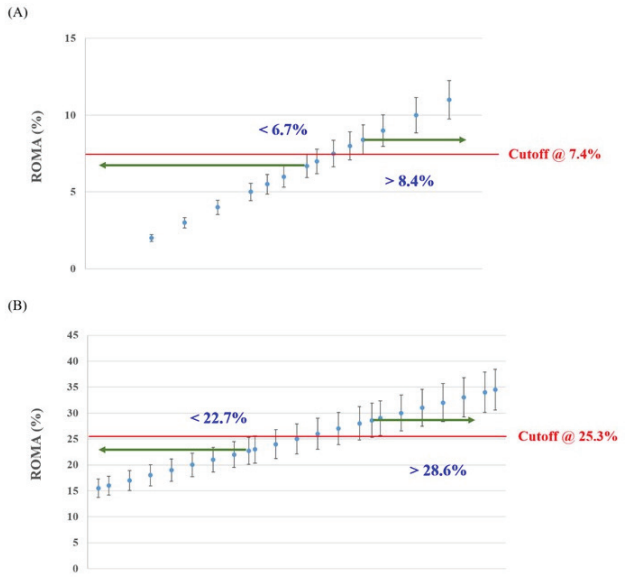


Figure 4: Measurement uncertainty-based estimation of cutoff for ROMA values for (A) premenopausal and (B) postmenopausal status. (A) Each data point denotes representative ROMA (%) values, such as 3, 4, 6, 7, 8, 9, etc. % with corresponding expanded uncertainties under the assumption of premenopausal status. For the ROMA value of 6.7%, 'ROMA + expanded uncertainty' is close to 7.4%, and for the ROMA of 8.4%, 'ROMA – expanded uncertainty' is close to 7.4%. Therefore, the reliable reporting range for ROMA is below 6.7% or above 8.4% in premenopausal status. (B) Each data point denotes representative ROMA (%) values such as 20, 22, 25, 28, 3,0, etc. % with corresponding expanded uncertainties under the assumption of postmenopausal status. For the ROMA value of 22.7%, 'ROMA + expanded uncertainty' is close to 25.3%, and for the ROMA of 28.6%, 'ROMA – expanded uncertainty' is close to 25.3%. Therefore, the reliable reporting range for ROMA is below 22.7% or above 28.6% in premenopausal status.

Reliable reporting range for CA 125 in patients' data:

3,501 CA 125 tests, consisting of 494 CA 125 tests independently performed and 3,007 ROMA value tests, were included. Among 3,501 CA 125 results, 2,195 results were below or equal to 35.0 U/mL, which was the cut-off value, and 1,306 results were above 35.0 U/mL. Among the 2,195 results, 2,121 results were below 32.5 U/mL, and therefore 74 results didn't belong to the reliable reporting range. Also, among the 1,306 results, 1,212 results were above 37.9 U/mL, which meant that 94 results were not included in the reliable reporting range. To sum up, 168 out of 3,501 (4.8%) test results could be regarded as those belonging to the borderline range.

Reliable reporting range for HE4 in patients' data:

Finally, 5,493 HE4 test results, composed of independent 2,486 HE4 tests and 3,007 ROMA value tests, were collected. Assuming that all the tests were obtained from premenopausal women, the cut-off value was determined as 70 pmol/L. Among the 4,414 results below or equal to 70 pmol/L, 4,285 results estimated below 64.5 pmol/L lay in the reliable reporting range, except 129 results out of this range. On the other hand, among the 1,079 results above 70 pmol/L, 952 results measured above 76.6 pmol/L were included in the reliable reporting range, while 127 results were not. Ultimately, 256 out of 5,493 (4.7%) test results could be regarded as those belonging to the borderline range.

Meanwhile, if all of them were obtained from postmenopausal female subjects, a 140 pmol/L cutoff value was applied. Of the 5,046 results below or equal to 140 pmol/L, 5,005 results below 128.9 pmol/L were within of reliable reporting range, and eventually, 41 results did not belong to this range. Also, among the 447 results above 140 pmol/L, 407 results above 153.1 pmol/L were included in the reliable reporting range, but 40 results were not. At last, 81 out of 5,493 (1.5%) test results were regarded as those belonging to the borderline range.

Reliable reporting range for the ROMA value in patients' data:

With the assumption of premenopausal status, the cut-off for high risk was defined as a ROMA value was 7.4% or more. A 2,204 of 3,007 test results were below 7.4%, and 803 were above or equal to 7.4%. Among the 2,204 results, 2,117 results below 6.7% were included in the reliable reporting range, while the other 87 results were not. As well, among the 803 results, 689 results above 8.4% fell within the reliable reporting range, and 114 results didn't. To sum up, 201 out of 3,007 (6.7%) test results were regarded as those belonging to the borderline range. With postmenopausal status assumed, 25.3% instead of 7.4% was applied as the cut-off value. Among the 2,328 results below 25.3%, aside from 2,244 results estimated below 22.7%, 84 results didn't belong to the reliable reporting range. Also, among the 679 results above or equal to 25.3%, 580 results above 28.6% were included in and 99 results were not included in the reliable reporting range. Consequently, 183 out of 3,007 (6.1%) test results were regarded as those belonging to the borderline range (**Figure 5**).

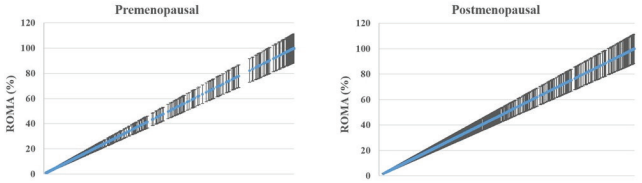


Figure 5: ROMA values in real patients' data with expanded uncertainties expressed using the formulae for premenopausal and postmenopausal status. Even for the same data pair of CA 125 and HE4, the ROMA values are different depending on the menopausal status. These differences are reflected in two figures.

Table 1: Summary of results.

Tumor Marker / Parameter	Standard Measurement Uncertainty (CV, %)	Expanded Uncertainty (%) (k=1.65)	Reliable Reporting Range
CA 125	4.59	7.57	<32.5 U/mL or >37.9 U/mL
HE4 (Premenopausal)	5.21	8.60	<64.5 pmol/L or >76.6 pmol/L
HE4 (Postmenopausal)	5.21	8.60	<128.9 pmol/L or >153.1 pmol/L
ROMA (Premenopausal)	6.94	11.46	<6.7% or >8.4%
ROMA (Postmenopausal)	6.94	11.46	<22.7% or >28.6%

Discussion:

According to this study, expanded MU of ovarian cancer related tumor markers were measured about 7.6-11.5% and about 4-6% of real patients' tumor markers were not included in reliable reporting range, which meant the assessment for ovarian cancer could not be decisively made by tumor markers level in the approximately 4-6% of all tests obtained in our laboratory.

Biomarkers, including tumor markers, play a role as risk factors and predictors of clinical outcomes and are usually presented as continuous types, which are often divided into two categories based on the cutoff point. Although this categorization provides physicians with distinctively presented criteria, the existence of a borderline around the cutoff value should be cautiously considered to avoid an unintended wrong decision. In a clinical practice setting, the test results placed near to cut-off points are not enough convincing to lead to a confirmatory decision.[19]

As well, although the cut-off value is decided in a way that test results of 95% among healthy subjects are included within the range bordered by this value, the intrinsic uncertainty of test results eventually alleviates its absolute significance. In this regard, a reliable reporting range is needed and determined using the MU obtained by CV. On the other hand, the area outside of the reliable range is defined as the borderline range, located close to the cut-off value. Test results within the borderline range enhance the necessity of retesting markers or referring to other laboratory or radiological results combined with clinical assessment.[2]

Especially because tumor markers act as the indicator of diagnosis, disease severity, progression, and treatment effects, misinterpretations of tumor markers can cause not only unwanted results such as misdiagnosis, unnecessary treatment, and economic burden, even serious or fatal consequences in comparison with other disease entities.[20,21] The application of MU in this field is supposed to be more crucial than other disease entities.

MU is an important indicator reflecting the performance characteristics of clinical testing, but its recent introduction into clinical laboratories has rendered the laboratory use limited.[1] Estimated uncertainties are often accompanied by measured reference values in reference laboratories, but they are rarely reported in routine clinical laboratory testing. This study applied MU to clinical use.

The top-down and bottom-up approaches were used for estimating MU. The study comparing the MU between the two approaches indicated that the MU obtained by the bottom-up approach was quite similar to that obtained by the top-down approach.[22] This result supports the top-down approach, used in this study, more than the bottom-up approach taken. The top-down approach is simpler and more practical in routine laboratory settings. The top-down approach is now officially endorsed by the ISO 20914, which provides practical guidance to be applied in medical laboratory settings for the purpose of estimating MU of values produced by measurement procedures intended to measure biological measurands.[23]

It estimates the MU of laboratory results by using IQC data to derive the random components of uncertainty and commercial calibrator information. It is based on the premise that long-term IQC results sufficiently reflect error factors that can change the test value, and is known to be useful in modern clinical laboratories, especially when closed measurement systems are used.[2] Therefore, we collected IQC results of CA 125 and HE4 for one year and determined the standard measurement uncertainties of CA 125 and HE4 based on CVs.

In general, when the 95% confidence interval is assumed, combined uncertainty is multiplied by the coverage factor 2 to calculate expanded uncertainty. But in case of the clinical tests used for comparison with reference interval or clinical decision limit, as in this study, a one-sided 95% level of confidence is assumed, so k=1.65 is multiplied by the combined uncertainty to obtain expanded uncertainty. As for the uncertainty estimation of the ROMA value, the combined uncertainty is calculated by the error propagation of the standard measurement uncertainties of CA 125 and HE4.[2]

Due to the different distribution of HE4 by menopausal status, the reference interval of HE4 is determined differently according to menopausal status. As uncertainty is expressed as CV percent, the bigger the test results, the bigger the uncertainty, even for the same test item. Therefore, values bigger than cutoffs have bigger uncertainty compared with those smaller than cutoffs, and, therefore, HE4 results in postmenopausal status have a broader borderline range compared with those in premenopausal status.

ROMA value is predominantly influenced by HE4 results in premenopausal status and by both CA 125 and HE4 in postmenopausal status.[24] Generally, higher CA 125 and HE4 results are obtained in postmenopausal women, and thus higher ROMA values are calculated in postmenopausal females with a broader range of borderline values. The interpretation of results far from cutoffs is generally unequivocal, but results close to cutoffs should be interpreted with caution, and the reliable reporting range should be considered as well as the measured values themselves. As MU is mostly determined by the precision parameter, CV performance characteristics, such as the imprecision of the testing method, are of critical importance for the reliability of test results. That is the reason why a more precise testing method should be utilized in clinical laboratory practices.

The current study's results indicated that about 4-6% of tumor markers of ovarian cancer were included in the borderline range. Considering that this range can be variable among laboratories depending on the assay systems or selected patients' characteristics, even values within the reliable reporting range should be prudently interpreted along with a comprehensive understanding of clinical aspects and other laboratory data.[25]

Applying MU estimation in clinical laboratories remains challenging due to the absence of standardized formulas and universally accepted limits for MU calculation and interpretation. Furthermore, communication of MU to clinicians is often inadequate, diminishing its clinical utility. Future efforts should focus on developing standardized, practical protocols for MU estimation, reporting, and interpretation, alongside

DOI: 10.36838/v8i3.56

educational initiatives to enhance awareness and effective use of MU data in clinical decision-making.

This study has several limitations. First, MU estimates were derived solely from IQC data on a single analytical platform, which may restrict the applicability of findings to other assay systems. Second, the MU estimation did not incorporate external quality assessment data or biological variation, both of which could provide complementary perspectives. Third, the inherent variability of tumor marker assays, combined with the absence of internationally standardized reference materials for ovarian cancer markers, might influence the comparability of results. Fourth, this was a single-center study, so the generalizability of the findings might be limited. Further multicenter studies would be warranted to validate and extend these results.

Conclusion

In this study, we assessed the statistical characteristics of ovarian cancer-related tumor markers using MU principles and determined a possible reliable reporting range for them. Also, the distribution of results belonging to the borderline range was assessed using real-world patients' data. The findings shown in this study could be a guide to the interpretation of test results in clinical practices, which could further help physicians to make diagnoses, establish treatment methods, and evaluate treatment outcomes. In the future, more precise MU would enhance the clinical value of tumor markers as accurate assessment tools.

Acknowledgments

This study has no grant or funding source.

References

1. BIPM, *Evaluation of measurement data — Guide to the expression of uncertainty in measurement, GUM 1995 with minor corrections*. JCGM: **2008**.
2. ISO, *Medical laboratories — Practical guidance for the estimation of measurement uncertainty*. **2019**.
3. CLSI, *Expression of measurement uncertainty in laboratory medicine, approved guideline*. CLSI EP29-A. Clinical and Laboratory Standards Institute: PA, **2012**.
4. Webb, P. M.; Jordan, S. J., Global epidemiology of epithelial ovarian cancer. *Nature reviews Clinical oncology* **2024**, *21* (5), 389-400.
5. American Cancer Society. Survival Rates for Ovarian Cancer. American Cancer Society. https://www.cancer.org/cancer/types/ovarian-cancer/detection-diagnosis-staging/survival-rates.html (accessed 25 Jan).
6. Menon, U.; Griffin, M.; Gentry-Maharaj, A., Ovarian cancer screening--current status, future directions. *Gynecologic oncology* **2014**, *132* (2), 490-5.
7. Zhang, M.; Cheng, S.; Jin, Y.; Zhao, Y.; Wang, Y., Roles of CA125 in diagnosis, prediction, and oncogenesis of ovarian cancer. *Biochimica et biophysica acta. Reviews on cancer* **2021**, *1875* (2), 188503, doi: 10.1016/j.bbcan.2021.188503.
8. Anastasi, E.; Farina, A.; Granato, T.; Colaiacovo, F.; Pucci, B.; Tartaglione, S.; Angeloni, A., Recent Insight about HE4 Role in Ovarian Cancer Oncogenesis. *International journal of molecular sciences* **2023**, *24* (13), 10479, doi: 10.3390/ijms241310479.
9. Granato, T.; Porpora, M. G.; Longo, F.; Angeloni, A.; Manganaro, L.; Anastasi, E., HE4 in the differential diagnosis of ovarian masses. *Clinica chimica acta; international journal of clinical chemistry* **2015**, *446*, 147-55.
10. Montagnana, M.; Danese, E.; Giudici, S.; Franchi, M.; Guidi, G. C.; Plebani, M.; Lippi, G., HE4 in ovarian cancer: from discovery to clinical application. *Advances in clinical chemistry* **2011**, *55*, 1-20.
11. Scaletta, G.; Plotti, F.; Luvero, D.; Capriglione, S.; Montera, R.; Miranda, A.; Lopez, S.; Terranova, C.; De Cicco Nardone, C.; Angioli, R., The role of novel biomarker HE4 in the diagnosis, prognosis and follow-up of ovarian cancer: a systematic review. *Expert review of anticancer therapy* **2017**, *17* (9), 827-839.
12. Matsas, A.; Stefanoudakis, D.; Troupis, T.; Kontzoglou, K.; Eleftheriades, M.; Christopoulos, P.; Panoskaltsis, T.; Stamoula, E.; Iliopoulos, D. C., Tumor Markers and Their Diagnostic Significance in Ovarian Cancer. *Life (Basel, Switzerland)* **2023**, *13* (8), 1689, doi: 10.3390/life13081689.
13. Moore, R. G.; McMeekin, D. S.; Brown, A. K.; DiSilvestro, P.; Miller, M. C.; Allard, W. J.; Gajewski, W.; Kurman, R.; Bast, R. C., Jr.; Skates, S. J., A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecologic oncology* **2009**, *112* (1), 40-6.
14. Zhang, L.; Chen, Y.; Wang, K., Comparison of CA125, HE4, and ROMA index for ovarian cancer diagnosis. *Current problems in cancer* **2019**, *43* (2), 135-144.
15. Yang, J.; Sa, M.; Huang, M.; Yang, J.; Xiang, Z.; Liu, B.; Tang, A., The reference intervals for HE4, CA125 and ROMA in healthy female with electrochemiluminescence immunoassay. *Clinical biochemistry* **2013**, *46* (16-17), 1705-8.
16. Bozkurt Yavuz, H.; Bildirici, M. A.; Yaman, H.; Karahan, S. C.; Aliyazıcıoğlu, Y.; Örem, A., Reference change value and measurement uncertainty in the evaluation of tumor markers. *Scandinavian journal of clinical and laboratory investigation* **2021**, *81* (7), 601-605.
17. Tan, R.; Yilmaz, M.; Kurtulmuş, Y., A practical approach example to measurement uncertainty: Evaluation of 26 immunoassay parameters. *Biochem Med (Zagreb)* **2022**, *32*(3), 030705, doi: 10.11613/BM.2022.030705.
18. Gu, H.; Lee, J.; Hong, J.; Lee, W.; Yun, Y.; Chun, S.; Lee, W.; Min. W., Practical Considerations for Clinical Laboratories in Top-down Approach for Assessing the Measurement Uncertainty of Clinical Chemistry Analytes. *Ann Lab Med* **2022**, *42*(6), 630-7.
19. Woo, S. Y.; Kim, S., Determination of cutoff values for biomarkers in clinical studies. *Precis Future Med* **2020**, *4* (1), 2-8.
20. Stollberg, S. M.; Näpflin, M.; Nagler, M.; Huber, C. A., Are Tumor Marker Tests Applied Appropriately in Clinical Practice? A Healthcare Claims Data Analysis. *Diagnostics (Basel, Switzerland)* **2023**, *13* (21), 3379, doi:10.3390/diagnostics13213379.
21. Al-Mughales, J. A.; Alahwal, M. S., Inappropriate practice in tumor marker requests at a university hospital in Western Saudi Arabia: A 3-year retrospective study. *The International journal of biological markers* **2020**, *35* (4), 35-43.
22. Lee, J. H.; Choi, J.; Youn, J. S.; Cha, Y. J.; Song, W.; Park, A. J., Comparison between bottom-up and top-down approaches in the estimation of measurement uncertainty. *Clin Chem Lab Med* **2015**, *53*(7), 1025-32.
23. Braga, F.; Panteghini, M., The utility of measurement uncertainty in medical laboratories. *Clin Chem Lab Med* **2020**, *58*(9), 1407-13.
24. Jeong, T. D.; Cho, E. J.; Ko, D. H.; Lee, W.; Chun, S.; Kwon, H. J.; Hong, K. S.; Kim, Y. M.; Min, W. K., A new strategy for calculating the risk of ovarian malignancy algorithm (ROMA). *Clinical chemistry and laboratory medicine* **2017**, *55* (8), 1209-1214.
25. Timbrell, N. E., The Role and Limitations of the Reference Interval Within Clinical Chemistry and Its Reliability for Disease Detection. *British journal of biomedical science* **2024**, *81*, 12339, doi: 10.3389/bjbs.2024.12339.

Author

Youn Joo Lee is a sophomore at Chadwick International School. One of her main interests is data science, a powerful tool that creates solutions for various problems. She hopes to utilize modern technology to tackle complex and sensitive problems and health issues like this study's topic.

DOI: 10.36838/v8i3.56

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Shadow Removal Based on Deep Learning

Heyuan Fang

Holy Trinity School, 11300 Bayview Avenue, Richmond Hill, ON, L4S 1L4, Canada; henryfang5908@gmail.com

ABSTRACT: This study evaluates ST-CGAN and RASM for shadow removal using the ISTD and ISTD+ datasets, comparing their accuracy, efficiency, and generalization. Results show ST-CGAN improves with training, reducing RMSE from 16.39 at epoch 1 to 9.64 at epoch 500, but gains plateau after 100 epochs. Training on ISTD+ lowers RMSE further, yet RASM significantly outperforms ST-CGAN, achieving an RMSE of 2.53 on ISTD+ compared to ST-CGAN's projected 5.02 at epoch 10,000. In addition to these two models, recent transformer-based methods such as ShadowFormer and HomoFormer have demonstrated state-of-the-art results on ISTD+ and SRD benchmarks. These findings highlight that RASM, which leverages a regional attention mechanism within a transformer framework, achieves superior accuracy and computational efficiency compared to earlier CNN-based approaches and other state-of-the-art transformer models, establishing it as a practical solution in the field of shadow removal.

KEYWORDS: Robotics and Intelligent Machines, Machine Learning, Computer Vision, Shadow Removal, ST-CGAN, ShadowFormer, HomoFormer, RASM.

Introduction

"Where there is light, there are shadows.":

Shadows occur when light is obstructed by objects–a natural and unavoidable phenomenon in human life that only rarely obstructs vision with any memorable significance. In robotics applications, however, shadows can cause frequent recognition errors in tasks such as object detection and tracking. These recurring failures, observed for example in competitive robotics systems, highlight the practical importance of effective shadow removal as a preprocessing step. Because robotic platforms typically operate under constrained computational resources, this motivates the exploration of lightweight shadow removal strategies that balance accuracy with efficiency.[1]

In the realm of computer vision, shadows present significant challenges for object detection, tracking, and segmentation. In finer image modification work, shadows impact light resolution—shadowed regions typically have lower luminance values, leading to altered color intensities and reduced accuracy in shadowed areas. Consequently, effective shadow removal is an essential prerequisite for nearly all computer vision applications and has been a subject of scholarly attention for decades.[2]

Artificial Intelligence models are poised to perform the majority of shadow removal in Image processing. While it is well known that AI models improve over time by incorporating new data and enhancing their predictive accuracy, the effectiveness of AI at shadow removal is constrained by several technical bottlenecks, most notably computational limitations within real-world constraints.[3] e.g., Autonomous vehicles will require exceptionally efficient image processing models to make sense of fast-moving objects in heterogeneous scenarios quickly. As of now, Deep Learning has achieved outstanding success in audio and speech processing, natural language processing (NLP), and numerical data analysis.[4] In recent years, deep learning-based shadow removal methods have demonstrated superior performance, primarily due to the availability of extensive training data.[5] With the rise in GPU capabilities, deep neural networks have become a central focus of modern shadow removal research. These models offer higher accuracy and efficiency compared to physical-model-based approaches. However, they also introduce new challenges, primarily the reliance on large and diverse datasets. Recent surveys[2,6] provide a comprehensive overview of shadow removal research from 2017 to 2023, illustrating a clear progression from early CNN-based methods toward transformer-based architectures[4,7] and diffusion-based approaches.[7,8] While these models achieve state-of-the-art results on widely used benchmarks such as ISTD+ and SRD, they generally require more computational resources than CNN or lightweight designs, making them less suitable for deployment in resource-constrained environments. This underscores the importance of evaluating new lightweight frameworks such as RASM not only against earlier CNN approaches but also within the broader trajectory of shadow removal research.

Although transformer- and diffusion-based architectures have recently advanced the field, embedded systems, mobile devices, and autonomous platforms still require methods that balance accuracy with efficiency. Within this context, two representative approaches illustrate distinct strategies. The first is the Stacked Conditional Generative Adversarial Network (ST-CGAN), an early CNN-based model that stacks two CGANs—one for shadow detection and one for shadow removal—thereby providing an end-to-end pipeline and demonstrating the benefits of multitask design under constrained computational budgets. By contrast, the Regional Attention Shadow Removal Model (RASM) reflects a newer, lightweight, region-aware paradigm: by enabling adaptive interaction between shadowed and non-shadowed areas, RASM leverages contextual correlation to restore shadowed content with improved accuracy while maintaining efficiency.

DOI: 10.36838/v8i3.63

While both models aim to improve shadow removal under limited resources, they embody different design philosophies—ST-CGAN as a canonical CNN-based baseline and RASM as a lightweight region-aware framework. This study compares these two approaches to analyze their respective strengths, limitations, and trade-offs in low-compute scenarios. The analysis begins with a brief review of shadow removal challenges, proceeds with a methodological comparison and experimental evaluation, and concludes with observations on practical implications.

Shadow Removal: A Scholarly Review:

Prior work has extensively examined shadow removal as a means to improve downstream computer vision tasks.[3,9] Traditional methods often relied on physical illumination and reflection models,[3] treating each pixel individually according to its lighting conditions. While these approaches can effectively restore images, they tend to be time-consuming and often require manual user interactions, limiting their practicality for large-scale or real-time applications. For example, breakthroughs in color detection for light restoration require a level of image comprehension that would be too costly to program into models running on home devices.[10] Patch-based shadow removal strategies[11] align more closely with deep learning models but would still be too slow for time-critical applications such as autonomous vehicles. In addition, filtering-based strategies like bilateral decomposition[12] were employed to separate base and detail layers for shadow removal. Arbel and Hel-Or[13] argued that shadow removal systems are inherently skewed towards inefficiency because "Shadows in images are typically affected by several phenomena in the scene, including physical phenomena such as lighting conditions, type and behavior of shadowed surfaces, occluding objects, etc.," suggesting profound detail orientation is necessary for complete artifact removal.

With the rise of deep learning, researchers began to address these inefficiencies. AI-assisted shadow removal has progressed beyond texture recognition and 3D modeling.[14,15] For instance, ST-CGAN[9] provided an end-to-end solution that simultaneously handled shadow detection and removal by stacking two conditional GANs, enhancing performance through mutual reinforcement of both tasks. Other studies emphasized context modeling: recent work highlighted concerns about artifact distortion at the shadow border, and proposed programs to model the correlation between shadowed and non-shadowed regions.[4,5] To further improve efficiency, several works[16] focused on increasing the number of training iterations and reducing computational demand. Wang, Li, and Yang[9] advanced this line of research by introducing a bijective mapping network, coupling the procedures of shadow removal and shadow generation in a unified parameter-shared framework. This approach effectively recovered the underlying background contents during the forward shadow removal process. However, their method still required additional programming layers to manage color-rich images, suggesting that dataset limitations remained unresolved.

Two primary strategies have been developed to overcome dataset limitations: Dataset Enhancement and Shadow Simulation Models. Dataset limitations. Dataset Enhancement involves creating shadow masks (binary images indicating shadowed and shadow-free areas) and shadow-free patches (manually edited versions of original images with shadows removed) to expand the dataset. While this method improves model performance, it requires significant human effort, particularly during the shadow removal and masking processes. Shadow Simulation Models artificially generate shadows on existing images to augment datasets. Though effective in increasing data volume, the quality of simulated shadows heavily depends on the diversity of the original dataset. A lack of variation in shadow patterns limits the effectiveness of this approach.

More recently, surveys[2, 17] have shown that shadow removal research from 2017 to 2023 has progressed from early CNN-based methods toward transformer-based architectures[4,6] and diffusion-based approaches[16]. These newer models achieve state-of-the-art performance on benchmarks such as ISTD+ and SRD, and their global context modeling and generative priors provide clear advantages in handling complex shadow patterns and boundary artifacts. While such models demonstrate clear advantages, this study restricts its scope—due to experimental constraints and the focus on robotics environments under limited computational resources—to two representative efficiency-oriented methods: ST-CGAN and RASM.

■ Methodology

This study evaluates the effectiveness of two shadow removal models: the Stacked-Conditional Generative Adversarial Network (ST-CGAN) and the Regional Attention Shadow Removal Model (RASM). The shadow removal process is divided into two main stages: (1) identifying shadowed regions and (2) reconstructing and refining these regions using different computational approaches. Each method is assessed independently to compare its effectiveness in deshadowing.

ST-CGAN is implemented using two separate Conditional Generative Adversarial Networks (CGANs) for the two stages of shadow removal. Python was chosen as the primary programming language due to its extensive machine learning libraries and the prevalence of prior implementations in Python.

As one of the transformers, the shadow removal model, using the Retinex-based model, ShadowFormer, introduced multiple channel-spatial attention mechanisms. Using the Shadow-Interaction Module along with the Shadow-Interaction Attention, ShadowFormer could build correlations between the shadowed and non-shadowed regions and use information from the shadow-free portion for the restoration of the image. ShadowFormer used the idea of a window, in which the transformer would only apply to a specific area (a channel), to compensate for the large calculation complexity and cost.[4]

As another development of implementing transformer use in the task of shadow removal, HomoFormer addressed the issue of the non-uniformity of the shadow in the given image.

Non-uniformity imposes a constraint on weight-sharing models, where they struggle to seek a compromise among regions of various degenerated degrees. HomoFormer implements a random shuffle mechanism in its encoding process to homogenize the degree of shadow degradation, while a de-shuffle in the decoder layer restores the image.[17]

RASM is a lightweight shadow removal model that leverages non-shadow areas to assist in restoring shadowed regions. By implementing a regional attention module, RASM has a regional attention module to continuously learning the correlation between adjacent shadow and non-shadow regions. Differentiating from ShadowFormer and HomoFormer, RASM took a step back and aggregated information from its adjacent non-shadowed regions in its regional attention module. This approach strikes a balance between model complexity and performance, optimizing both accuracy and computational efficiency.

To evaluate the performance of these models, we used the ISTD+ dataset throughout the training and testing phases. Model effectiveness was measured using Root-Mean-Square Deviation (RMSE) and PSNR, MAE, and SSIM. Additionally, we varied the number of training epochs to analyze the relationship between task complexity and image reconstruction quality.

ISTD Dataset:

The ISTD dataset is a widely used benchmark for shadow removal tasks, comprising 1,870 image triplets across 135 diverse scenes that feature various shadow shapes and lighting conditions. The dataset is divided into 1,330 triplets for training and 540 triplets for testing, facilitating model evaluation and generalization.

Each triplet in the dataset comprises the following components:

1. A shadowed image – the original input image containing natural shadows.
2. A shadow mask – a binary mask delineating the shadowed regions.
3. A shadow-free image – an image where shadows have been manually removed following the shadow mask.

The shadow removal in ISTD is performed manually, ensuring precise adherence to the shadow mask. This high-quality annotation enables Convolutional Neural Networks (CNNs) and Conditional Generative Adversarial Networks (CGANs) to effectively learn shadow localization from the shadow mask, while also identifying differences in luminance, texture, and resolution between shadowed and shadow-free images. By leveraging these structured triplets, the models can be trained to accurately detect and reconstruct shadowed areas, improving overall deshadowing performance.

Figure 1: Example of a triplet: Shadow-Free Image, Shadow Mask, and Original Image.[9]

Assessing the Accuracy of Models:

We use the root mean square error (RMSE) as the metric to quantify the discrepancy between the ground-truth shadow-free image and the recovered image, with lower values indicating higher accuracy or lower distortion.

The RMSE is determined by the equation:[18]

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2},$$

Figure 2: Root Mean Square Error (RMSE) Equation.[18]

We consider two input sets of images, I1 and I2, where each image consists of n pixels. For each corresponding pixel in the two images, we compute the difference between their RGB values, take the average of these differences, and then compute the square root of this average to determine the overall error. This results in the Root Mean Squared Error (RMSE), which quantifies the difference between the two images. The detailed RMSE calculation is provided in the Appendix.

Methodologies for Shadow Removal: ST-CGAN, ShadowFormer, HomoFormer, and RASM:

Shadow Removal Using ST-CGAN:

Shadow removal from a single image involves two fundamental tasks: shadow detection and shadow removal. The Stacked Conditional Generative Adversarial Network (ST-CGAN) is an architecture designed to perform both tasks jointly, enabling end-to-end learning for improved accuracy and efficiency.

The ST-CGAN architecture consists of two stacked Conditional Generative Adversarial Networks (CGANs):

1. Shadow Detection Network – Identifies shadowed regions and generates a shadow mask.
2. Shadow Removal Network – Uses the shadow mask along with the original image to reconstruct a shadow-free version.

Each CGAN comprises a generator and a discriminator, working in tandem to enhance the realism and accuracy of the shadow removal process. The generator aims to produce an image where shadows are effectively removed, while the discriminator evaluates the authenticity of the generated output, pushing the model toward higher-quality reconstructions.

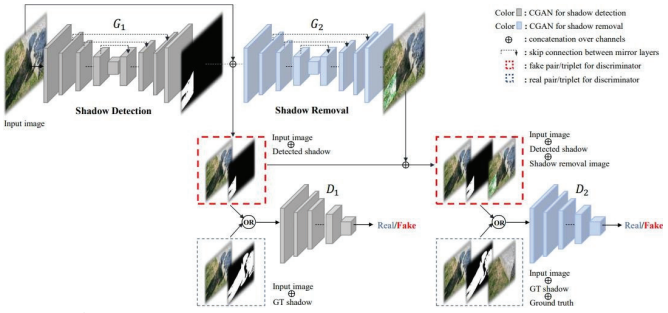


Figure 3: Architecture of the Proposed ST-CGAN: Stacked CGANs for Shadow Detection and Removal.[9]

Shadow Removal Using Shadowformer:

Shadowformer introduces a multiple-scale channel-spatial attention mechanism within a Transformer framework. This

DOI: 10.36838/v8i3.63

design addressed the challenge of other transformer models of integrating global semantics from deep feature layers with local details from shallow feature layers.

ShadowFormer started with the Retinex-based shadow model that handles shadow degradation to allow models to draw information from non-shadow regions to restore the shadow region. The restoration process is done through an encoder-decoder under a single-stage transformer framework via channel attention to efficiently multiple-stack the hierarchical information. Then, the correlation between the shadow vs non-shadow region is exploited by the Shadow-Interaction Module with Shadow-Interaction Attention. This allowed ShadowFormer to address the challenge of colour inconsistency and boundary trace in the restored shadow-free images.

Here, we present a quick overview of the architecture of ShadowFormer. The process begins with 2 inputs, the shadow image along with the shadow mask, which is linearly projected, which maps the inputs into a latent feature space. In the encoder stage, features are processed using a channel attention model and a standard transformer structure. Then, the features enter the decoder stage, which mirrors the encoder by using channel attention modules to reconstruct the spatial details. A linear projection layer transformation is connected to feature it back into image space.

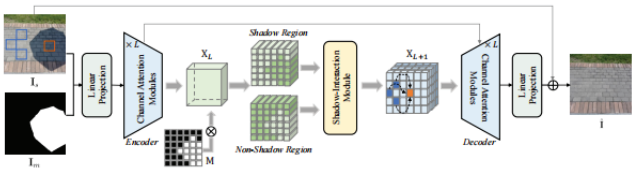


Figure 4: Architecture of the Proposed ShadowFormer.[4]

Shadow Removal Using HomoFormer:

The HomoFormer is an advanced shadow removal model that leverages a homography-inspired Transformer architecture to model the relationship between shadow and non-shadow regions. A foundation based on other transformer architectures that would focus on the non-uniform issue of shadow and as well as the classic self-attention method. The image shuffle strategy involves upsampling and downsampling to exchange information between channels and space, while preventing the spatial arrangements of pixels. At the end, with an implementation of self-attention to concentrate on regional attention, HomoFormer decreased computing complexity compared to other models.

This model compromises a conflict of previous transformer models: previous models suffer from either a quadratic increase in complexity as the resolution of the image increases, or the weight sharing when dealing with non-uniform shadow degradation. HomoFormer implemented the strategy to homogenize the non-uniform distribution.

Here, we present a brief overview of the HomoFormer structure. The process begins with the input project stage, where the images with shadow and the shadow masks are mapped into a feature space, which is the core parts: the HomoBlocks integrates Layer Normalization, local self-attention with random shuffle, and SMLP (standard transformer used). The regional attention is also implemented in this section to minimize complexity in calculation. During the down and then up-sampling layers, progressive reduction in the spatial resolution while increasing feature richness is applied with downsampling, while the upsampling strategy is used to restore the spatial resolution. This process is able to preserve important details lost during the traditional encoder-decoder architecture. The features would lastly pass through an output projection layer to generate the image.

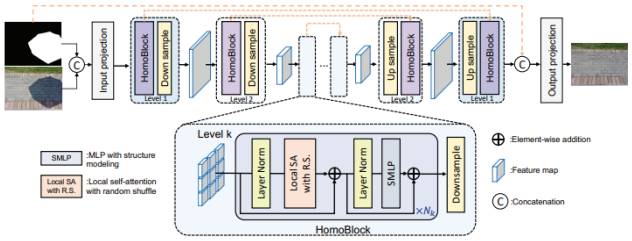


Figure 5: Architecture of the Proposed HomoFormer.[17]

Shadow Removal Using RASM:

The Regional Attention Shadow Removal Model (RASM) is a lightweight yet effective shadow removal approach that utilizes non-shadow regions to enhance the reconstruction of shadowed areas. By incorporating regional attention mechanisms, RASM enables context-aware interactions between shadowed and non-shadowed areas, facilitating a more accurate and natural restoration process.

This model is designed to strike an optimal balance between computational efficiency and accuracy, ensuring effective shadow removal while maintaining manageable model complexity. Specifically, the regional attention module could focus on a specific area of the matrix and therefore avoid being excessive yet non-informative. In doing so, RASM achieved only 1/4 of GFLOPs. Through its regionally contextual approach, RASM enhances the overall quality of shadow-free images while reducing computational overhead.

Here, we present a brief overview of the RASM structure. On the left side (a) was the process of model and shadow contraction area, which employs the Channel Attention Module (b). After allowing the global information to interact, Channel Attention enters and eventually concludes in a spatial information interaction.

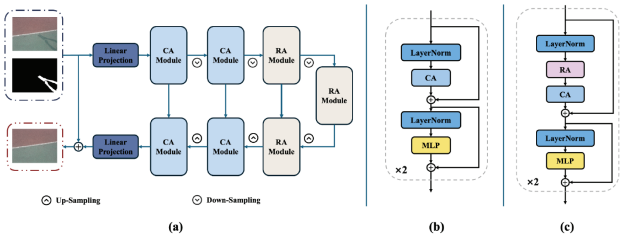


Figure 6: Demonstration of RASM.[19]

Experiment

For computational resources, we employed Google Colab as the primary cloud-based virtual machine. The training duration per epoch ranged from 3,087 to 3,850 seconds on a CPU and approximately 90.79 to 95.69 seconds on a GPU. However, GPU availability on Colab was limited, restricting the

number of training epochs to well below 500. To supplement this, we conducted additional tests on a consumer-grade laptop equipped with an NVIDIA GeForce GTX 1650 GPU with 16GB of memory. The GPU performance on this system was comparable to that of Colab, making it a reasonable benchmark for evaluating model feasibility on consumer hardware.

To systematically evaluate the performance of the shadow removal models, we conducted a series of experiments under varying computational environments and training durations. The experiments were performed using both Google Colab and a personal computing device, with training durations of 1, 100, and 500 epochs. The effectiveness of the models was assessed using both qualitative visual analysis and quantitative evaluation via Root-Mean-Square Deviation (RMSE).

To investigate the potential influence of computational environments on model performance, we trained the model for 100 epochs using both Google Colab and a personal laptop. Figure 4 and Figure 5 illustrate six representative examples from these experiments, where each row corresponds to a different test sample and each column represents the original shadowed image, the ground truth shadow-free image, the shadow-free image generated by the model, the ground truth shadow mask, and the shadow mask produced by the model. A qualitative analysis of the results revealed no significant differences between the outputs generated on the two platforms. Samples that were successfully processed on one platform were also successfully processed on the other, while failure cases remained consistent across both environments. This observation was further confirmed by the RMSE values, which indicated that the personal laptop achieved performance comparable to the Colab-based model. These findings suggest that the choice of computational environment does not substantially impact the model's performance, validating the robustness of the implementation across different hardware configurations.

Results and Discussion

Experiment Results:

To assess the impact of training duration on shadow detection and removal, we compared the results obtained after 1,100 and 500 epochs, all conducted on a personal laptop (Figures 6 and 7). After a single epoch of training, the model exhibited poor generalization, failing to accurately detect shadow boundaries or reconstruct shadow-free images. The generated shadow masks were imprecise, and the overall visual quality of the outputs was suboptimal. This was reflected in a high RMSE of 16.39, indicating a significant deviation from the ground truth. These results highlight the necessity of extended training for the model to learn meaningful representations of shadow regions and their corresponding shadow-free reconstructions.

After 100 epochs, the model demonstrated substantial improvements in both shadow detection and removal. The generated shadow-free images closely resembled the ground truth, with only minor imperfections in certain cases. This was quantitatively supported by a significant reduction in RMSE to 10.09, indicating enhanced accuracy and improved reconstruction quality. However, increasing the training duration further to 500 epochs yielded only marginal improvements. While the RMSE decreased slightly from 10.09 to 9.64, the visual differences were negligible. This suggests that beyond a certain point, additional training yields diminishing returns, as the model reaches a plateau in performance where further refinement offers minimal perceptible enhancement in shadow removal quality.

The experimental results lead to three key observations:

- The computational environment does not significantly affect model performance, as both Colab and a personal laptop produced comparable outputs and RMSE values.
- Training duration has a substantial impact on model effectiveness, particularly in the early stages, as demonstrated by the significant improvements between 1 and 100 epochs.
- Beyond 100 epochs, further training yields only incremental gains, with minimal reductions in RMSE and imperceptible improvements in visual quality.

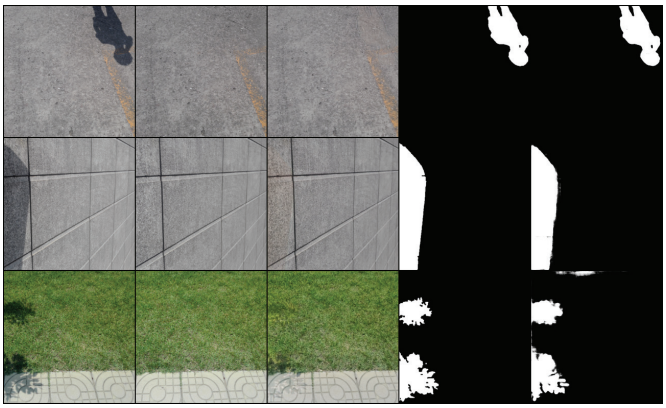


Figure 7: Completed with Colab, 100 epochs. Here, we demonstrate 3 examples (corresponding to 3 rows) of the output of our program. Each column, from left to right, corresponds to: the shadowed picture from the dataset, the shadow-free image from the dataset, the shadow-free image the code generated, the shadow mask the dataset contains, and the shadow mask the code generated.

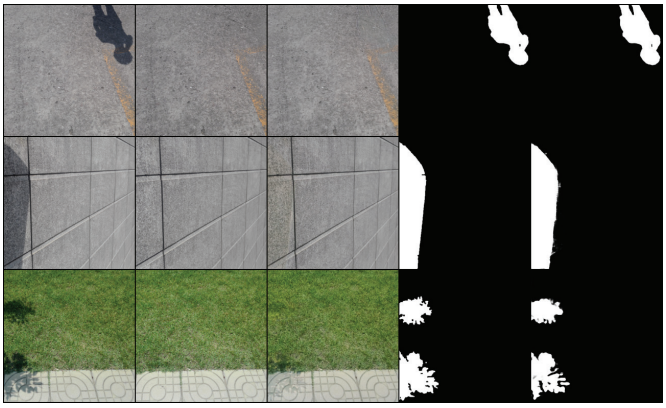


Figure 8: Completed on personal laptop, 100 epochs. As a result, between a personal laptop (Figure 8) vs Colab (Figure 7), we could eyeball and conclude that there is almost no difference. Therefore, we conclude that the computational environment does not significantly affect model performance.

DOI: 10.36838/v8i3.63

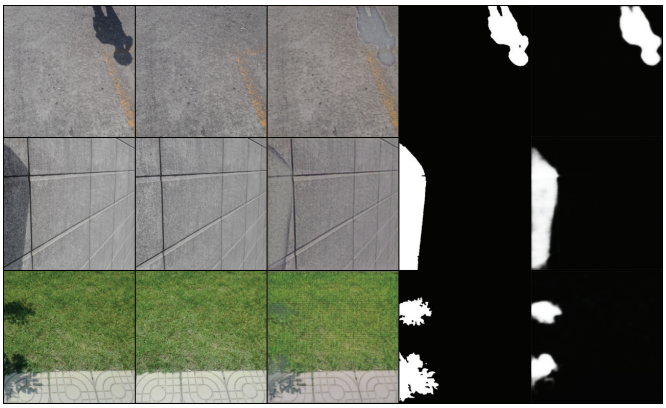


Figure 9: Completed on personal laptop, 1 epoch. Compared with the personal laptop after 100 epochs (Figure 8), the output after 100 epochs on a personal laptop (Figure 8) shows a significantly higher quality.

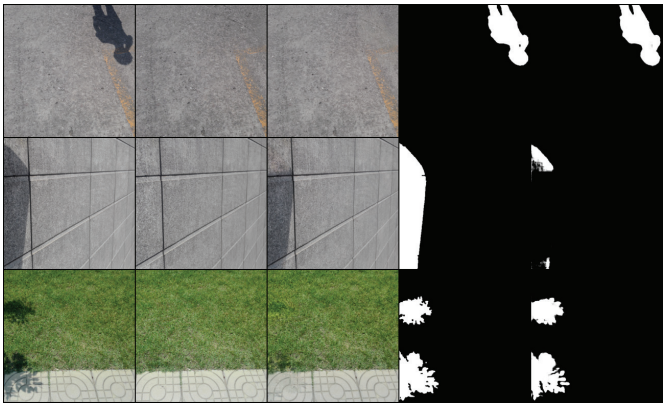


Figure 10: Completed on personal laptop, 500 epochs. At this point, it is difficult to eyeball any difference between 100 epochs (Figure 8) and 500 epochs (Figure 10), which means the improvements are gradually decreasing during this process; however, there could still be a non-obvious improvement.

Effect of Training Epochs on RMSE with ST-CGAN:

Table 1: RMSE was tested with different training epochs using ST-CGAN.

Dataset	epoch=1	epoch=10	epoch=100	epoch=300	epoch=500	epoch=10000
ISTD	16.39	12.83	10.09	9.88	9.64	7.47

Table 1 illustrates the Root Mean Squared Error (RMSE) values obtained at different training epochs for the ISTD dataset. RMSE serves as a quantitative measure of deshadowing effectiveness, where lower values indicate better performance. The results demonstrate a clear downward trend in RMSE as the number of training epochs increases, suggesting that extended training enhances the model's ability to remove shadows.

At epoch 1, the RMSE is 16.39, indicating a relatively high error in shadow removal. As training progresses to epoch 10, the RMSE decreases to 12.83, reflecting an improvement in the model's capacity to reconstruct shadow-free images. A more significant reduction is observed at epoch 100, where the RMSE drops to 10.09, indicating substantial progress in learning. However, after epoch 100, the reduction in RMSE becomes less pronounced. At epoch 300, the RMSE is 9.88, and at epoch 500, it further declines slightly to 9.64. This diminishing improvement suggests that the model approaches its performance plateau, where additional training yields only marginal enhancements.

For comparison, the ST-CGAN paper[9] reports an RMSE of 7.47 at epoch 10,000, which is significantly lower than the RMSE values achieved in our experiments. This discrepancy suggests that extended training beyond 500 epochs may further enhance the model's performance.

Effect of Training Epochs on RMSE between Models:

Table 2: RMSE compared with different training epochs between ST-CGAN and RMSE.

Dataset	Model	epoch=1	epoch=10	epoch=100	epoch=300	epoch=500	epoch=10000
ISTD	ST-CGAN	16.39	12.83	10.09	9.88	9.64	7.47
ISTD+	ST-CGAN	14.06	9.14	6.97	6.67	6.48	5.02(estimated)
ISTD+ RMSE / ISTD RMSE		0.858	0.712	0.691	0.675	0.672	0.672

To ensure a fair comparison between the ST-CGAN and RASM models, we used a consistent dataset for evaluation. The ST-CGAN model was originally trained and tested on the ISTD dataset. In contrast, the RASM model was evaluated using the ISTD+ dataset, a refined version of ISTD that addresses illumination inconsistencies between shadowed and shadow-free images. To facilitate direct comparison, we recalculated the RMSE values for the ST-CGAN model using the ISTD+ dataset at training epochs 1, 10, 100, 300, and 500. The recalculated RMSE values for ST-CGAN are derived from our experimental results, while the RMSE values for RASM are taken from "Regional Attention for Shadow Removal."[19]

According to Table 2, the comparison between the ST-CGAN model trained on ISTD vs. ISTD+ reveals notable differences in shadow removal effectiveness. At the initial training stage (epoch 1), the RMSE for ISTD+ is 14.06, slightly outperforming ISTD's 16.39, yielding an RMSE ratio of 0.858. A more pronounced improvement is observed at epoch 10, where the RMSE for ISTD+ decreases to 9.14, compared to 12.83 for ISTD, resulting in an RMSE ratio of 0.712. However, beyond epoch 10, the reduction in RMSE for ISTD+ becomes less significant in comparison to ISTD. By epoch 100, the ratio stabilizes at 0.691, then further decreases to 0.675 at epoch 300, and 0.672 at epoch 500. This trend suggests that the RMSE ratio is approaching a steady state. Assuming this ratio remains consistent up to epoch 10,000, we estimate the RMSE for ISTD+ at epoch 10,000 to be approximately 5.02.

According to the RASM paper,[19] the RMSE of the RASM model trained on ISTD+ is reported as 2.53, which is significantly lower than the projected RMSE of ST-CGAN from our experiments. This substantial difference indicates that RASM achieves superior performance in shadow removal compared to ST-CGAN, particularly in terms of quantitative accuracy. The results suggest that the regional attention mechanism employed in RASM is more effective in leveraging contextual information for shadow reconstruction, leading to a more precise removal process.

Comparison with Advanced Transformer Models:

Table 3: The performance of the 3 models mentioned, with measurements of PSNR, SSIM, and RMSE, using the ISTD+ dataset.[6]

Method	Shadow Region			Non-Shadow Region			All-Region		
	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
ShadowFormer	39.48	0.992	5.23	38.82	0.983	2.30	35.46	0.971	2.78
HomoFormer	39.49	0.993	4.73	38.75	0.984	2.23	35.35	0.975	2.64
RASM	40.73	0.993	4.41	39.23	0.985	2.17	36.16	0.976	2.53

Table 4: The performance of the 3 models mentioned, with measurements of PSNR, SSIM, and RMSE, using the SRD dataset.[6]

Method	Shadow Region			Non-Shadow Region			All-Region		
	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
ShadowFormer	35.55	0.982	6.14	36.82	0.983	3.54	32.46	0.957	4.28
HomoFormer	38.81	0.987	4.25	39.45	0.988	2.85	35.37	0.972	3.33
RASM	37.91	0.988	5.02	38.70	0.992	2.72	34.46	0.976	3.37

Table 5: The comparison of 4 models mentioned about the number of parameters and FLOPs.[4,20]

Method	#Params. (M)	FLOPs (G)
ST-CGAN	29.2	17.9
ShadowFormer	11.4	63.1
HomoFormer	17.8	38.4
RASM	5.2	25.2

Beyond CNN-based methods and lightweight designs such as RASM, recent studies have introduced transformer and diffusion architectures that achieve state-of-the-art results on benchmark datasets. For example, ShadowFormer reports an RMSE of 2.78, PSNR of 35.46 dB, and SSIM of 0.971 on the ISTD+ dataset.[4] Similarly, HomoFormer achieves an RMSE of 2.64, PSNR of 35.35 dB, and SSIM of 0.975,[7] reflecting its ability to model non-uniform degradation through homogenized attention blocks.

These figures illustrate that transformer-based models currently surpass CNN approaches in absolute accuracy, particularly in handling complex shadow boundaries and color inconsistencies. However, they are also characterized by significantly higher computational demands, often requiring multiple GPUs or extended training times.

RASM has a small number of parameters and low FLOPs, utilizing a negligible amount of computational resources while achieving superior performance, demonstrating that RASM effectively balances model complexity and model performance.

In this study, the experimental comparison is limited to ST-CGAN and RASM due to hardware constraints typical of robotics environments, where platforms must balance accuracy against strict efficiency requirements. Nevertheless, as robotic systems increasingly adopt more powerful GPUs, the incorporation of advanced architectures may become both feasible and advantageous, enabling higher-fidelity perception in dynamic real-world tasks.

■ Conclusion

This study compared ST-CGAN and RASM for shadow removal, evaluating their effectiveness, computational efficiency, and accuracy using the ISTD and ISTD+ datasets. Our experiments demonstrated that ST-CGAN benefits from extended training, with RMSE decreasing from 16.39 at epoch 1 to 9.64 at epoch 500 on ISTD. However, beyond 100 epochs, improvements became marginal, indicating a performance plateau. Additionally, results confirmed that the choice of computational environment (Google Colab vs. personal laptop) had no significant impact on model performance.

Training ST-CGAN on ISTD+ consistently resulted in lower RMSE values, highlighting the role of dataset refinement in improving shadow removal accuracy. However, RASM significantly outperformed ST-CGAN, achieving an RMSE of 2.53 on ISTD+, compared to ST-CGAN's projected 5.02 at epoch 10,000. This suggests that RASM's regional attention mechanism more effectively restores shadow-free images while maintaining computational efficiency.

Beyond these two models, transformer-based methods such as ShadowFormer, HomoFormer have recently set new benchmarks, with RMSE values under 2.8 and SSIM above 0.97. These results indicate that advanced architectures offer superior absolute accuracy, particularly for complex shadow boundaries. However, they remain computationally demanding, making them less feasible for current low-compute robotic platforms.

Taken together, our findings suggest that lightweight region-aware frameworks like RASM currently provide the best trade-off between efficiency and accuracy in resource-constrained settings. Future research should investigate strategies related to RASM approaches, emphasizing the use of regional channels and a continuously modifying window to minimize calculation complexity and thereby lower hardware requirements for its further use in applications under the current hardware level of robots. As robotics and embedded systems increasingly gain access to high-performance GPUs, it will become both feasible and advantageous to deploy these advanced models, enabling higher-fidelity visual perception in dynamic real-world environments.

■ Acknowledgments

I would like to sincerely thank all those who contributed to this paper, including all guidance teachers/professors for their invaluable support and for providing this exceptional environment. I would like to specifically thank Professor Yashtini from Georgetown University for her guidance, patience, and encouragement during the summer of 2024.

■ References

1. Alzayat Saleh, Alex Olsen, Jake Wood, Bronson Philippa, & Mostafa Rahimi Azghadi. (2025). FieldNet: Efficient real-time shadow removal for enhanced vision in field robotics. Expert Systems with Applications, 127442. https://doi.org/10.1016/j.eswa.2025.127442
2. Zhu, X., Chow, C.-O., & Chuah, J. H. (2024). From darkness to clarity: A comprehensive review of contemporary image shadow removal research (2017–2023). Image and Vision Computing, 148, 105100. https://doi.org/10.1016/j.imavis.2024.105100
3. Rubinger, L., Gazendam, A., Ekhtiari, S., & Bhandari, M. (2023). Machine learning and artificial intelligence in research and healthcare. Injury, 54(Suppl 3), S69–S73. https://doi.org/10.1016/j.injury.2022.01.046

DOI: 10.36838/v8i3.63

4. Guo, L., Huang, S., Liu, D., Cheng, H., & Wen, B. (2023, June). Shadowformer: Global context helps shadow removal. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 1, pp. 710–718). https://doi.org/10.1609/aaai.v37i1.25148
5. Liang, J., Wang, J., Xu, W., Hou, L., & Zhou, B. (2025). Towards hard and soft shadow removal via dual-branch separation network and vision transformer. arXiv. https://doi.org/10.48550/arXiv.2501.01864
6. Guo, L., Wang, C., Wang, Y., Yu, Y., Huang, S., Yang, W., Kot, A. C., & Wen, B. (2024). Single-Image Shadow Removal Using Deep Learning: A Comprehensive Survey (arXiv:2407.08865). arXiv. https://doi.org/10.48550/arXiv.2407.08865
7. Mei, K., Figueroa, L., Lin, Z., Ding, Z., Cohen, S., & Patel, V. M. (2023). Latent feature-guided diffusion models for shadow removal. arXiv. https://doi.org/10.48550/arXiv.2312.02156
8. Luo, J., Li, R., Jiang, C., Zhang, X., Han, M., Jiang, T., Fan, H., & Liu, S. (2024). Diff-Shadow: Global-guided diffusion model for shadow removal. arXiv. https://doi.org/10.48550/arXiv.2407.16214
9. Wang, J., Li, X., & Yang, J. (2018). Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1788–1797. https://doi.org/10.1109/CVPR.2018.00192
10. Finlayson, G. D., Hordley, S. D., Lu, C., & Drew, M. S. (2006). On the removal of shadows from images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(1), 59–68. https://doi.org/10.1109/TPAMI.2006.18
11. Zhang, L., Zhang, Q., & Xiao, C. (2015). Shadow remover: Image shadow removal based on illumination recovering optimization. IEEE Transactions on Image Processing, 24(11), 4623–4636. https://doi.org/10.1109/TIP.2015.2465159
12. Yang, Q., Tan, K.-H., & Ahuja, N. (2012). Shadow removal using bilateral filtering. IEEE Transactions on Image Processing, 21(10), 4361–4368. https://doi.org/10.1109/TIP.2012.2208976
13. Arbel, E., & Hel-Or, H. (2011). Shadow removal using intensity surfaces and texture anchor points. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(6), 1202–1216.https://doi.org/10.1109/TPAMI.2010.157
14. Finlayson, G. D., Drew, M. S., & Lu, C. (2009). Entropy minimization for shadow removal. International Journal of Computer Vision, 85(1), 35–57. https://doi.org/10.1007/s11263-009-0243-z
15. Liu, F., & Gleicher, M. (2008). Texture-consistent shadow removal. In D. Forsyth, P. Torr, & A. Zisserman (Eds.), Computer Vision – ECCV 2008 (Vol. 5305). Lecture Notes in Computer Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-88693-8_32
16. Le, H., & Samaras, D. (2020). From shadow segmentation to shadow removal. In A. Vedaldi, H. Bischof, T. Brox, & J. M. Frahm (Eds.), Computer vision – ECCV 2020 (Vol. 12356). Lecture Notes in Computer Science. Springer, Cham. https://doi.org/10.1007/978-3-030-58621-8_16
17. Xiao, J., Fu, X., Zhu, Y., Li, D., Huang, J., Zhu, K., & Zha, Z.-J. (2024). HomoFormer: Homogenized Transformer for Image Shadow Removal. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 25617–25626. https://doi.org/10.1109/CVPR52733.2024.02420
18. Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. Geoscientific Model Development, 15(14), 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022
19. Liu, H., Li, M., & Guo, X. (2024, October). Regional attention for shadow removal. In Proceedings of the 32nd ACM International Conference on Multimedia (pp. 5949–5957). https://doi.org/10.1145/3664647.3681126
20. Xiao, J., Wang, C., Mei, K., Ding, Z., Lin, Z., Pfister, H., & Wen, B. (2024). HomoFormer: Homogenized transformer for image shadow removal. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1-11). IEEE. https://doi.org/10.1109/CVPR52733.2024.02420

Authors

Henry Fang is a grade 11 student at Holy Trinity School in Canada. He is passionate about STEM subjects and hopes would obtain a degree in the field of Engineering. He is also typically interested in the function of AI in the fields of medicine and Engineering and hopes he will participate in related jobs.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Uncovering Subtype-Specific Biomarkers in Breast Cancer through Bioinformatics Analysis

Mihika Deora[1], Nirupma Singh[2]

1) Oberoi International School, Mumbai, Maharashtra, India
2) Department of Biological Sciences and Engineering, Netaji Subhas University of Technology (Formerly NSIT), Dwarka, Delhi, India; nirupmajadaun21@gmail.com

ABSTRACT: Breast cancer is the most common cancer in the world, accounting for over 28.2% of all female cancers. However, there are still no effective subtype-specific biomarkers to help in diagnosis and more targeted therapies for patients with breast cancer. This study uses advanced bioinformatics approaches to identify subtype-specific biomarkers for *four* molecular subtypes and analyse their potential role in treatment processes. To accomplish this objective, differential gene expression analysis (DGE) was conducted using the GEO2R (Gene Expression Omnibus) tool to gain different data sets. Using the data, a network was constructed in the database STRING, which was then analysed using Cytoscape to identify the topological parameters. Pathway analysis was conducted in the Reactome database to determine the top-enriched pathways in which the significant hub genes for breast cancer are present. The study identified the top significant genes and hub genes in breast cancer subtypes, assessing their ability as biomarkers for more personalised treatments through detailed DGE, network, and pathways analysis. Notably, *RPS27A* emerged as the top significant gene in all the subtypes, with its presence in the EML4 and NUDC in the mitotic spindle formation pathway for all 4 subtypes showing its potential for therapy. These findings will enhance understanding of the treatment processes of breast cancer and aim for more targeted therapies for different subtypes.

KEYWORDS: Computational Biology and Bioinformatics, Computational Biomodelling, Cancer Biology Analysis, Network Biology, Pathway Analysis.

Introduction

Breast cancer (BC), the most commonly diagnosed cancer in women worldwide, leads to significant morbidity and mortality, placing a considerable strain on healthcare systems. Breast cancer affects millions of women globally, with approximately 1.5 million new cases annually, making it a leading cause of cancer-related deaths.[1] In India, some foundations such as the ICGA (Indian Cancer Genome Atlas) are developing technologies to identify the genetic basis of cancer in the Indian population and genetic biomarkers that will improve the rate of detection and better-targeted therapies.[2]

Breast cancer can be classified into several subtypes, which are grouped according to the immunohistochemical expression of hormone receptors. Luminal A is characterised by the presence of the ER and PR receptors and the absence of the HER2 (human epidermal growth factor receptor 2) receptor. Clinically, this subtype grows at a slower rate, has a lower chance of relapse, and has a higher survival rate compared to others. It presents a positive and faster response to hormone therapy in comparison to chemotherapy.[3] According to the European Society for Medical Oncology, genetic platforms identify the preferred treatment for the patient based on the severity, risk of relapse, and survival rate.[4] Luminal B grows faster and is harder to predict than Luminal A, but is also characterised by the presence of PR+ and sometimes PR- receptors.[5] Hormonal therapy, along with chemotherapy, can be beneficial to it. The presence of HER2 expression characterises HER2 and causes it to grow at a faster rate compared to the luminal types. The prediction has improved since the introduction of more HER2-targeted therapies, specifically directed drugs, and a high response to chemotherapy. Triple-negative (TNBC) has ER-, PR-, and HER2- receptors, which cause it to have highly aggressive behaviour, early relapses, a higher proliferation rate, changes within the repair genes, and genomic stability. *BRCA1* mutation carriers often have the basal-like subtype, which is comparable to TNBC but has different genetic markers.[6]

By identifying new biomarkers and the genetic basis of the disease, its risk and progression can be monitored and better understood. Studies have used different bioinformatic approaches to focus on the molecular heterogeneity of breast cancer progression. The National Institute of Biomedical Genomics (NIBMG) uses biomedical genomics to identify the genetic markers associated with the disease. Although various biomarkers have been proposed, the severity of breast cancer requires more efficient data methods, such as bioinformatics approaches that can help bring data from diverse sources together and offer a more holistic view of the disease. Research to identify molecular biomarkers that would be more efficient for therapies has been done, which has helped improve the progression of the disease.[7] By using gene expression profiling to uncover intrinsic subtypes, researchers carried out groundbreaking research on the molecular classification of breast cancer, which has since impacted therapeutic approaches.[8] To gain a better understanding of tumor heterogeneity, this work was extended by dividing breast cancer into ten different subgroups using integrative genomic analysis.[9]

DOI: 10.36838/v8i3.71

While there has been significant progress in breast cancer research, it lacks deeper bioinformatic analysis. These studies have relied on genomic data, overlooking the proteomic and transcriptomic factors, which are not able to capture tumor heterogeneity. Subtype-specific biomarker identification is required for deeper analysis. To find new biomarkers and treatment targets, advanced computational techniques are needed due to the complexity of breast cancer subtypes. Bioinformatics tools such as STRING, CYTOSCAPE, and REACTOME will be employed in the research to perform differential gene expression analysis, pathway analysis, and network analysis, allowing researchers to explore the disease-related pathways in the body and genetic mutations that are related to the disease. The use of such bioinformatic tools allows vast datasets to be analysed together quickly and efficiently, providing a deeper understanding of the progression and the possible treatments. Furthermore, research has looked at the mutational signatures of breast cancer; however, further research is required to determine the functional implications of these mutations and the effect they have on the body's response to the treatment.[10] Additionally, there is still research on triple-negative breast cancer, mainly reliable therapeutic targets and personalised treatments for this subtype's aggressiveness.

However, previous research studies have primarily relied on genomic data and ignored the incorporation of proteomic and transcriptomic data, which could provide a more in-depth understanding of the disease. This gap in the research shows that a more comprehensive approach with more in-depth biological research would be beneficial. For example, whereas genetic mutations have been researched in great detail, little is known about how they affect the function and patterns of protein production. Filling in these gaps will improve our knowledge of the illness and result in better methods for diagnosis and treatment. Thus, this paper argues that the leverage of bioinformatic tools that incorporate multi-omics data will help identify new biomarkers, creating better therapeutic targets, personalised treatments, and earlier detection of the disease. This study intends to identify formerly unknown molecular patterns that might be useful treatment targets by examining a variety of datasets. By enabling more accurate disease characterisation and customised treatments, the discoveries will support the expanding field of personalised therapies.

Methods

Data Collection:

The gene expression data were retrieved from the NCBI GEO platform. Specifically, the subtypes, Luminal A, Luminal B, HER2-Positive, and Triple negative RNA-Seq data are extracted for analysis. *GSE233242 -'Tumor circadian clock strength influences metastatic potential and predicts patient prognosis in Luminal A breast cancer*' investigates the circadian clocks in human breast tumors by conducting an expression profiling using high-throughput sequencing. *GSE214344 - 'A genome-wide cell-free DNA methylation analysis identifies an episignature associated with metastatic luminal B breast cancer.*' aims to discover non-invasive biomarkers of the disease using an epigenomic approach. *GSE52194 - 'mRNA-sequencing of breast cancer subtypes and normal tissue*' uses RNA sequencing technology to identify the digital transcriptome. *GSE167152 - 'Comparative Characterisation of 3D Chromatin Organisation in Triple-Negative Breast Cancers [RNA-seq]'* detected CTCF-dependent TNBC-susceptible loss/gain of 3D chromatin organisations using expression profiling by high-throughput sequencing.

Differential Gene Expression:

Differential gene expression analysis was conducted using the GEO2R analysis tool for the RNA-Seq datasets obtained for each subtype. GEO2R applies a default normalisation to each of the datasets (log2 transformation and quantile normalisation for microarray data, variance-stabilising transformation for RNA-seq) before it undergoes differential expression analysis. This allows for fewer technical biases and comparability across samples, which increases the reliability of the identified DEGs. For the RNA-Seq data of subtype Luminal A, 29 tumor samples and 42 normal samples were assigned as test and control groups, respectively. The second type, Luminal B, had 7 cell-free DNA samples from luminal B patients for the test group and 5 normal cell-free DNA samples for the control group. The third subtype, TNBC, had 18 triple-negative breast cancer cell samples and 2 normal samples, which were assigned as the test and control groups, respectively. The last subtype, HER2+, had 5 tumor samples and 3 that matched normal samples. The raw data were normalised using DESeq2 (Differential Expression analysis based on the Negative Binomial distribution), and the DEGs (Differentially Expressed Genes) were selected based on an adjusted p-value < 0.05 and a fold change > 2. To ensure robustness and reduce the possibility of background noise, a threshold of FC > 2 was applied to identify those genes with significant, biologically meaningful expression changes together with an adjusted p-value < 0.05.

Network Construction and Analysis:

A network was constructed using the STRING database by identifying gene-gene interactions for the significant genes of each subtype. A list of the top 2000 differentially expressed genes was used as input to generate a network. The interactions with a high confidence of 0.7 were retained. The network was exported as a short tabular text output to visualise in Cytoscape to analyse the topological parameters and visualise the network. The software provided gene interactions and positive topological parameters, which were downloaded for further analysis.

Pathway Analysis:

From the topological parameters file created in Cytoscape, the degree was set in a descending manner to identify the top genes. Pathway enrichment analysis was performed using Reactome for the significant DEGs (differentially expressed genes) to identify biological pathways significantly affected by the differentially expressed genes. The top enriched pathways with a p-value < 0.05 were considered significant. The pathway analysis results adopted from Reactome were then used to

search for the top 10 hub genes of the network to check their presence in the top enriched pathways.

Results and Discussion

This section represents the results obtained from the bioinformatics analysis carried out for different subtypes of breast cancer and the top genes and pathways identified to gain insights into personalised treatment approaches. The focus is on understanding the specific biological processes, genes, and pathways associated with the subtypes of breast cancer. After performing DGE analysis for Luminal A - GSE233242, 13450 significant genes were identified, for Luminal B - GSE270967, 2088 significant genes were found, for TNBC - GSE167152, 4546 significant genes were identified, and for HER2+ - GSE52194, 4575 significant genes were found. The overexpressed and underexpressed genes were visualised in the form of volcano plots for each subtype, as shown in Figure 1. The analysis revealed a total of 22,572 genes associated with different subtypes of breast cancer. The key findings also included the identification of different genes and pathways that play a role in breast cancer.

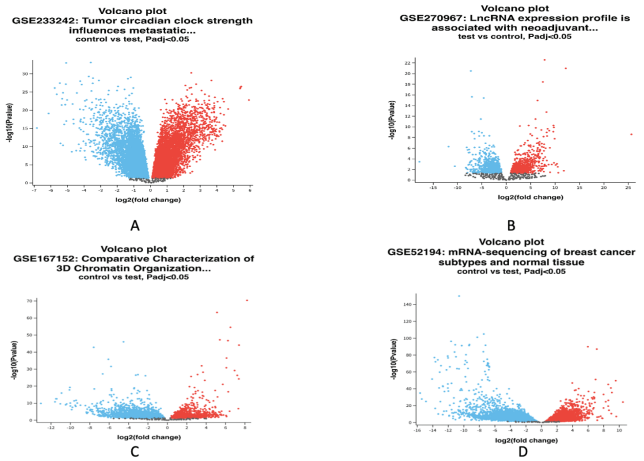


Figure 1: Volcano plots for the significant genes identified from differential gene expression analysis by GEO2R tool. (A) This plot shows the significant DE genes for Luminal A subtype, (B) This plot shows the significant DE genes for Luminal B subtype, (C) This plot shows the significant DE genes for TNBC subtype, (D) This plot shows the significant DE genes for HER2+ subtype. This analysis helped in identifying the differentially expressed genes for each subtype of breast cancer as compared to the normal individuals. The x-axis shows the fold change of the genes and y-axis represents the adjusted p-value.

The network construction for gene interactions of DE genes in STRING showed densely connected networks for each subtype with high-confidence interactions of >0.7, as shown in Figure 2. For Luminal A, the gene interaction network visualised in Cytoscape had 1567 nodes and 2117 edges. The network for the Luminal B subtype had 1564 nodes and 1844 edges. For TNBC, the network constructed and visualised included 1432 nodes and 4122 edges. For the last subtype, HER2+, the network had a total of 1398 nodes and 2059 edges. After performing network analysis in Cytoscape, four different topological measures, namely, degree, betweenness centrality, clustering coefficient, and closeness centrality, helped in identifying top hub genes of the network for each subtype. The top 10 hub genes identified for each breast cancer subtype are listed in Table 1.

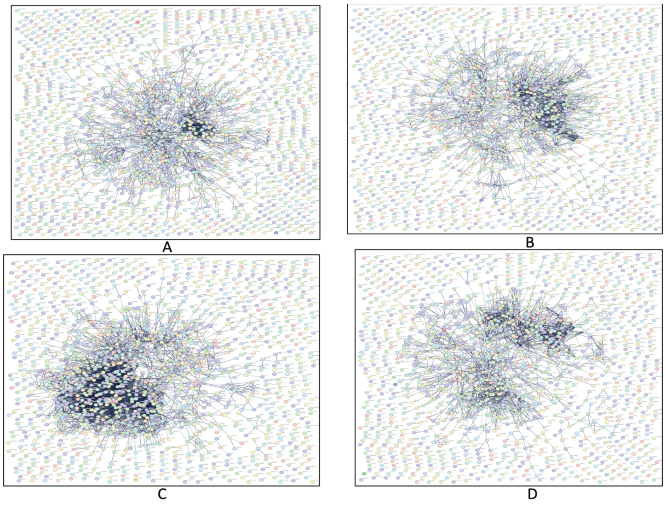


Figure 2: Networks constructed in STRING and visualized in Cytoscape for each of the four subtypes using the significant genes from the DGE analysis. (A) Gene-interaction network of Luminal A subtype, (B) Gene-interaction network of Luminal B subtype, (C) Gene-interaction network of TNBC subtype, (D) Gene-interaction network of HER2+ subtype. Network analysis was carried out to identify the most significant genes of each subtype of breast cancer.

Table 1: A list of top 10 hub genes for each subtype of breast cancer computed on the basis of topological parameters of gene-interaction network.

S.No.	Luminal A	Luminal B	TNBC	HER2+
1.	*RPS27A*	*IL1B*	*RPS27A*	*RPS27A*
2.	*HDAC1*	*IFNG*	*H3C13*	*IL1B*
3.	*RPL11*	*IL10*	*H3C12*	*H3C13*
4.	*RPL5*	*CXCL8*	*CENPA*	*PTPRC*
5.	*MRPL24*	*CXCL10*	*CCNA2*	*RPS8*
6.	*RPS8*	*HSP90AA1*	*H4C5*	*BUB1*
7.	*RPS7*	*CCL2*	*H4C8*	*RPS27*
8.	*CENPA*	*CALML5*	*CDC20*	*RPL5*
9.	*MRPS5*	*ESR1*	*CCNB1*	*FN1*
10.	*RPS27*	*TLR2*	*BUB1*	*CDC20*

The pathway analysis highlighted the top enriched, over-represented Reactome pathways for each subtype of breast cancer. The top 5 enriched pathways for subtype Luminal A were Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal, Amplification of signal from the kinetochores, Chromatin modifying enzymes, Chromatin organisation, and Cytokine Signalling in the Immune system. The top 5 enriched pathways for subtype Luminal B were Interleukin-10 signalling, Signalling by Interleukins, Chemokine receptors bind chemokines, Peptide ligand-binding receptors, and Cytokine Signalling in the Immune system. The top 5 enriched pathways for subtype HER2+ were EML4 and NUDC in mitotic spindle formation, Cytokine Signalling in the Immune system, RHO GTPases Activate Formins, Amplification of signal from the kinetochores, and Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal. The top 5 enriched pathways for subtype TNBC

DOI: 10.36838/v8i3.71

were Cell Cycle, Mitotic, Cell Cycle, Cell Cycle Checkpoints, M Phase, and G2/M Checkpoints.

The top 25 hub genes of the gene-interaction network were checked for their presence in the top 25 pathways. For the Luminal A subtype, of the top 25 genes, the top 13 were found in the pathway, and 12 were not in the enriched pathways. For Luminal B, 15 of the top 25 genes were found in the pathways, and 10 were not found. In the triple-negative (TNBC) subtype, of the top 25 genes, 24 were found, and 1 was not found in the top 25 pathways. Lastly, for the HER2+ subtype, 16 genes were found, and 9 genes were not visible in the top pathway.

This study is about different subtypes of breast cancer and makes use of different bioinformatic approaches, such as DGE analysis, network analysis, and pathway analysis, to identify key biomarkers related to breast cancer. The primary objective of this paper was to perform a comprehensive set of analyses to identify subtype-specific biomarkers and determine their potential role in treatment processes. The results indicated that the genes found were significantly upregulated in individuals with breast cancer. Gene enrichment analysis revealed that 'Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal' is heavily involved in the progression of luminal A breast cancer. 'Cell Cycle, Mitotic' is most involved for triple-negative, and 'EML4 and NUDC in mitotic spindle formation' for HER2+ breast cancer. These results provided insights into more personalised treatments and targeted therapies for patients.

In previous studies conducted by ASC Omega suggested that immune-related gene expression has a pivotal role in the progression of this disease. The identification of key prognostic gene signatures, including those derived from WGCNA and LASSO analysis, serves as a critical biomarker for breast cancer, which adds to the already existing evidence that suggests there is a genetic predisposition to breast cancer. In the study, a similar approach was used as this research paper, which involved reliance on TCGA datasets contributing to the growing knowledge of subtype-specific biomarkers' relevance in treatments.

The identification of the subtype-specific biomarkers as diagnostic markers suggests their role in developing more personalised treatments for breast cancer subtypes. Additionally, these findings could also lead to the discovery of more therapeutic targets for patients with this disease. Conducting pathway and network analysis provided an overall view of the systems biology behind different subtypes of breast cancer.

A major strength of this study is the comprehensive approach and use of multiple bioinformatic analyses and tools than other literature, which allowed for deeper analysis of the genomic data while focusing on each breast cancer subtype. One limitation could be the use of datasets that are available to anybody, as they may not fully represent the diversity of the global population and may not be classified. The biomarkers identified in this study can further be subjected to the identification of the lead compounds by using more advanced bioinformatic approaches, such as molecular docking and molecular dynamics simulation. The viability of these biomarkers can be further experimentally validated for their suitability in the clinical setting. Additionally, expanding the dataset to include a more diverse population could enhance the generalizability of the findings.

In order to enhance the therapeutic relevance of these results, wet-lab confirmation of the hub genes and pathways would be crucial. In order to precisely quantify changes in gene expression, methods like quantitative polymerase chain reaction (qPCR) might be employed to confirm the differential expression of candidate genes at the mRNA level between samples of breast cancer tissue and matched controls. Additionally, patient-derived tumor sections may be subjected to immunohistochemistry (IHC) to verify the protein-level expression of these biomarkers, enabling the spatial localization of the proteins within the tissue microenvironment. To guarantee that the discovered biomarkers are not only statistically significant but also functionally confirmed for their potential in diagnosis and treatment, these techniques would close the gap between in silico predictions and biological relevance.

■ Conclusion

In conclusion, this study identifies the biomarkers for each subtype of breast cancer by making use of bioinformatic approaches such as differential gene expression analysis, network analysis, and pathway analysis, which could further be experimentally validated for their potential in personalised treatments. *RPS27A* was identified as a key biomarker across 3 subtypes (Luminal A, HER2+, and TNBC), determining its potential as a therapeutic target. *RPS27A*'s recurring occurrence across several subtypes has biological significance because, despite being traditionally thought of as a housekeeping gene necessary for ribosome function, new research indicates it also plays oncogenic roles, including promoting proliferation and altering the ubiquitin–proteasome pathway. Reactome pathways showed enrichment of pathways related to mitotic spindle formation, cytokine signalling, chromatin organisation, and cell cycle regulation. Hub network genes, *HDAC1*, *IFNG*, *H3C13*, and *IL1B*, were identified as unique biomarkers for each subtype, Luminal A, Luminal B, HER2+, and TNBC, respectively. Overall, this study paves the way for therapeutic intervention of key biomarkers for each specific subtype of breast cancer that can help in personalised treatments.

■ Acknowledgments

MD would like to acknowledge Aashna Saraf, Founder of CreatED, for providing valuable feedback and guidance throughout the project.

■ References

1. Wilkinson, L.; Gathani, T., Understanding breast cancer as a global health concern. *British Journal of Radiology* **2022**, *95* (1130), 20211033.
2. Dhup, S.; Ramamurthy, R., A genome atlas mapping cancers in India. *Nature India* **2024**.
3. Eroles, P.; Bosch, A.; Alejandro Pérez-Fidalgo, J.; Lluch, A., Molecular biology in breast cancer: Intrinsic subtypes and signaling pathways. *Cancer Treatment Reviews* **2012**, *38* (6), 698-707.

4. Zhou, G. Q.; Lv, J. W.; Tang, L. L.; Mao, Y. P.; Guo, R.; Ma, J.; Sun, Y., Evaluation of the National Comprehensive Cancer Network and European Society for Medical Oncology Nasopharyngeal Carcinoma Surveillance Guidelines. *Front Oncol* **2020**, *10*, 119.
5. Gajbe, B.; Das Kurmi, B.; Kenwat, R.; Paliwal, R.; Paliwal, S. R., Chapter 1 - Breast cancer: introduction. In *Targeted Nanomedicine for Breast Cancer Therapy*, Paliwal, S. R.; Paliwal, R., Eds. Academic Press: 2022; pp 3-26.
6. Derakhshan, F.; Reis-Filho, J. S., Pathogenesis of Triple-Negative Breast Cancer. *Annu Rev Pathol* **2022**, *17*, 181-204.
7. Alam, M. S.; Sultana, A.; Reza, M. S.; Amanullah, M.; Kabir, S. R.; Mollah, M. N. H., Integrated bioinformatics and statistical approaches to explore molecular biomarkers for breast cancer diagnosis, prognosis and therapies. *PLoS One* **2022**, *17* (5), e0268967.
8. Zhang, X., Molecular Classification of Breast Cancer: Relevance and Challenges. *Archives of Pathology & Laboratory Medicine* **2022**, *147* (1), 46-51.
9. Guo, L.; Kong, D.; Liu, J.; Zhan, L.; Luo, L.; Zheng, W.; Zheng, Q.; Chen, C.; Sun, S., Breast cancer heterogeneity and its implication in personalized precision therapy. *Exp Hematol Oncol* **2023**, *12* (1), 3.
10. Jiménez-Santos, M. J.; García-Martín, S.; Fustero-Torre, C.; Di Domenico, T.; Gómez-López, G.; Al-Shahrour, F., Bioinformatics roadmap for therapy selection in cancer genomics. *Mol Oncol* **2022**, *16* (21), 3881-3908.

Authors

Mihika Deora is presently a Grade 11 student at Oberoi International School, OGC. She has a keen interest in biology and chemistry and has completed two internships and courses in the field. Additionally, her academic curiosity extends to areas such as neuroscience, genetics, and biochemistry, reflecting her passion for interdisciplinary learning and research.

Nirupma Singh is a Bioinformatics Scientist with a doctorate in Biotechnology and Bioinformatics from the University of Delhi, with six years of hands-on research and development experience. Her journey is marked by a robust foundation in Machine Learning and Python, with five years of expertise. Proficient in Linux and cloud servers like AWS. She excels in structural biology, systems biology, protein/gene network analysis, data mining, and computational genomics.

DOI: 10.36838/v8i3.71

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Selective Inhibition of MMP-2 Protein for Preventing Decorin Cleavage Linked with Hypermobile Ehlers-Danlos Syndrome

Shreejith Krishnamoorthy

Central Bucks High School South, 1100 Folly Road, Warrington, Pennsylvania, 18976, USA; krishnamoorthyshreejith@gmail.com
Mentor: Ryan Ford

ABSTRACT: Hypermobile Ehlers-Danlos Syndrome (hEDS), a connective tissue and joint disorder that affects millions of people, causes chronic pain, increases the risk of injury, and other diseases. Up to 25% of the world population is also predicted to be affected by this condition in the future, making targeted therapeutics essential. Recent discoveries in the methylenetetrahydrofolate biological pathway associated with hEDS have enabled this study to attempt to develop targeted therapeutics. This study leverages and develops the rationale behind the inhibition of the matrix metalloprotease-2 (MMP-2) enzyme to act as a therapy for hEDS. Small molecule inhibitors were chosen for the mode of inhibition due to ease of production, effectiveness, and analysis. Identification of top inhibitors/drug candidates for MMP-2 was conducted *in silico* utilizing thermochemical and structural metrics computed in Maestro, a computational chemistry software. Drug candidates were also sorted based on the biological and cellular context of the disease rather than just potency. The top inhibitor, ChemBridge 3040, is predicted to have a binding affinity of -80.93 kilocalories per mole to the binding site of MMP-2. Experimental validation of the drug targets required recombinant MMP-2 expression. Two novel expression cultures were developed to validate *in vitro* efficacy.

KEYWORDS: Chemistry, Computational Chemistry, Protein Inhibitors, Protein Expression, Matrix Metalloproteinase.

Introduction

Hypermobile Ehlers-Danlos Syndrome (hEDS) is a connective tissue and joint disorder that is diagnosed in 1 in 500 people globally and is predicted to affect up to 25% of people in an undiagnosed and mild manner.[1,2] Inflicted patients suffer from chronic muscular and bodily pain; increased risk of injury and disease; loss of connective tissue function; joint pain/instability; and various other complications.[3] Primarily caused by genetic mutations and inherited through family lineages, the incidence of hEDS cases is steadily increasing. Despite the increasing need for therapeutics, there are currently no targeted therapies—only retrospective painkillers and vitamin supplements.[4,5]

The predominant mutation in hEDS patients is the suppression of the methylenetetrahydrofolate reductase (MTHFR) gene, resulting in haltered expression of the MTHFR enzyme.[6] Metalloprotease-2 (MMP-2) is an enzyme that aids in the catalysis of cleavage within various extracellular matrix (ECM) peptides and proteins. Recent discoveries have heavily associated MMP-2 with a complex biological pathway involved in the advancement of hEDS.[7]

Typically, MTHFR would catalyze a reaction to form 5-methylenetetrahydrofolate (5-MTHF) from 10-methylenetetrahydrofolate (10-MTHF), as seen in Figure 1.[7]

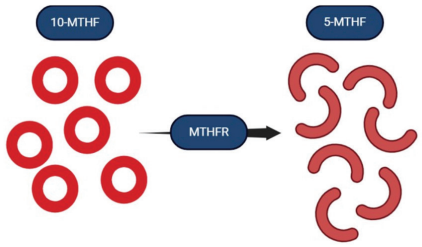


Figure 1: Visualization of the MTHFR enzyme cleaving the 10-MTHF vitamin into the 5-MTHF vitamin.

5-MTHF would subsequently inhibit the MMP-2 enzyme. This inhibition would result in an organized extracellular matrix (ECM) and inactivated fibroblast, a cell mechanism responsible for injury repair, as seen in Figure 2.[7]

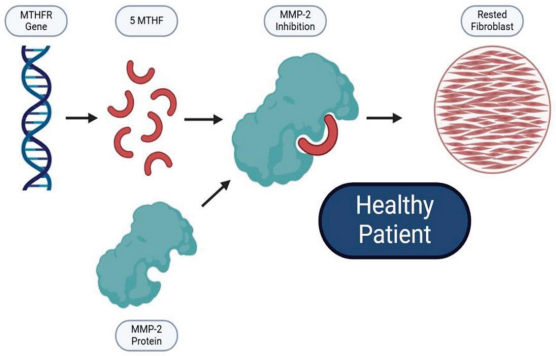


Figure 2: Visualization of the MTHFR gene mechanism, the MTHFR gene produces MTHFR, which produces 5-MTHF, which then inhibits the MMP-2 enzyme, leading to proper regulation of the debris in the ECM and resting fibroblast.

Downregulation mutations in the MTHFR gene, found in hEDS patients, result in reduced MTHFR expression, leading

DOI: 10.36838/v8i3.76

to the reduced production of 5-MTHF. This allows MMP-2 to be overactivated and cleave decorin, a peptide that increases the availability of transforming growth factor β (TGFβ) and thus, activates fibroblasts without any injury to repair. This leads to fibrosis, which is the scarring and damage of tissue caused by falsely activated fibroblasts, as seen in Figure 3.[7]

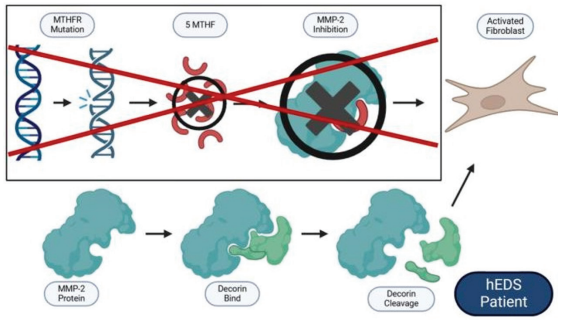


Figure 3: Visualization of the downregulated MTHFR gene mechanism, the MTHFR gene produces less MTHFR, which produces less 5-MTHF, which then cannot inhibit the MMP-2 enzyme, leading to MMP-2 cleaving decorin, resulting in increased TGFβ and, in turn, activating fibroblast or inducing fibrosis.

The mechanism described above allows the supplementation of 5-MTHF to be a possible therapeutic for hEDS. However, clinical supplementation of this compound had minimal to no results.[7] Further analysis determined that another biological process involving methionine synthase actively converts 5-MTHF into 10-MTHF, reducing its efficacy drastically, as seen in Figure 4.[7]

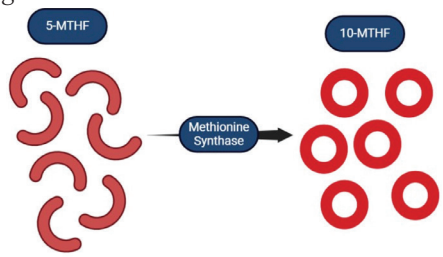


Figure 4: Visualization of the methionine synthase mechanism, methionine synthase catalyzes a reaction to form 10-MTHF from 5-MTHF.

As a method to counteract these results, testing synthetic inhibitors of MMP-2 could prove to be an effective therapeutic for hEDS. These inhibitors would be selective yet unique, allowing them to inhibit MMP-2 without being catalyzed by methionine synthase, as seen in Figure 5.

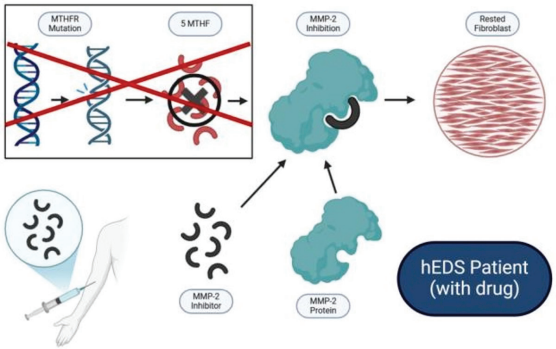


Figure 5: Visualization of the downregulated MTHFR gene mechanism with the proposed MMP-2 inhibitors as drug candidates, the synthetic MMP-2 inhibitors could compensate for the reduced 5-MTHF production by inhibiting MMP-2 and returning normal function in the ECM.

Methods

Methods Overview:

The basic procedure of the experiment consisted of 3 main phases as follows: MMP-2 Protein Expression, Computational Inhibitor Screening, and Experimental/Activity Assays (Figure 6).

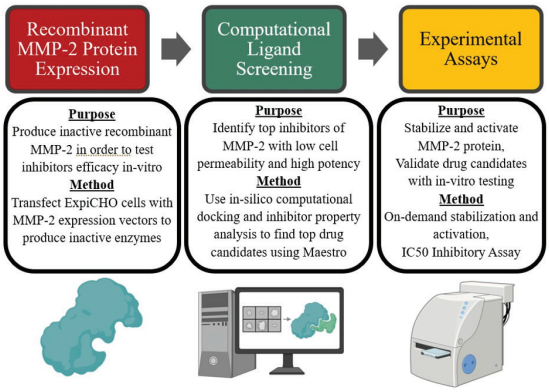


Figure 6: Methodology phases are separated into Recombinant MMP-2 Protein Expression, Computational Ligand Screening, and Experimental Assays, with purposes and methodologies briefly described.

Expression of Recombinant MMP-2 Enzyme:

Validation of potential inhibitors of the MMP-2 protein could only be done utilizing *in vitro* testing. MMP-2 protein could not be obtained from patient donors due to logistical and legal constraints. Replication of *in vivo* activity could be shown through recombinantly expressed proteins. The study utilized Expi Chinese Hamster Ovary (ExpiCHO™) cells and protocols to produce MMP-2 using two distinct expression vectors, recombinant DNA used to produce protein.[8] CAUTION: The mentioned methods and protocols can expose individuals to biological, chemical, and mechanical hazards.

In-vivo expression of the MMP-2 protein produces a propeptide region along with the protein itself to fold it properly, so recombinant expression vectors were ordered with the propeptide amino acid sequence along with the protein amino acid sequence. The propeptide region keeps the MMP-2 enzymatically inactive until cleaved using the membrane type 1 matrix metalloproteinase (MT1-MMP) enzyme *in vivo* (Figure 7).[9] However, with no access to MT1-MMP, the study mutated the amino acid sequence to include a Furin enzyme cleavage site (RAKR) since the study had access to Furin (Figure 8).[10] This resulted in the study producing two separate MMP-2 proteins, one with the wild-type amino acid sequence known as MMP-2-Native and one with the Furin cleavage site implemented between the propeptide sequence and the MMP-2 protein sequence (Table 1). The study chose to produce MMP-2-Native even though MT1-MMP was not available, since there are possibilities of autocatalysis and autoactivation, which would not require MT1-MMP. Furthermore, MMP-2-Furin could alter the protein's structure, resulting in a completely different protein, so MMP-2-Native was produced as a backup.

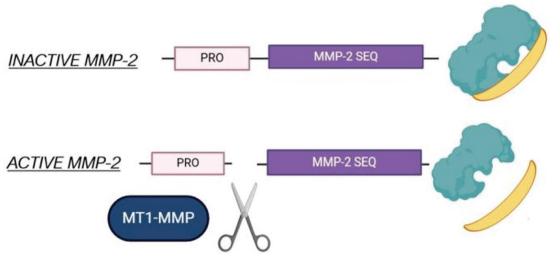


Figure 7: Visualization of *in vivo* activation of MMP-2, MMP-2 remains inactive until MT1-MMP cleaves off the propeptide sequence.

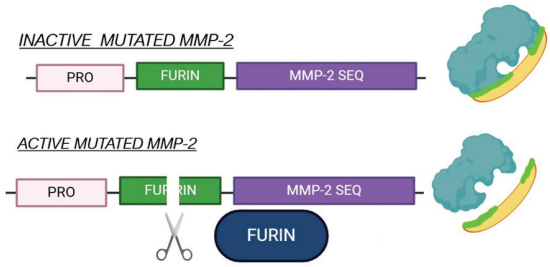


Figure 8: Visualization of suggested *in vitro* activation of MMP-2, MMP-2 remains inactive until Furin cleaves off the propeptide using the Furin cleavage site, which was implemented.

Table 1: The table displays the two variants of MMP-2, which will be produced in this study, along with the plasmid amino acid sequence color-coded by each segment of the protein, labeled in the notes section.

MMP-2-Native and MMP-2-Furin Amino Acid Sequence

MMP-2 Variant	Amino Acid Sequence
MMP-2-Native	MGVKVLFALICIAVAEAAPSPIIKFPGDVAPKTDKELAVQYLNTFY GCPKESCNLFVLKDTLKKMQKFFGLPQTGDLDQNTIETMRKPRCGN PDVANYNFFPRKPKWDKNQITYRIIGYTPDLDPETVDDAFARAFQV WSDVTPLRFSRIHDGEADIMINFGRWEHGDGYPFDGKDGLLAHAFA PGTGVGGDSHFDDDELWTLGEGQVVRVKYGNADGEYCKFPFLFNGK EYNSCTDTGRSDGFLWCSTTYNFEKDGKYGFCPHEALFTMGGNAEG QPCKFPFRFQGTSYDSCTTEGRTDGYRWCGTTEDYDRDKKYGFCPE TAMSTVGGNSEGAPCVFPFTFLGNKYESCTSAGRSDGKMWCATTAN YDDDRKWGFCPDQGYSLFLVAAHAFGHAMGLEHSQDPGALMAPIYT YTKNFRLSQDDIKGIQELYGASPD**H6**
MMP-2-Furin	MGVKVLFALICIAVAEAAPSPIIKFPGDVAPKTDKELAVQYLNTFY GCPKESCNLFVLKDTLKKMQKFFGLPQTGDLDQNTIETMRKPRCGN P**RAKR**YNFFPRKPKWDKNQITYRIIGYTPDLDPETVDDAFARAFQV WSDVTPLRFSRIHDGEADIMINFGRWEHGDGYPFDGKDGLLAHAFA PGTGVGGDSHFDDDELWTLGEGQVVRVKYGNADGEYCKFPFLFNGK EYNSCTDTGRSDGFLWCSTTYNFEKDGKYGFCPHEALFTMGGNAEG QPCKFPFRFQGTSYDSCTTEGRTDGYRWCGTTEDYDRDKKYGFCPE TAMSTVGGNSEGAPCVFPFTFLGNKYESCTSAGRSDGKMWCATTAN YDDDRKWGFCPDQGYSLFLVAAHAFGHAMGLEHSQDPGALMAPIYT YTKNFRLSQDDIKGIQELYGASPD**H6**

Note. RAKR = furin cleavage site - MMP2 propeptide - MMP2 catalytic domain - His6 tag

Table 2: The table displays the amino acid sequence of both MMP-2 Native and MMP-2 Furin, which have been aligned against each other to emphasize the differences.

```
Native 1    MGVKVLFALICIAVAEAAPSPIIKFPGDVAPKTDKELAVQYLNTFYGCPKESCNLFVLKD  60
Furin  1    MGVKVLFALICIAVAEAAPSPIIKFPGDVAPKTDKELAVQYLNTFYGCPKESCNLFVLKD  60

Native 61   TLKKMQKFFGLPQTGDLDQNTIETMRKPRCGNPDVANYNFFPRKPKWDKNQITYRIIGYT  120
Furin  61   TLKKMQKFFGLPQTGDLDQNTIETMRKPRCGNPRAKRYNFFPRKPKWDKNQITYRIIGYT  120

Native 121  PDLDPETVDDAFARAFQVWSDVTPLRFSRIHDGEADIMINFGRWEHGDGYPFDGKDGLLA  180
Furin  121  PDLDPETVDDAFARAFQVWSDVTPLRFSRIHDGEADIMINFGRWEHGDGYPFDGKDGLLA  180

Native 181  HAFAPGTGVGGDSHFDDDELWTLGEGQVVRVKYGNADGEYCKFPFLFNGKEYNSCTDTGR  240
Furin  181  HAFAPGTGVGGDSHFDDDELWTLGEGQVVRVKYGNADGEYCKFPFLFNGKEYNSCTDTGR  240

Native 241  SDGFLWCSTTYNFEKDGKYGFCPHEALFTMGGNAEGQPCKFPFRFQGTSYDSCTTEGRTD  300
Furin  241  SDGFLWCSTTYNFEKDGKYGFCPHEALFTMGGNAEGQPCKFPFRFQGTSYDSCTTEGRTD  300

Native 301  GYRWCGTTEDYDRDKKYGFCPETAMSTVGGNSEGAPCVFPFTFLGNKYESCTSAGRSDGK  360
Furin  301  GYRWCGTTEDYDRDKKYGFCPETAMSTVGGNSEGAPCVFPFTFLGNKYESCTSAGRSDGK  360

Native 361  MWCATTANYDDDRKWGFCPDQGYSLFLVAAHAFGHAMGLEHSQDPGALMAPIYTYTKNFR  420
Furin  361  MWCATTANYDDDRKWGFCPDQGYSLFLVAAHAFGHAMGLEHSQDPGALMAPIYTYTKNFR  420

Native 421  LSQDDIKGIQELYGASPDHHHHHH  444
Furin  421  LSQDDIKGIQELYGASPDHHHHHH  444
```

The protein was purified using the His-6 tag at the end of the protein and immobilized metal affinity chromatography (IMAC) as stated in well-known protocols.[11] To ensure that the protein produced was indeed MMP-2-Native and MMP-2-Furin, a protein Ultraviolet (UV) Scan and a sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were conducted according to well-established protocols.[12,13]

Computational Identification of Synthetic Inhibitors of MMP-2:

Identification of effective inhibitors manually through trial-and-error processes is time-consuming, ineffective, and expensive. As a result, the study utilized a computational chemistry program called Maestro to computationally identify the best inhibitors for the MMP-2 protein. Maestro was the only computational chemistry software available for the study. Various arrays of analysis were used. The ligands that were tested come from a bank of 150,000 compounds that are available for use in the study. The purpose is to find inhibitors with low IC50 values. (Inhibitory Concentration 50%)—a metric of inhibitory potency.

The study strategically filtered through the 150,000 compound library using a workflow that is shown in Figure 9. The figure will include the number of compounds the study expects to remain in each section of the workflow. Refer to the Maestro resource catalog for specific information (such as methodology and protocols) on any specific section.[14]

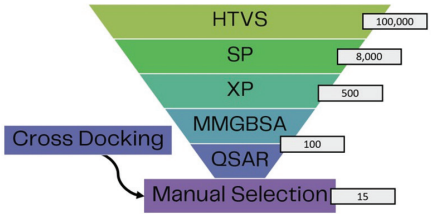


Figure 9: Visualization of computational workflow for filtering compounds with the highest predicted inhibitory potency in Schrodinger Maestro's application.

The computational workflow, which was used to identify drug candidates, can be seen in a visual representation above in Figure 9.

High Throughput Virtual Screening (HTVS): Screens 150,000 compounds and eliminates 50,000 compounds based on the physical size of the compound in comparison to the active site of the protein, i.e., whether the compound is too large or too small for the active site.[14]

Standard Precision (SP) Docking: Screens 100,000 compounds and eliminates 92,000 compounds based on predicted thermochemical favorability presented as a docking score (the more negative, the more thermochemically favorable). This metric is produced through computational rigid pose docking, rigidly placing various compounds in the active site and predicting the various forces of attraction and repulsion (hydrogen bonds, pi-cation bonds, etc.).[14]

Extra Precision (XP) Docking: Screens 8,000 compounds and eliminates 7,500 based on docking scores similar to SP Docking, however, with more computational power and time provided.[14]

DOI: 10.36838/v8i3.76

Molecular Mechanics Generalized Born Surface Area (MMGBSA) Solvation Model: Screens 500 compounds and eliminates 400 compounds based on thermochemical favorability presented as a ΔG Bind Score (the more negative, the more thermochemically favorable). The ΔG Bind Score is expressed in kcal/mol of reaction. This metric is produced through computational dynamic free binding docking, which allows the protein and the compound to rotate and move (similar to in-vivo expectations) and records the various forces of attraction and repulsion (hydrogen bonds, pi-cation bonds, etc.).[14]

Quantitative Structure Activity Relationship (QSAR): Creates a Machine Learning Model given known inhibitors of MMP-2 in order to predict the IC50 values of screened compounds. Around 6,000 known MMP-2 Inhibitors were used to generate this model. All 100 remaining compounds had IC50 values predicted for them.[14]

Manual Selection: 15 compounds were chosen based on the cell permeability of the compound and the QSAR-generated IC50 value.[14,15]

The study utilized the crystal structure and control inhibitor for the computational analysis from the Protein Data Bank (access code: 7XGJ) for all computational analyses.[16] The known inhibitor cited in the paper was also isolated and screened through the workflow above for control validation.[16] This control was chosen because it was one of the few inhibitors of the MMP-2 protein, which had recorded crystal structures in complex with MMP-2.

In Vitro Experimental Assays of MMP-2-Furin and MMP-2-Native:

In order to prepare the recombinantly produced proteins for an IC50 Assay, the study needs to ensure that the proteins produced are stable and enzymatically active.[17] The study chose to use a Thermal Shift Assay to determine the protein's stability through the proxy of how rapidly it denatures using well-documented protocols.[18] The study then chose to activate a portion of the MMP-2-Furin protein using 1 μL of 1 μM sample active Furin and incubate the sample at 32 degrees Celsius for 30 minutes. After that, an enzymatic activity assay was set up to determine whether the activated MMP-2-Furin was indeed active according to a well-documented protocol.[19-20] The specific substrate that the study chose was the MMP-2/MMP-7 Fluorogenic Substrate due to easy access and relatively cost-efficient pricing.[21]

The study aimed to run an IC50 Assay using active MMP-2 and all 15 identified compounds to determine the best-performing drug candidate for further experimentation.[17]

■ Result and Discussion

Expression of Recombinant MMP-2 Enzyme: Results:

After producing and purifying both MMP-2-Furin and MMP-2-Native proteins using well-known protocols elaborated on in the methodology portion of the paper, the purified protein was analyzed using a UV absorbance scan and an SDS-PAGE gel, respectively. The UV absorbance scan initially validated that the study successfully purified the protein at high concentrations. As seen in Figures 9 and 10, the UV Absorbance scan measured the amount of light absorbed at various wavelengths. Typically, all proteins absorb light at 280 nanometers. Both MMP-2-Native (Figure 10) and MMP-2-Furin (Figure 11) had multiple fractions that had high absorbance at the 280-nanometer mark, indicating a high concentration of protein.

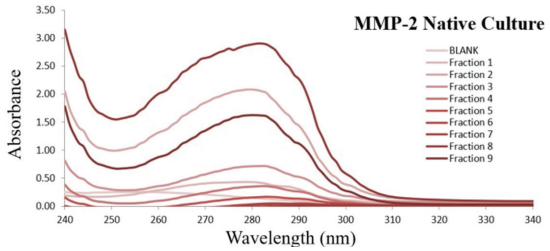


Figure 10: This graph shows the absorbance of 10 different protein fractions of the MMP-2-Native protein culture at different wavelengths. The graph shows that multiple fractions of protein have high absorbance at 280 nanometers, indicating that there is sufficient protein in this purified culture of MMP-2-Native.

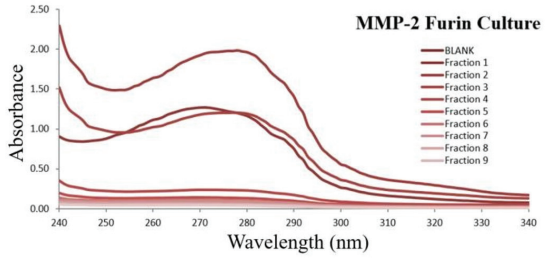


Figure 11: This graph shows the absorbance of 10 different protein fractions of the MMP-2-Furin protein culture at different wavelengths. The graph shows that multiple fractions of protein have high absorbance at 280 nanometers, indicating that there is sufficient protein in this purified culture of MMP-2-Furin.

After validating that both purified protein cultures have protein, the study chose to validate that the protein was indeed MMP-2-Native and MMP-2-Furin by conducting an SDS-PAGE gel, which would reveal the molecular weight of the protein tested. If the study successfully produced MMP-2-Native and MMP-2-Furin, then the SDS-PAGE will indicate that the molecular weight of both proteins would be around 50 kilodaltons.[16] The SDS-PAGE for both MMP-2-Native and MMP-2-Furin is shown below in Figure 12 and Figure 13, respectively.

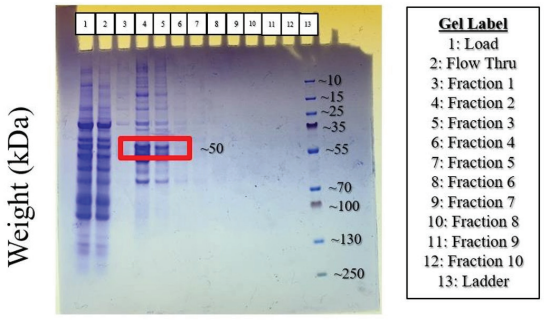


Figure 12: This graph shows the molecular weight of the MMP-2-Native protein culture to be around 50 kilodaltons for fractions 4 and 5 when compared against the protein weight ladder, which shows various milestones of weight in kilodaltons. This validates that the MMP-2-Native culture expressed is MMP-2.

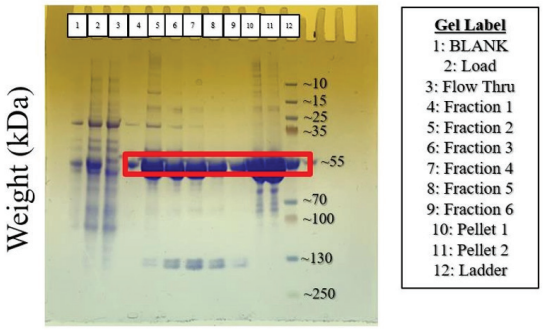


Figure 13: This graph shows the molecular weight of the MMP-2-Furin protein culture to be around 50 kilodaltons for fractions 4 through 12 when compared against the protein weight ladder, which shows various milestones of weight in kilodaltons. This validates that the MMP-2-Furin culture expressed is MMP-2.

The study successfully produced the MMP-2 protein in fractions 4 and 5 in the MMP-2-Native culture and in fractions 4 through 12 in the MMP-2-Furin culture. Fractions 4 and 5 were consolidated into a master stock for MMP-2-Native. Fractions 4 through 12 were consolidated into a master stock of protein for MMP-2-Furin. The concentration of the MMP-2-Native and MMP-2-Furin cultures was determined through a UV nanodrop with well-known protocols for more precision than the typical UV absorbance scan.[22] The nanodrop revealed that the MMP-2-Native and MMP-2-Furin had concentrations of 3.87 milligrams/milliliter and 3.96 milligrams/milliliter, respectively.

From the two SDS-PAGE gels that were conducted, it can be concluded that the two protein samples have molecular weights of approximately 50 kilodaltons, the theoretical weight of MMP-2. Although there are clear, concentrated markings at the desired protein weight markings for both proteins, the various erroneous markings can indicate impurities that may affect the further downstream end of experimentation on these proteins.

Computational Identification of Synthetic Inhibitors of MMP-2: Results:

After conducting all phases of the computational workflow cited in the methodology, the study narrowed down 15 compounds that have a high likelihood of being successful inhibitors for MMP-2 and, in turn, a successful drug candidate for hEDS. The top 15 compounds with their respective name designations, XP docking scores, and MM-GBSA binding affinity scores can be seen below in Table 3.

Table 3: The table displays 15 top compounds that were computationally identified to be potent inhibitors for the MMP-2 protein. This data was attained through computational chemistry metrics identified and explained in the methodology portion.

Compound	XP Docking Score	MM-GBSA Binding Affinity (kcal/mol)
5871-1859	-10.330	-45.42
ChemBridge.3040	-12.454	-44.94
BAS 0649085	-10.324	-44.90
BAS 2282287	-10.302	-43.73
BAS 2825385	-12.469	-42.69
BAS 1538434	-10.803	-41.58
ASN 3794129	-10.806	-36.65
4180-0304	-10.176	-35.00
Y200-1346	-10.371	-34.84
ChemBridge.14986	-11.092	-34.50
ChemBridge.11749	-11.141	-33.14
BTB 14099	-10.256	-33.07
3664-0103	-12.236	-33.00
BAS 0791468	-10.735	-32.85
4358-2895	-10.628	-31.97

When looking at various poses (physical positions in correspondence to the binding site) of all 15 compounds, the top compound, 5871-1859, had poses that seemed improbable, as well as many aromatic rings, which make it a larger compound, making it unideal. Due to those reasons, ChemBridge 3040 was considered the top inhibitor for this study. ChemBridge 3040's position, as well as the various bonds that occur while inhibiting MMP-2, visualized by Maestro, can be seen below in Figure 14.

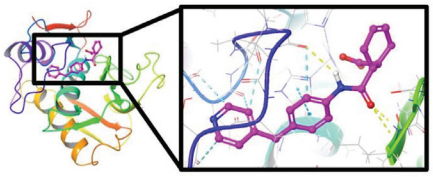


Figure 14: This visualization showcases the MMP-2 protein (rainbow) in complex with the top inhibitor Chembridge 3040 (pink). The various bonds (hydrogen, aromatic, etc.) are also showcased.

The individual compound of ChemBridge 3040 can also be seen isolated alone in Figure 15 below.

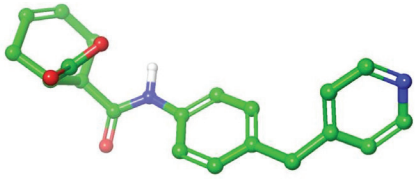


Figure 15: This visualization showcases the ChemBridge 3040 compound with carbons identified in green, nitrogens in blue, oxygens in red, and polar hydrogens in white. Non-polar hydrogens are not displayed.

The interactions between ChemBridge 3040 and the MMP-2 protein can be further analyzed in a ligand interac-

DOI: 10.36838/v8i3.76

tion diagram, shown below in Figure 17. The diagram labels can be seen in Figure 16.

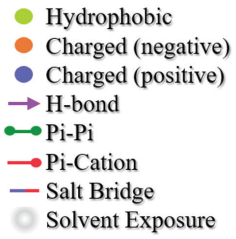


Figure 16: This figure shows the diagram labels for the following ligand interaction diagrams between inhibitory compounds and the MMP-2 protein.

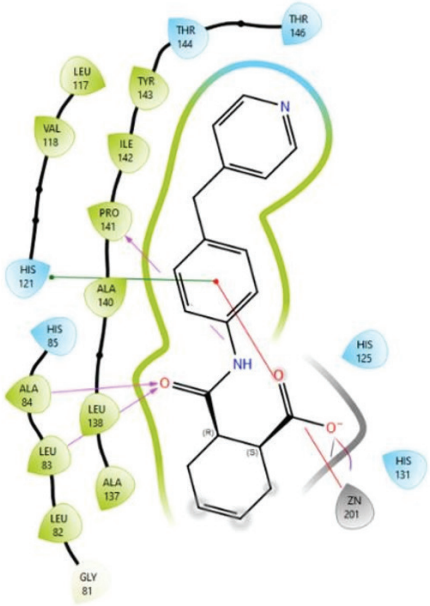


Figure 17: This figure is a ligand interaction diagram between MMP-2's various amino acids and ChemBridge 3040, which was the top inhibitor identified in this study. Please refer to Figure 16 for diagram labels.

As a comparison to ChemBridge 3040, the study also developed a ligand interaction diagram for the control inhibitor, which was identified in the methodology section. This control inhibitor ligand interaction diagram can be seen below in Figure 18.

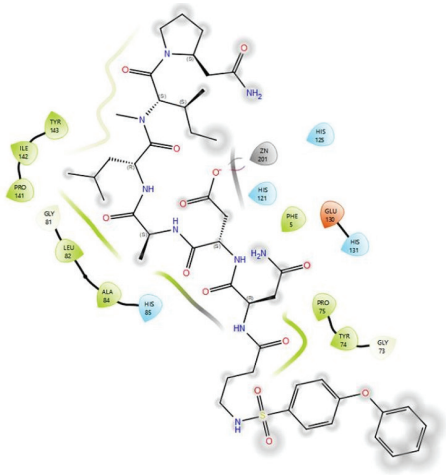


Figure 18: This figure is a ligand interaction diagram between MMP-2's various amino acids and the control inhibitor, which was identified previously by other researchers. Please refer to Figure 15 for diagram labels.

The QSAR Model was created from 6,000 MMP-2 inhibitors.[23] The model was trained to predict the IC50 value for MMP-2 of given compounds. The model used 75% of the dataset to train the model, and 25% of the dataset was used to test the accuracy. The model was deemed sufficiently accurate with an r^2 value of 0.91. A visualization of the model can be seen below in Figure 19.

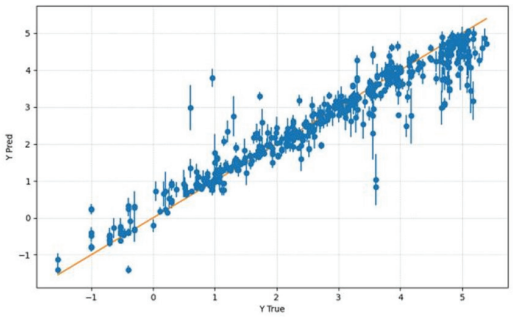


Figure 19: This figure is a graph of the predicted IC50 value of various inhibitors on the Y axis and the experimental IC50 values of the very same inhibitors. The prediction model is represented by the line of best fit shown in yellow.

The model predicted that ChemBridge 3040 would have an IC50 value of 7.877 x 10^{-5} nanomolar. This would be much more potent compared to the control inhibitor's IC50 value of 0.20 nanomolar.

The study then used SwissADME properties to determine the cell permeability of the top compound in comparison to the control inhibitor. A half-boiled egg plot can represent permeability, the ability to pass through various materials, by placing the compound as a point on the plot. The plot has a gray outer region, which represents impermeability through most biological barriers, including the gastrointestinal tract, which is how drugs are typically delivered. The inner white region, as well as the inner yellow region, represents gastrointestinal permeability, meaning that the compound has the potential to be delivered as a drug orally. The innermost yellow region also means that the compound can pass through the blood-brain barrier and can act as a proxy for cell membrane permeability. Since MMP-2 is an extracellular protein, the ideal range for the MMP-2 inhibitory compound is in white since it will be a deliverable drug yet not too permeable such that it permeates cells unnecessarily. Figure 20 below shows the half-boiled egg plot with the top and control compound.

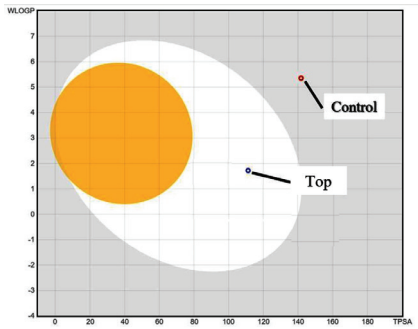


Figure 20: This figure is a graph of the WLOGP on the y axis, which is the octanol-water partition coefficient, which can predict a compound's lipophilicity, and TPSA on the x axis, which is the topological polar surface area, which, in combination with the WLOGP, can predict bioavailability as described above.

The control inhibitor can be seen in the gray area, meaning that the compound is not permeable through the gastrointestinal tract. The top inhibitor, ChemBridge 3040, can be seen in the white region of the graph, meaning that the compound is the ideal permeability that the study sought after in MMP-2 inhibitors.

All these factors indicate that ChemBridge 3040 will be a better inhibitor for this study in comparison to the control inhibitor.

In Vitro Experimental Assays of MMP-2-Furin and MMP-2-Native:

The stability of MMP-2-Native and MMP-2-Furin was assessed through a thermal shift assay, with increases in the temperature of a protein with a fluorescent dye along with it. As the protein denatures, the dye will bind to the amino acids and emit fluorescence, indicating when the protein unravels. A stable protein will have a defined peak in this range of temperatures to indicate its ideal state. If a peak is not defined, that could indicate an improperly folded protein. The Thermal Shift Assay of MMP-2-Native can be seen below in Figure 21.

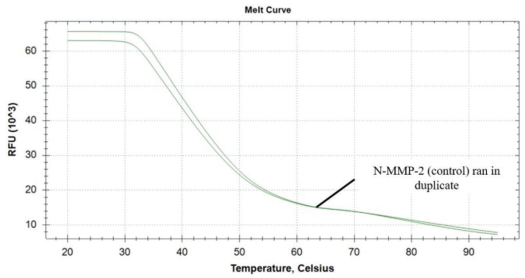


Figure 21: This figure is a graph of the Thermal Shift Assay of MMP-2-Native as described above. RFU stands for relative fluorescent units.

Since MMP-2-Native had no defined peak, the study determined that the protein could be unstable. This resulted in the study supplementing the protein with zinc and calcium ions since they were structural ions that were missing in the expression feeds, leading to unstable protein. The study predicted that the addition of these ions could induce spontaneous folding. The Thermal Shift Assay was run again with the MMP-2-Native and MMP-2-Furin, both before and after ion supplementation. This can be seen below in Figure 22.

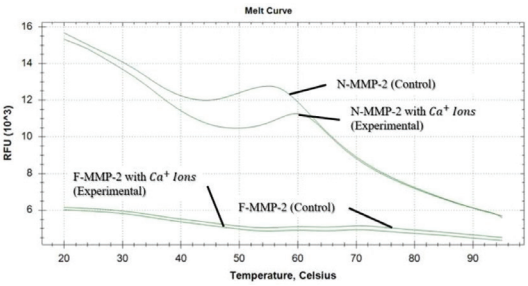


Figure 22: This figure is a graph of the Thermal Shift Assay of MMP-2-Native and MMP-2-Furin as described above. RFU stands for relative fluorescent units.

The supplementation of ions resulted in a more defined curve for MMP-2-Native, but MMP-2-Furin still seems unstable and denatured. Despite unideal conditions, the study moved forward with enzyme activity assays since any doubts on protein stability can be answered if the study could see whether the protein is enzymatically active.

The MMP-2 protein is a protease, meaning that its enzymatic activity involves cutting other proteins. This led the study to use a fluorogenic substrate identified in the methodology, which would typically be cleaved by MMP-2 and, in turn, release fluorescence. The MMP-2-Furin protein was activated with Furin. The MMP-2-Native protein is activated using Tween-20 (detergent) and a reducing agent (DTT) to reduce the thiol group between the propeptide and the native sequence. The enzyme activity assays for both MMP-2-Furin and MMP-2-Native can be seen, respectively, in Figure 23 and Figure 24 below.

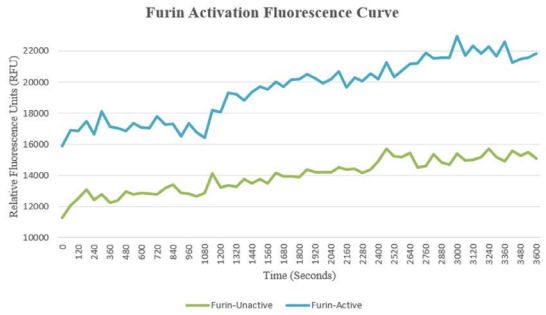


Figure 23: This figure is a graph of the enzyme activity assay of MMP-2-Furin as described above. RFU stands for relative fluorescent units.

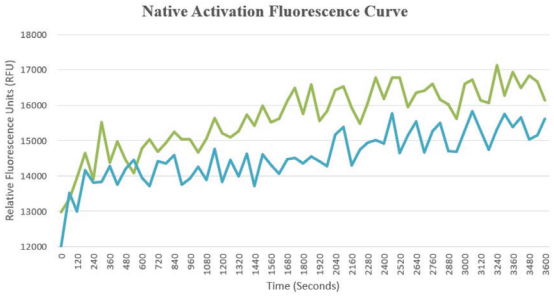


Figure 24: This figure is a graph of the enzyme activity assay of MMP-2-Native as described above. RFU stands for relative fluorescent units.

The increase in RFU may show that both MMP-2-Furin and MMP-2-Native, though to different degrees, were active and properly produced. This also validates that the novel activation mechanism that the study used with the MMP-2-Furin protein might be functional. The enzymatic kinetics of this activity assay need to be further validated with further in vitro testing in order to have sufficient evidence of activation.

Future Works:

In the future, the study plans on conducting the IC50 assay in combination with the active MMP-2 proteins and the fluorogenic substrate to measure the decrease in relative fluorescence emitted, as stated in the methodology. Crystallization of proteins will allow the study to analyze the structure of the protein and how it interacts with the inhibitory compound. This crystallization and x-ray diffraction will occur using known crystallization methods of the MMP-2 protein.[9] The study also plans to crystallize the inactive and active versions of

DOI: 10.36838/v8i3.76

MMP-2-Furin and MMP-2-Native. The crystallization will also be attempted with the MMP-2 active protein in complex with all tested inhibitors. This allows the study to find definitive answers on the effectiveness of various MMP-2 inhibitors in the context of hEDS disease. Further adaptations of this research would include *in vivo* effectiveness in mammalian cells to see if the MMP-2 inhibitors have the anticipated effect, which was hypothesized in the context of hEDS.

Conclusion

This study aimed to develop a therapeutic relief drug for hypermobile Ehlers-Danlos Syndrome (hEDS). Through literature review and research, it was revealed that the MTHFR gene mechanism mutation could be the leading cause of hEDS symptoms. This study chose to target a specifically overexpressed portion of this mechanism, which was the MMP-2 protein. The study analyzed various small-molecule inhibitors to help inhibit the activity of MMP-2 and regulate its function in hEDS patients to provide relief from hEDS symptoms. The study successfully identified 15 potential drug candidates or MMP-2 inhibitors computationally. The study also developed a novel activation mechanism for the MMP-2 protein using a Furin enzyme cleavage site for ease of use in *in vitro* settings. This successful identification of MMP-2 inhibitors and *in vitro* expression of the MMP-2 protein will allow future researchers to validate the study's results and potentially develop a drug utilizing the hypothesis developed in the study.

Acknowledgments

I would like to thank my adult sponsor, Mark Hayden, for allowing me to participate in the Central Bucks School District Biotechnology Partnership Program. I would like to thank my family and my sister for their help in encouraging me. I would also like to thank my mentors, Ryan Ford and Scott Willett, for being patient, guiding me, and educating me through this process.

References

1. The Ehlers-Danlos Society. hEDS Body System. The Ehlers-Danlos Society. https://www.ehlers-danlos.com/heds/.
2. Grahame, R. Joint Hypermobility and Genetic Collagen Disorders: Are They Related? *Arch. Dis. Child.* **1999**, *80* (2), 188–191. DOI: 10.1136/adc.80.2.188.
3. Mayo Clinic. Ehlers-Danlos Syndrome - Symptoms and Causes. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/ehlers-danlos-syndrome/symptoms-causes/syc-20362125.
4. Malfait, F.; Francomano, C.; Byers, P.; Belmont, J.; Berglund, B.; Black, J.; Bloom, L.; Bowen, J. M.; Brady, A. F.; Burrows, N. P.; et al. The 2017 International Classification of the Ehlers-Danlos Syndromes. *Am. J. Med. Genet. C Semin. Med. Genet.* **2017**, *175* (1), 8–26. DOI: 10.1002/ajmg.c.31552.
5. Demmler, J. C.; Atkinson, M. D.; Reinhold, E. J.; Choy, E.; Lyons, R. A.; Brophy, S. T. Diagnosed Prevalence of Ehlers-Danlos Syndrome and Hypermobility Spectrum Disorder in Wales, UK: A National Electronic Cohort Study and Case–Control Comparison. *BMJ Open* **2019**, *9* (11), e031365. DOI: 10.1136/bmjopen-2019-031365.
6. Courseault, J.; Kingry, C.; Morrison, V. E.; Edstrom, C.; Morrell, K.; Jaubert, L.; Elia, V.; Bix, G. J. Folate-Dependent Hypermobility Syndrome: A Proposed Mechanism and Diagnosis. *Heliyon* **2023**, *9* (4), e15387. DOI: 10.1016/j.heliyon.2023.e15387.
7. Ehlers-Danlos News. Mutations That Impair Folate Processing May Be Cause of Hypermobile EDS. https://ehlersdanlosnews.com/news/mutations-impair-folate-processing-cause-hypermobile-study/?cn-reloaded=1.
8. ThermoFisher. *ExpiCHOTM Expression System: User Guide*; 2018. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0014337_expicho_expression_system_UG.pdf.
9. Morgunova, E. Structure of Human Pro-Matrix Metalloproteinase-2: Activation Mechanism Revealed. *Science* **1999**, *284* (5420), 1667–1670. DOI: 10.1126/science.284.5420.1667.
10. Zimmer, G.; Conzelmann, K.-K.; Herrler, G. Cleavage at the Furin Consensus Sequence RAR/KR(109) and Presence of the Intervening Peptide of the Respiratory Syncytial Virus Fusion Protein Are Dispensable for Virus Replication in Cell Culture. *J. Virol.* **2002**, *76* (18), 9218–9224. DOI: 10.1128/jvi.76.18.9218-9224.2002.
11. KTH Royal Institute of Technology. IMAC - Immobilized Metal-Affinity Chromatography. https://static.igem.org/mediawiki/2017/a/a6/Stockholm_IMAC.pdf.
12. Caers, J.; Van Der Juegt, B. Quantification of Proteins by UV-Vis Absorbance Using VICTOR Nivo with Microvolume Plates. https://resources.revvity.com/pdfs/app-quantification-of-proteins-by-uvvis-absorbance-using-victor-nivo.pdf (accessed 2025-06-20).
13. SDS-PAGE. Assay Protocol. http://www.assay-protocol.com/molecular-biology/electrophoresis/denaturing-page.html.
14. Schrödinger. Maestro. Life Science Content Library. https://www.schrodinger.com/life-science/resources/ (accessed 2025-06-20).
15. SwissADME. SwissADME. http://www.swissadme.ch/.
16. RCSB Protein Data Bank. Crystal Structure of Human MMP-2 Catalytic Domain in Complex with Inhibitor. *RCSB PDB* https://www.rcsb.org/structure/7XGJ (accessed 2025-06-20).
17. Coscueta, R.; Pintado, M. M. ACE-Inhibitory Activity Assay: IC50 V1. *protocols.io* **2022**. DOI: 10.17504/protocols.io.q26g74q5kgwz/v1.
18. Bio-Rad. Protein Thermal Shift. https://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_7180.pdf.
19. Knight, C. G.; Willenbrock, F.; Murphy, G. A Novel Coumarin-Labelled Peptide for Sensitive Continuous Assays of the Matrix Metalloproteinases. *FEBS Lett.* **1992**, *296* (3), 263–266. DOI: 10.1016/0014-5793(92)80300-6.
20. Fields, G. B. Using Fluorogenic Peptide Substrates to Assay Matrix Metalloproteinases. *Methods Mol. Biol.* **2010**, 393–433. DOI: 10.1007/978-1-60327-299-5_24.
21. Sigma-Aldrich. Benzil. *Merck* **2025**, *1* (1).
22. Phizicky Lab. Measuring Nucleic Acid/Protein Concentration with NanoDrop 1000 V3.6 Spectrophotometer. https://www.urmc.rochester.edu/MediaLibraries/URMCMedia/labs/kielkopf-lab/documents/nanodrop2022update.pdf.
23. Binding Database. Type IV Collagenase 72kDa. https://www.bindingdb.org/rwd/jsp/dbsearch/PrimarySearch_ki.jsp?tag=pol&submit=Search&target=72%20kda%20type%20iv%20collagenase&polymerid=112,8378,10557,50001223.

Authors

Shreejith Krishnamoorthy is a passionate student at Central Bucks High School South. He is interested in the interdisciplinary field between biotechnology and business. He hopes to

be involved in biotech start-ups that can help patients with rare diseases, such as Ehlers-Danlos Syndrome, get the attention and research that they deserve.

DOI: 10.36838/v8i3.76

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Towards a New Hybrid Stochastic Model for Enhanced Stock Price Estimation: The Heston-GARCH-Levy Model

David S.Y. Park

Campbell Hall Episcopal School, 4533 Laurel Canyon Blvd, Studio City, CA, 91607,, USA; parkd0257@gmail.com
Mentor: Rajit Chatterjea

ABSTRACT: The objective of this paper is to look at a stochastic framework for stock price estimation that builds upon specific advanced structures used in the financial industry. Specifically, we look at and simulate the Heston Model (with correlated Wiener processes with correlation coefficients), and other variants that are considered more accurate with respect to Black-Scholes. Then we look at an improved version of this model with jump conditions and simulate it, then look at the Stochastic Alpha Beta Rho (SABR) model and the Rough Volatility Model. We finally construct a possible Heston-GARCH-Levy Model with Jump Diffusions, which has a Heston Stochastic Volatility, a GARCH for Conditional Heteroskedasticity, as well as Jump Diffusions that introduce discontinuities in the price process, with Levy processes for tailed behavior. This combined model is then analyzed as a Hybrid Stochastic System specifying assumptions, the function spaces and norms needed, existence and uniqueness proofs, handling of hybrid components, regularity and stability bounds, counterexamples when solutions fail, and looking at BIBO stability of the system. Our analysis has resulted in a final stochastic model that has passed conditions for uniqueness and existence of solutions.

KEYWORDS: Mathematics, Analysis, Stochastic, Ito Calculus, Heston-Model, Cox-Ingersoll-Ross, GARCH & Levy, Processes for Jump Diffusions.

Introduction

The modeling of financial markets has seen significant advances, from the seminal Black-Scholes model to stochastic volatility frameworks such as the Heston model. However, real-world asset prices exhibit characteristics that traditional models fail to capture, such as volatility clustering and heavy tails. This paper explores the integration of the Heston model with GARCH processes and Levy noise to address these limitations. The combined model leverages:

- Stochastic volatility dynamics for market fluctuations
- GARCH processes to account for volatility clustering
- Levy noise to model jumps and heavy-tailed return distributions

Background and Past Innovations

1. The Black-Scholes Model: A Starting Point:

The Black-Scholes model, made in 1973, was groundbreaking in financial mathematics, providing a closed-form solution for options pricing.[1] The stock price S(t) is modeled as a geometric Brownian motion:

$$dS_t = S_t\,(rdt + \sigma dW_t)$$

Here, r is the risk-free interest rate, the Greek letter sigma represents the constant volatility, and W(t) is a standard Brownian motion. The Black-Scholes model assumes constant volatility and a lognormal price distribution, making it analytically tractable but inconsistent with real market behavior.[1] Price distributions in real-world asset returns are often skewed, and thus, a normal assumption is inaccurate for accurate price modeling.

2. Limitations of the Black-Scholes Model:

Empirical studies have highlighted major shortcomings in the Black-Scholes framework.

- Volatility smiles and skews: Real option prices show implied volatilities vary with strike price and maturity, contradicting the constant volatility assumption
- Fat Tails in Return Distributions: Empirical stock returns exhibit heavy tails and leptokurtic behavior, unlike the normal distribution predicted by Black-Scholes.
- Sudden price jumps: Financial markets experience large discontinuous moves like crashes and new shocks, which a pure diffusion process fails to capture.
- Volatility clustering: Periods of high volatility tend to be followed by more volatility, a missing feature in the constant σ assumption.

Below is a sample simulation of the Black-Scholes model tested against Apple Stock prices.

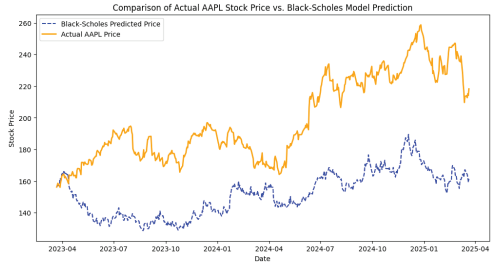


Figure 1: Predicted AAPL prices against Black Scholes. Black-Scholes model simulation was tested against AAPL stock prices for the last 2 years, showing extreme errors and only a few solutions at the beginning of the simulation. Simulation was done with the Monte-Carlo simulation method, and significant gaps and high residuals are present between the two simulations.

DOI: 10.36838/v8i3.85

Notice the significant gaps between the predicted and actual prices. 1's comparisons have nearly no solutions, with only incredibly small solutions (where intersections occur) happening at the beginning. Thus, while the figure demonstrates that the Black-Scholes model can predict in the short run, it is eventually ineffective over a span of weeks. The residuals are very large, and thus the model has weak predictions due to the reasons mentioned previously. It can, however, be argued that this is simply due to computational errors. The simulation is utilizing incredible amounts of approximation methods to predict stock prices in the following period. It is analogous to Euler's method or linear approximation for ordinary calculus, using the tangent line and derivative at step size increments to predict future values of solutions to differential equations. As we span out in time, approximations capture more inaccuracies. However, as we will show in the results section, this error does not inhibit the Heston-GARCH-Levy to the extent of the Black Scholes errors, meaning the HGL model is indeed more mathematically effective.

Methods

The Heston model here considers stochastic volatility, adding a dynamic stochastic evolution of variance with a second differential equation.[2] The continuous Heston variance evolution follows a mean-reverting Cox-Ingersoll Ross process.[3]

$$dv_t = \kappa(\theta - v_t)\mathrm{d}t + \sigma\sqrt{v_t}\,dW_t^v$$

This allows volatility to fluctuate randomly over time, improving the model's fit to market data. Here, κ is the mean-reverting rate, θ is the long-term memory, and the product of the square root volatility, and σ gives us the volatility of volatility. However, the Heston model still assumes continuous price paths, failing to capture market jumps and fat tails.

Thus, to capture market irregularities, we incorporate a Levy jump process following a Poisson process, leading to our Stochastic Differential Equation (SDE) for the dynamics of the stock price as:

$$dS_t = S_t(rdt + \sigma dW_t + J_t dN_t)$$

The additional J(t) term is this jump process following a Poisson distribution N(t) with intensity denoted by the Greek letter lambda. J(t) are specifically i.i.d. jump sizes. Levy processes further generalize jumps by allowing infinite activity jumps, such as the Variance Gamma and Normal Inverse Gaussian, capturing fat tails and skewness in stock returns more effectively than Merton's model.[4]

Lastly, we incorporate a GARCH component into our model. To better model long-memory effects in volatility, we incorporate a GARCH (Generalized Autoregressive Conditional Heteroskedasticity) process that innovates upon the general ARCH(1,1) model for volatility clustering.[5] The GARCH model is defined as follows:

$$v_{t+1} = w + \alpha\epsilon_t^2 + \beta v_t,\ \epsilon_t = \sqrt{v_t}Z_t$$

Unlike the Heston model, GARCH models account for discrete-time volatility clustering, making them effective in high-frequency financial modeling. Notice its recursive behavior, as there is a v_{t+1} that is dependent on the previous period ϵ_t to predict future volatility fluctuations. Therefore, the Hybrid-GARCH-Levy Model with Jump Diffusions integrates:

- Heston's Stochastic volatility (continuous reverting behavior)
- GARCH dynamics (discrete volatility clustering)
- Levy-driven jumps (heavy tails and skewness)
- Poisson processes (sudden price jumps)

Advantages over previous models include capturing volatility clustering (GARCH and Heston components ensure time-varying volatility), modeling extreme market moves (Levy jumps introduce fat tails and rare events), flexible skewness and kurtosis (the model accommodates asymmetric return distributions), and better option pricing fits (the combination of stochastic volatility and jumps corrects the implied volatility smile). The Heston-GARCH-Levy model with Jump Diffusions represents an improvement over classical models, allowing for realistic asset price dynamics. By addressing volatility clustering, jumps, and heavy tails, it better explains market phenomena such as crashes, skewed option prices, and persistent volatility shocks. The model is also a hybrid model, meaning that there are continuous and discrete dynamics occurring simultaneously. We will prove such a structure exists and contains unique solutions in the following section. And so, we propose a new system of SDEs to capture real-world market dynamics, for which we will prove in the following section:

$$dS_t = \mu S_t dt + \sqrt{v_t}S_t dW_t + S_t J_t dN_t\ ,$$

$$dv_t = \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}\,dW_t^v$$

$v_t = \alpha + \beta v_{t-1} + \gamma\epsilon_{t-1}^2$ (Note: The v(t+1) was simply rewritten as v(t), moving the starting point of the recursion sequence back by 1).

- S(t) is the stock price at time t,
- v(t) is the stochastic variance process,
- μ is the drift of the stock price,
- κ is the mean-reversion rate of variance
- σ is the volatility of variance
- W(t) and Z(t) are correlated Brownian motions with a correlation coefficient ρ
- J(t) models the jump sizes,
- N(t) is a Poisson process modeling the jump occurrences,
- λ is the intensity of the Poisson process.

Theorems for Stability Proofs (Methods)

A major component of this paper will be to extend results from Ordinary Differential Equations to Stochastic Differential Equations, under appropriate conditions. The key intuition is that an SDE is an ODE perturbed by noise, often in the form of Brownian motion. In particular, we are concerned with four specific ideas:

- Existence and Uniqueness: How ODE theorems extend to SDEs
- Stability and Convergence: How solutions behave under perturbations
- Gronwall's Inequality for SDEs: A key inequality that carries over from ODEs

• Flow properties and Diffeomorphisms: How ODE flow maps extend to stochastic settings

Consider a deterministic ODE of the form:

$$\frac{dx_t}{dt} = f(x_t),\ \ x_0 = x$$

By Picard's existence and uniqueness theorem, if f(x) is Lipschitz continuous, there exists a unique solution x(t) for all t.[6] Now consider the corresponding stochastic differential equation:

$$dX_t = f(X_t)dt + g(X_t)dW_t$$

Here, W(t) is standard Brownian motion, and g(x(t)) represents the diffusion terms. Here we have the following theorem. If (Lipschitz Condition), $|f(x) - f(y)| + |g(x) - g(y)| \leq C|x-y|$ and (Linear Growth Condition), there exists a C>0 such that $|f(x)|^2 + |g(x)|^2 \leq C(1+|x|)^2\ \forall x$ then there exists a strong unique solution X(t) to the SDE.

In terms of proving this, there is a form of Picard Iteration for ODEs and for SDEs (which is simply an extension). For ODEs, we define a sequence as follows:

$$x_{n+1}(t) = x + \int_0^t f(x_n(s))ds$$

This converges under the Lipschitz condition. In the case of SDEs, we extend the above to SDEs:

$$x_{n+1}(t) = x + \int_0^t f(x_n(s))ds + \int_0^t g(X_n(s))dW_s$$

The deterministic part follows from Picard iteration for ODEs. The stochastic integral exists due to Ito's Isometry. In the case of proving convergence, we use Banach's fixed-point theorem in the space of stochastic processes.[7]

$$\mathrm{E}\sup_{0\leq t\leq T}|X_{n+1}(t) - X_n(t)|^2 \to 0$$

This guarantees that the stochastic sequence converges, proving existence and uniqueness with the difference in iterated guesses being zero.[8] Note, for this paper, the capital letter "E" will denote taking the expectation or mean of some process.

Next, let us consider stability and convergence for SDEs. Consider a deterministic stability case with Lyapunov's method for ODEs. For the ODE:

$dx(t)/dt = f(x(t))$

$$\frac{dV}{dt} \leq -cV(x)$$

for some c>0, then x(t) is globally stable and converges to an equilibrium.

Now consider stochastic stability (Ito's Lemma for SDEs). Suppose we have the SDE:

$$dX_t = f(X_t)dt + g(X_t)dW_t$$

Using Ito's Lemma, the stochastic analog of Lyapunov's condition is:

$$\mathcal{L}V(x) = f(x)V'(x) + 1/2\, g^2(x)V''(x)$$

If we can show that $\mathcal{L}V(x) \leq -cV(x)$ then X(t) is stochastically stable. A bit of background is needed here. In the context of SDEs, the operator $\mathcal{L}$ is known as the infinitesimal generator of stochastic processes. It is the key idea in stochastic analysis and helps in deriving stability conditions and solving PDEs associated with SDEs.[9] Two equations in this regard would be the well-known Fokker-Planck equation and the Hamilton-Jacobi-Bellman (HJB) equation, through which the stochastic differential equations can be re-formed into a solvable partial differential equation.[10]

Gronwall's inequality for SDEs is well known, as it is for ODEs. Given:

$$y(t) \leq C + \int_0^t ky(s)ds$$

then $y(t) \leq Ce^{kt}$

The stochastic version is extremely similar. Suppose we have the SDE

$$dX_t = f(X_t)dt + g(X_t)dW_t$$

We apply Ito's Lemma to the process Y_t = |X_t|^2, which gives $dY_t = CY_t dt + g^2(X_t)dt$

This gives

$$Y_t \leq C + \int_0^t kY_s ds$$

So Gronwall's inequality in this case is
$\mathrm{E}[X_t]^2 \leq Ce^{kt}$

This is crucial for bounding solutions for ODEs to ensure solutions do not explode.

Proofs

In order to look at the existence and uniqueness of solutions for this type of Hybrid model, where there are discrete and continuous systems of equations at play, we need to understand the mathematical frameworks underlying its components. Understand that the model combines the following aspects: SDEs from the Heston model volatility, GARCH dynamics, which are discrete-time processes modeling conditional variance and volatility clustering, Levy processes, for tailed behavior, and jump diffusions, which add discontinuities in the price dynamics. We will show well-posedness of our model in this section, show boundedness and uniqueness of solutions, and demonstrate the ways in which we arrive at our final integrated solution of our system of SDEs.

1. Existence and Uniqueness of Solutions:

Let us analyze the variance process for existence and uniqueness, given the framework laid out in Section 4. The stochastic variance process is governed by $dv_t = \mathcal{K}(\theta - v_t)dt + \sigma\sqrt{v_t}Z_t$

The drift term $\mathcal{K}(\theta - v_t)$ and diffusion term $\sigma\sqrt{v_t}$ satisfy the following conditions:

• Local Lipschitz Continuity: The square root volatility function is locally Lipschitz for v(t)>0. This ensures that small changes in v(t) result in small changes in dv(t).

• Linear Growth Condition: The terms $\mathcal{K}(\theta - v_t)$ and $\sigma\sqrt{v_t}$ grow linearly in v(t), satisfying:

$||\mathcal{K}(\theta - v_t)|| + ||\sigma\sqrt{v_t}|| \leq C(1+||v_t||)$ for some constant C>0.

DOI: 10.36838/v8i3.85

2. Stock Price Process:

The stock price process includes a jump component:

$$dS_t = S_t(\mu d_t + \sqrt{v_t} dW_t + J_t dN_t)$$

The terms in this equation satisfy:

• **Local Lipschitz Continuity:** The three dynamic functions (jump, stock price, and volatility) are Lipschitz in their respective domains, ensuring stability in the evolution of S(t).

• **Linear Growth Condition:** The drift, diffusion, and jump terms grow linearly in, satisfying:

$$|| \mu S_t || + || \sqrt{v_t} S_t || + || J_t S_t || \leq C(1 + || S_t)$$

In hindsight, the Lipschitz condition implies that no matter what sequence of events is occurring, and despite the position of the current real-world price, the stock price dynamics cannot change unnaturally. Without such a condition, the model becomes unstable and can predict changes in the price that are nowhere near realistic.

3. Jump Conditions:

The jump component J(t)dN(t) is modeled as a compound Poisson process. The following conditions ensure the existence and uniqueness of solutions:

• The Poisson process N(t) has finite intensity λ over any finite time interval.

• The jump sizes J(t) are modeled to have finite variance, i.e., $E[J_t^2] < \infty$.

• The jump coefficient J(t) is locally Lipschitz, ensuring the stability of the jump term.

4. The Feller Condition, Simulation Parameters, and GARCH Constraints:

To ensure volatility remains strictly positive, ensuring stability for our model and avoiding negative volatility predictions, we require the Feller condition to be satisfied.[11]

$2\mathcal{K}\theta \geq \sigma^2$

This condition above guarantees the model avoids any undefined behavior at [v(t)]^0.5. The boundary classification results for the CIR process are already conventionally discussed in literature, and the Feller condition is sufficient to uphold positive volatility.[12] The uniqueness aspect of SDEs is usually seen with Lipschitz continuity as well as linear growth conditions, and thus validates the use of the Feller condition.

For the GARCH dynamics, which has a discrete time variance, we need non-negativity through the following parameters: $w > 0, \alpha, \beta \geq 0, \alpha + \beta < 1$. If $\alpha + \beta < 1$ then the process is stationary and ergodic, which guarantees a well-defined sequence.

In this paper, the simulation will mainly be based on Monte Carlo simulation methods. The parameters for the simulation of our stochastic dynamics will run 100-500 stochastic paths, then take the median of those paths. We avoid the mean as generally done in Monte Carlo simulations for resistance to outliers and to help the simulation run time. Additionally, the parameters and variables are the same as mentioned in section 3. We constrain volatility from Heston's equations to be strictly positive, and we constrain the GARCH parameters as discussed above.

4.1. Data Collection and Preprocessing::

We download historical stock price data for the desired period. For example, we use Yahoo Finance or other financial data sources to retrieve the adjusted closing prices. We then calculate logarithmic daily returns as r(t) = ln (S_t / S_{t-1}), where S_t is the adjusted closing price at time t. We optimize the model parameters with the Mean Squared Error (MSE) objective function. It is defined as

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{S}_i - S_i\right)^2.$$

We formulate the stock price dynamics under the model

$$S_{t+1} = S_t \exp((\mu - 0.5\, v_t)\Delta t + \sqrt{v_t}\Delta t Z_t + \text{Jumps}),$$

where we have variance at time t, and Z_t is Brownian Motion. "Jumps" are indeed Poisson-driven.

5.Picard Iteration for Existence and Uniqueness:

In our proof of existence, we use the Picard Iteration Scheme to construct the solution to the stochastic hybrid system. This method is well-suited due to the following reasons:

• It ensures the existence of a unique fixed-point solution under appropriate contraction conditions.

• It aligns naturally with the structure of our SDE.

• It provides a constructive approach to approximating the solution, which is beneficial for both theoretical analysis and numerical implementation.

We consider a stochastic hybrid system of the form:

$$dS_t = f(S_t, v_t, N_t, J_t)dt + g(S_t, v_t, N_t, J_t)dW_t + h(S_t, v_t, N_t, J_t)dN_t$$

where:

• f, g, and h are measurable functions ensuring well-posedness.

• W(t) is a Wiener process.

• N(t) is a Poisson process representing jumps.

• J(t) are i.i.d. jump sizes with finite variance.

Given an initial condition S(0), we seek a solution satisfying:

$$S_t = S_0 + \int_0^t f(S_u, v_u, N_u, J_u)du + \int_0^t g(S_u, v_u, N_u, J_{uj})dW_u + \int_0^t h(S_u, v_u, N_u, J_u)dN_u$$

The Picard Iteration Scheme constructs an approximate sequence $\{S_t^n\}$ recursively as follows:

$$S_t^{n+1} = S_0 + \int_0^t f(S_u^n, v_u, N_u, J_u)du + \int_0^t g(S_u^m, v_u, N_u, J_u)dW_u + \int_0^t h(S_u, v_u, N_u, J_u)dN_u$$

where $S_t^0 = S_0$ is the initial guess.

To justify why such an iteration is allowed, we will showcase a contraction mapping property.

For a Picard iteration to converge, we need to show that the mapping:

$$T(S)(t) = S_0 + \int_0^t f(S_u^n, v_u, N_u, J_u)du + \int_0^t g(S_u^m, v_u, N_u, J_u)dW_u + \int_0^t h(S_u, v_u, N_u, J_u)dN_u$$

is a contraction in the normed function space $L^2(\Omega; C([0,T];\mathrm{R}))$ Consider two iterates S_t^n and S_t^{n+1} Taking the difference between two iterates and applying Lipschitz conditions gives us the following: $\mathrm{E}\sup_{t\in[0,T]}|S_t^{n+1} - S_t^n|^2 \leq CT\mathrm{E}\sup_{t\in[0,T]}|S_t^n - S_t^{n-1}|^2$. We then force T to be sufficiently small to achieve a contraction as follows: $\mathrm{E}\sup_{t\in[0,T]}|S_t^{n+1} - S_t^n|^2 \leq \theta\mathrm{E}\sup_{t\in[0,T]}|S_t^n - S_t^{n-1}|^2$. For Some 0< θ< 1 And finally, as it is a contraction, by Banach's Fixed-Point Theorem, (S(t))^n converges to a unique solution. Now, to apply such a concept fully to our model, we show that the solution of our system of SDEs,

which are denoted as X(t) = (S(t), v(t)) with an initial guess of the zeroth term of the iterated sequence of X(t) for all t on the interval [0, T], converges by estimating the difference between consecutive approximations, and showing that such differences tend towards 0. Above was the framework that we now apply to our H.G.L. model. We define the n-th approximation X_t^{n+1} to be defined recursively as:

$X_t^{n+1} = x_0 + \int_0^t b(s, X_s^{(n)})ds + \int_0^t \sigma(s, X_s^{(n)})dW_s$. We want to show that the difference between consecutive approximations

$$\| X_t^{(n+1)} - X_t^{(n)} \| \leq \int_0^t \| b(s, X_s^{(n)}) - b(s, X_s^{(n-1)}) \| ds + \int_0^t \| \sigma(s, X_s^{(n)}) - \sigma(s, X_s^{(n-1)}) \| dW_s$$

goes to zero. Using Lipschitz conditions, we find an inequality that follows from Itô Isometry:

$\| X_t^{(n+1)} - X_t^{(n)} \| < C \int_0^t \| X_s^{(n)} - X_s^{(n-1)} \| ds$ in which $||X_t^{(n+1)} - X_t^{(n)}|| \to 0$ as $n \to \infty$ uniformly in $t \in [0, T]$ by Gronwall's Inequality.

What this means is that if we take any two guesses of predicted stock prices at the n-th time in the future, the difference between the guesses would be zero, meaning those guesses are forced to be the same.

Thus, there is a unique solution, or no two guesses that differ as the limit goes to infinity in terms of n. And so, we further justified the existence of a unique solution to our stochastic dynamics.

6. Justification for L^2 Normed Space

L^2 provides a reasonable space for our model to work in. Firstly, this normed space has Hilbert Space properties, as L^2 is a Hilbert space, meaning that it has inner products that allow for powerful tools like orthogonality, projections, and energy estimates. We also have Ito Isometry: Stochastic processes, especially those involving Brownian motion, are naturally well behaved in L^2 due to Ito Isometry, which ensures that expectations of squared integrals behave nicely. Well-posedness and many existence and uniqueness results in SDEs and PDEs rely on energy methods in L^2. Below is a demonstration of how other normed spaces fail. Let us look at integral control for the normed space L^(infinity). Such a norm is given by

$$\| f \|_{L^\infty} = \sup_x |f(x)|$$

A key issue is that the integral of f(x) can diverge, making standard norm-based convergence arguments in L^2 inapplicable.

Counterexample: Consider the sequence of functions g_n of x where it is 1 for x on [0,1] and 1/n for x in (1, n]. In this normed space, all of these functions satisfy $||g_n||_{L^\infty} = 1$. Thus, they do not vanish in the norm as n goes to infinity, which prevents the use of limit arguments that are valid in L^2.

7. Moment Matching for the Hybrid Nature of the Model:

In our system of SDEs, we indeed have a discrete and continuous interplay between the SDEs. The GARCH process and the Levy Jumps are discrete events, which are in place to capture market irregularities that happen in discrete time events. The Heston volatility dynamics are continuous. To ensure stability and the model's accuracy, we need to ensure that the hybrid components match with each other in time. The first consistency condition requires that the moments of discrete-time variance align with the moments of the continuous-time variance process in our system of SDEs. For the Discrete-Time Variance, which is from the GARCH equation, we have: $v_{t+1} = w + \alpha\epsilon_t^2 + \beta v_t$, where $\epsilon_t^2 = \sqrt{v_t} Z_t$ and Z(t) follows a standard normal distribution. We compute the expectation of the GARCH volatility defined by $E[v_{t+1}] = w + \alpha E[\epsilon_t^2] + \beta E[v_t]$. Since $E[\epsilon_t^2] = E[v_t Z_t^2] = v_t$ we have $E[v_{t+1}] = w + \alpha E[v_t] + \beta E[v_t]$. It can be shown that the expectation of the squared GARCH volatility process has a summation in the form of a geometric series, particularly, $\omega \sum_{i=1}^{\infty} (\alpha + \beta)^i$ which converges to $\frac{\omega}{1-(\alpha+\beta)}$ for $\alpha+\beta < 1$ due to the convergence rule for a geometric series. It is why we also set the condition $\alpha+\beta < 1$ for this process. For the Heston variance process, the Fokker-Planck equation implies that the expected variance satisfies: $\frac{dE[v_t]}{dt} = \kappa(\theta - E[v_t])$. Of course, the Feller condition stated earlier guarantees that this volatility process remains strictly positive. To see this visually, the following graph showcases our model's volatility remaining strictly positive over a time interval.

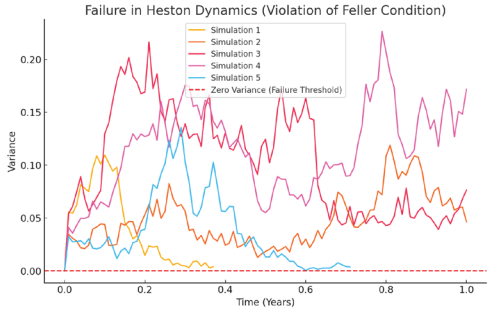


Figure 2: Variance over time of the hybrid model with a threshold for when volatility does not remain stable. The figure showcases 5 different probable expectations of simulations that are above the 0-volatility line, indicating that volatility is strictly positive and the HGL does not blow up.

Figure 2 demonstrates that our volatility dynamics of the Heston aspect of the model will strictly remain positive. Our model is thus steady and will not predict negative volatility, which is both unintuitive and causes model breakdowns. The Feller condition is then satisfied, and we are in a steady state of the model.

In a steady state, $E[v_t] = \theta$. And therefore, for our consistency condition to be met, we require $\theta = \frac{\omega}{1-(\alpha+\beta)}$.

As for the conditions for the jump term J(t), we simply require that the expectation of 1+J(t) is finite and greater than 0 almost surely.

■ Result and Discussion

We now take our three SDEs in the system of stochastic differential equations for the Heston-GARCH-Levy, and arrive at an equivalent integrated form of the model:

$$S_t = S_0 \exp\left(rt - \frac{1}{2}\int_0^t v_s ds + \int_0^t \sqrt{v_s}\, dW_s\right) \prod_{i=1}^{N_t} (1 + J_{t_i})$$

Here is how the product term arrived in this final form: When a jump occurs in the process S(t) due to the Poisson process, the stock price experiences a discontinuous change. The cumulative effect of jumps on the stock price over the interval [0, T] is described by $\int_0^T J_t dN_t = \sum_{i=1}^{N_T} J_{t_i}$. With this established, and using Monte Carlo simulations, the HGL model was tested against AAPL prices, and the error was analyzed as

DOI: 10.36838/v8i3.85

well. Following the conventional work on Heston's model, the variables are defined as follows: the drift function will include" r," which is the risk-free interest rate that is derived using Ito's Formula. We have two separate volatility integrals, one following an ordinary differential, and the other being with respect to Brownian motion. These are, of course, governed by both GARCH and Heston dynamics.

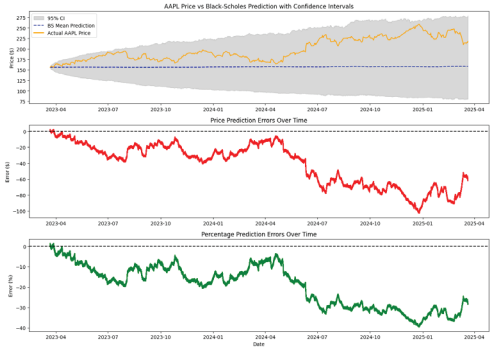


Figure 3: Black Scholes model was tested against AAPL stock prices, and percentage and numerical error charts. A maximum percentage error of 40 percent was shown in this sample simulation.

Figure 3 demonstrates an error analysis simulation of the Black Scholes model when tested against AAPL stock prices. In particular, the error percentages and specific price residuals are displayed over time. The error maximizes at around 40 percent, and the simulation itself has no solutions. The simulation utilized 99 percent confidence intervals, where we are 99 percent confident that the true expected simulation lies within such a range determined by the shaded region.

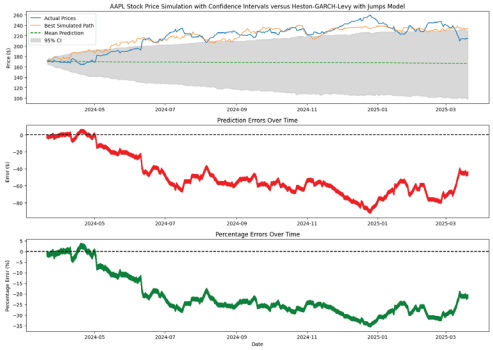


Figure 4: HGL model tested against AAPL stock prices, with percentage and numerical error charts. The maximum percentage error of the HGL occurs at 35%, and several solutions of intersections between simulated and actual prices are visually displayed above. A 99% confidence interval was used to capture the true expected simulation of all probable outcomes.

Figure 4 is an error analysis simulation of our own model, the HGL. In this case, we have a maximum error at around 35 percent, and we once again use 99 percent confidence intervals to capture the true expected simulation of our stock price dynamics. Note, I say expected as there are many possibilities for predictions. The one we choose will be the average or expected simulation of our process in the HGL model. When compared side-by-side, the HGL model has a significant improvement in approximations compared to the Black-Scholes model. The Black-Scholes model has a maximum error of 40% deviations, while the HGL model has a 35% maximum error. The HGL model is also better at tracking price behaviors, as tendencies to move in certain directions are mimicked by the HGL. To further see this, look at the figure below of a broader simulation of the HGL against AAPL prices.

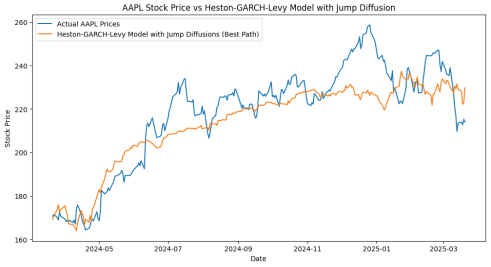


Figure 5: HGL sample simulation against AAPL prices with enhanced jump diffusions. The HGL tends to follow the growth trend of AAPL prices in early 2024, and early in the simulation, the drops and spikes and often mimicked. In real-world trading, options trading is significantly enhanced with our model's improvements on previous versions of stochastic processes.

Large Deviations in the Poisson Process

Furthermore, in discussing the resulting simulations, we must note that the large deviations matter in predicting abnormal stock price dips and spikes.[13] In section 5.2, it has been noted that the Lipschitz condition prevents the model's explosive growth (of course, with the linear growth condition allowing the existence of solutions). However, immense changes in the stock price have occurred in the past, and thus, we cannot rule out the ability to predict such instances.

The number of jumps in the stock price, defined by N(t), follows a Poisson process with:

$$P(N_t = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

Using Cramér's theorem, we estimate $P(N_t > \lambda_t + \delta_t)$. We define the scaled process as $\frac{N_t}{t} = \lambda + \frac{1}{t}\sum_{i=1}^{N_t} J_i$ and we use Sanov's theorem to give $P(N_t > \lambda_t + \delta_t) \approx e^{-tI(\delta)}$, where the rate function is: $I(\delta) = \sup_{\theta}[\theta\delta - \log \mathrm{E} e^{\theta N_t}]$

For Poisson-distributed N(t), $\Lambda(\theta) = \lambda (e^{\theta} - 1)$, yielding:

$$I(\delta) = \delta \log\left(\frac{\delta}{\lambda}\right) - \delta + \lambda$$

What this states, intuitively, is that jump clustering is exponentially rare but still possible. Thus, we do not rule out a possible extreme market crash or market spike. Additionally, if J(t) follows a heavy-tailed Lévy distribution, the rate function decays more slowly, increasing the likelihood of extreme price movements.

As for the GARCH volatility process, to show that volatility can also have such extreme predictions, we can showcase the large deviations proof by rewriting the GARCH volatility dynamics as a discrete sum: $v_t = \sum_{k=0}^{\infty} \beta^k(\omega + \alpha\epsilon_{t-k}^2)$. The empirical mean is then defined as $\frac{1}{T}\sum_{t=1}^{T} v_t$. The Gärtner-Ellis theorem gives us the large deviation rate function as $I(x) = \sup_{\theta>0}[\theta x - \Lambda(\theta)]$ where $\Lambda(\theta)$ is the moment-generating function for this volatility process. For large x-values, $P(v_t > x) \approx e^{-TI(x)}$. For heavy-tailed innovations (e.g., Student-t distributions), I(x) decays more slowly, increasing the probability of extreme volatility. Thus, extreme volatility states can persist and thus make risk assessment crucially involved in trading decisions.

To best understand this from a non-mathematical standpoint, consider that while our model has a low probability of

enacting high motion jump dynamics, there still "is" a probability that is on the tail end of the Poisson. At this level, the distribution values that we obtained are incredibly high, and will thus affect the price prediction significantly away from the drift function with the risk-free interest rate.

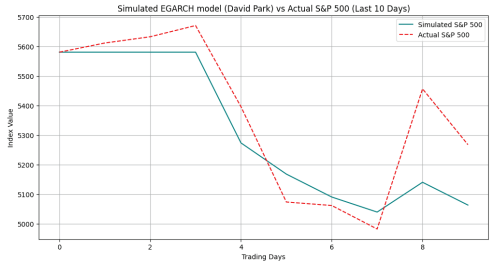


Figure 6: HGL tested against the S&P 500 for an extreme market crash under Trump tariffs. HGL simulation demonstrates mimicking patterns with the AAPL stock prices.

Figure 6 showcases an example of how our model applies to real-life scenarios. During the recent Trump administration's tariffs, the stock market experienced a period of severe decline. The HGL predicted the exact moment at which this fall begins, and thus, a put option would safely put the investor in a winning scenario despite extreme market crashes. Notice, the predictions within this time frame also match the characteristic behaviors. When the prices are falling, so does the HGL prediction, and when they shift up, so does the HGL. These characteristics are particularly powerful in options trading, where there are fundamental strategies for buying puts or calls that are easy to apply with this model's behaviors.

And so, the model is successful in incorporating extreme market events as part of its predictive capabilities and is also stable. The model's simulations are both characteristically and statistically more accurate in price predictions than previous versions and thus can be utilized effectively in options and algorithmic trading.

Conclusion

The analysis presented in this paper confirms that the stochastic differential equations (SDEs) of the Heston-GARCH-Lévy model satisfy the local Lipschitz continuity and linear growth conditions for the drift, diffusion, and jump terms. As a result, the existence and uniqueness of a strong solution to the model are established based on standard stochastic calculus theorems.[14] Beyond numerical accuracy, this study provides a theoretical foundation for the use of specific structures in financial modeling. In particular, we justified the Cox-Ingersoll-Ross (CIR) process for variance modeling, which ensures non-negative volatility paths, making it more suitable than standard Brownian motion for financial applications, as seen in ref [8]. Furthermore, we established the necessity of working within the L^2 normed space, providing counterexamples that illustrate why alternative normed spaces fail to maintain the desired mathematical properties for our SDE framework. These justifications strengthen the mathematical rigor behind the model's construction and ensure its reliability for practical applications. The Heston-GARCH-Lévy model offers a more robust and flexible framework for asset price modeling by addressing key limitations of traditional models like Heston, SABR, Rough Volatility, and Variance Gamma. Unlike these models, our approach effectively captures volatility clustering, stochastic mean reversion, and discontinuous price jumps, providing a more comprehensive reflection of financial market behaviors. However, despite its accuracy, the computational cost remains a challenge. Future work will focus on calibrating the model using historical market data and implementing it in real time for derivative pricing applications. Additionally, since Monte Carlo simulations remain computationally expensive, efforts will be made to improve simulation efficiency through optimized numerical methods, reduced variance techniques, and potentially leveraging high-performance computing for faster simulations without compromising accuracy. By refining both the theoretical and computational aspects of the model, we will aim to enhance its applicability in quantitative finance and risk management. Additionally, wavelet simulations will be attempted to cover extreme noise in different markets. Future work will also incorporate the extension of our model toward other markets, such as cryptocurrencies, with more sets of SDEs. Lastly, new rigorous proofs of stability will be studied for hybrid stochastic systems for markets outside of tech-based stocks.

References

1. Black, F., & Scholes, M. The pricing of options and corporate liabilities. *Journal of Political Economy*. 1973;81(3):637–654.
2. Heston, S. L. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud*. 1993;6(2):327–343.
3. Cox, J. C., Ingersoll, J. E., Jr., Ross, S. A. A theory of the term structure of interest rates. *Econometrica*. 1985;53:385–407.
4. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics.* 1986;31(3):307–327.
5. Skorokhod, A. V. *Asymptotic methods in the theory of stochastic differential equations*. Mineola, NY: Dover Publications; 2009.
6. Øksendal, B. *Stochastic differential equations: An introduction with applications*. 6th ed. Berlin: Springer; 2003.
7. Khasminskii, R. *Stochastic stability of differential equations*. 2nd ed. Berlin: Springer; 2011. (Stochastic Modelling and Applied Probability; Vol. 66).
8. Revuz, D., Yor, M. *Continuous martingales and Brownian motion*. 3rd ed. Berlin: Springer; 1999. (Grundlehren der mathematischen Wissenschaften; Vol. 293).
9. Krylov, N. V. An analytic approach to SPDEs. *Stoch. Partial Differ. Equ.: Six Perspect*. 1999;64:185–242.
10. Karatzas, I., Shreve, S. E. *Brownian motion and stochastic calculus*. 2nd ed. New York: Springer; 1991. (Graduate Texts in Mathematics; Vol. 113).
11. Feller, W. Two singular diffusion problems. *Ann. Math*. 1951;54:173–182.
12. Jaeckel, A. G., Yor, M. Boundary classification for CIR processes. *Bernoulli*. 2003;9(6):957–983.
13. Gyöngy, I., Krylov, N. V. Stochastic differential equations in Lp spaces and applications. *Stoch. Process. Their Appl*. 1994;50(1):1–21.
14. Lee, S. S. Jumps and information flow in financial markets. *Working Paper*. 2024.

DOI: 10.36838/v8i3.85

Author

David Park is a high school senior from Campbell Hall, a private school in Studio City, CA. He is a current student of the class of 2026, and he is entering the class of 2030 for his undergraduate career. He is enthusiastic about researching the depths of stochastic processes and how randomness and stochastic differential equations can model several different aspects of financial derivatives and securities. Additionally, he is passionate about the fields of Measure Theory and probability, which make up the foundation of stochastic research. His research has been acknowledged by The David J. Park Foundation for Innovation & Technology and has been granted funding to pursue the hybrid model even further in the world of computer science. He aspires to pursue the fields of financial engineering and applied mathematics moving forward, and is an eager researcher with multiple recognitions in the field of stochastic processes and their applications.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Inequality and the Gig Economy in Baltimore

Brandon Wu

Gilman School, 5407 Roland Ave, Baltimore, MD, 21210, USA; brandon2048wu@gmail.com

ABSTRACT: This research paper examines the intersection of inequality and the gig economy in Baltimore, a city shaped by segregation, deindustrialization, and systemic barriers to mobility. Through interviews with Uber drivers and secondary data analysis, the study explores the dual nature of gig work: providing quick income, scheduling autonomy, and entry points to entrepreneurship while also perpetuating precarity through low pay, earnings volatility, and limited protections. Baltimore's stark disparities offer a critical lens for understanding these dynamics. Findings indicate most drivers use gig work as a transitional tool rather than a permanent career, showcasing the need for policies that preserve flexibility while also adding security through upward mobility programs, universal protections, and digital equity initiatives.

KEYWORDS: Behavioral and Social Sciences, Sociology and Social Psychology, Gig Economy, Baltimore, Uber.

Introduction

Baltimore has long struggled with persistent inequality, which continues to limit economic mobility and create sharp racial and economic disparities. These entrenched divides have shaped livelihoods for generations.[1-3] These conditions set the stage for examining how new forms of work interact with long-standing disparities.

At the same time, the gig economy has emerged as a new force in the labor market, defined by short-term, flexible work facilitated by platforms such as Uber and DoorDash.[4-6] Its rapid growth is particularly relevant for urban areas like Baltimore, where traditional labor markets have failed to provide sufficient opportunities.[7,8] By offering income, autonomy, and entrepreneurial entry points, gig work appears to offer a new pathway for addressing entrenched inequality.[9]

This paper examines the intersection of inequality and the gig economy in Baltimore. Drawing on interviews with Uber drivers and secondary data, it analyzes the opportunities and risks of platform work. The paper highlights drivers' use of gig work as a temporary bridge rather than a permanent career, and it proposes policy measures to preserve flexibility while strengthening protections. The next sections provide context on Baltimore's structural inequalities, review the literature on gig work, outline the research methods, present the findings, and discuss their policy implications.

Context

Baltimore's inequality is rooted in historical segregation, discriminatory policies, deindustrialization, and cycles of poverty. Practices such as redlining in the 1930s created lasting residential and resource divides. Today, these structural barriers continue to shape the city's economic and racial landscape.

Black residents, who make up 63% of Baltimore's population, fare worse than the national African American average in nearly every socioeconomic category, while White residents, at 28% of the population, perform better than national averages.[3] Unemployment among Black residents is three times higher than among Whites, and the percentage of Whites with a bachelor's degree is 3.2 times higher. These figures underscore the sharp disparities that reflect limited economic mobility.

This background makes Baltimore a critical case for examining how new forms of work intersect with inequality. The city's combination of entrenched racial disparities and economic precarity provides a revealing lens for analyzing the opportunities and risks of gig work. These broader dynamics are the focus of the following literature review.

Literature Review

The gig economy, represented by platform businesses such as Uber and DoorDash, reflects the contradictions of inclusion and exclusion central to Baltimore's socioeconomic landscape. Scholars have long debated whether gig work mitigates or worsens inequality, and this study adds to that conversation. Platform-based work offers quick income, autonomy, and entrepreneurial entry points, making it a low-barrier pathway to employment, especially in places where systemic barriers block access to traditional jobs.[7] Research on Uber's staggered entry into U.S. metropolitan areas found that the platform increased labor participation among groups excluded from competitive markets due to a lack of skills.[10] Uber's arrival also contributed to local business activity, with ~5% growth in new business registrations and similar increases in small business lending.[11]

Gig platforms also stimulate entrepreneurship. As Wolla[7] notes, flexible scheduling and supplemental income allow individuals to pursue ventures with reduced financial risk. Nationally, gig platforms are associated with a ~7% rise in internet searches for entrepreneurship-related terms, especially in constrained regions.[11] In Baltimore, programs such as Innovation Works and the Techstars Equitech Accelerator have sought to leverage this potential by supporting underserved entrepreneurs and fostering an inclusive innovation economy.

Another strand of research highlights skill development. The 2023 Baltimore Digital Equity Report shows that digital literacy directly influences economic outcomes. By requiring

DOI: 10.36838/v8i3.93

drivers to use apps and GPS, Uber can help build competencies in technology, time management, and customer service. For residents affected by digital redlining, gig work can be a gateway to bridging divides.[12,13]

At the same time, scholars warn that gig work may exacerbate inequality without safeguards. Workers face unstable earnings, limited protections, and risks tied to independent contractor classification.[6,8,14] Such structures deprive workers of healthcare, paid leave, and retirement security, leaving them exposed to financial shocks. Critics argue that socioeconomic stratification pushes low-income populations into precarious gig jobs, perpetuating rather than alleviating inequality.[15]

Proposed reforms aim to balance flexibility with protection. These include creating a third category of "dependent contractors," granting partial benefits and bargaining rights,[16] or designing universal benefit schemes decoupled from traditional employment, ensuring baseline protections such as health coverage and retirement contributions.[5] These reforms seek to preserve the accessibility of gig work while mitigating its vulnerabilities. To examine these issues in practice, this study focuses on Uber drivers in Baltimore.

Methods

This study relies on semi-structured interviews with ten Uber drivers in Baltimore to capture how gig work intersects with inequality. Baltimore was chosen as a case study due to its suitability for examining how platform work interacts with entrenched inequality. Participants were recruited through personal outreach and informal referrals. Each interview lasted about ten to twenty minutes and included questions on drivers' backgrounds, income reliance, motivations, benefits, challenges, and future goals. Inequality was measured through drivers' accounts of economic precarity, including reliance on Uber to meet basic expenses, lack of stable employment alternatives, variability in weekly earnings, and limited access to career advancement. Interviews were transcribed with consent and systematically reviewed, with drivers organized into four categories using a two-by-two framework distinguishing between primary versus supplementary income reliance and temporary versus long-term orientation.

Secondary sources, including census data, policy reports, and prior studies on gig work and digital equity, were used to contextualize the interview findings. These materials provided a backdrop against which drivers' experiences could be interpreted, helping to situate interview findings within wider patterns identified in prior research. The results below show how drivers themselves described these experiences.

Result and Discussion

Results:

Table 1: Key interview findings. This table presents responses from ten Uber drivers to questions covering their backgrounds, driving habits, motivations, challenges, and future goals. The answers show wide variation in how drivers use Uber, from full-time dependence to part-time flexibility, shaped by differing financial needs and career aspirations.

Questions	Driver 1	Driver 2	Driver 3	Driver 4	Driver 5	Driver 6	Driver 7	Driver 8	Driver 9	Driver 10
Where do you live?	Baltimore	Rosedale, MD	Baltimore	Baltimore	Baltimore	Millersville, MD	Baltimore	NE Baltimore	Patterson Park, Baltimore	Virginia
How long have you been driving for Uber?	Two years	A year and three months	Five months	One year	One year	On and off for a few years	Not specified	Since 2016	13 months	Almost two years
How much do you drive?	Four to five hours a day	Depends on time available	Four to five hours a day	Seven to eight hours a day	Around 20 hours a week	8–12 hours, with breaks	2–3 hours on weekdays; 5–6 on weekends	About 35 hours/week	About 70 hours a week	Few hours after Whole Foods; more on day off
Do you have any other jobs?	Yes, has another job	Yes, works at Hertz	Yes	Previously worked in a hospital; now full-time Uber	No, Uber is only job	Yes, full-time IT engineer	Yes, works for the city	Retired from City; Uber to fill income gap before SS	No	Yes, works at Whole Foods
Do you think gig jobs like Uber are good for society?	Yes	Yes	Partly	Yes		Yes, builds skills and bridges careers	Yes, but pay has become unfair	Yes, especially where public transit is poor	Yes	Yes, helps those without cars
How has Uber positively impacted your life?	Provides extra money and offer scheduling flexibility	Provides financial freedom and flexibility	Positive, more money	Helps balance work and family; covers daily expenses		Funds business ventures; supports family	Financial and social benefits	Built significant savings	Flexible to do other stuff	Extra cash; enjoy exploring while driving
What do you dislike about Uber?	GPS issues; grateful not to have had bad customers	Disrespectful passengers	Some people don't care	Long driving hours affect health	3-4% Customers are hard	Some stressful passenger situations	Pay taxes on income	Drivers earn less despite higher rider charges	Rates could be better	Can't control ride destinations
Has Uber improved your professional skills?	No	Yes; learned digital platforms & future of apps	Yes		Yes	Yes, social and digital skills	Yes, both social and professional skills	No, previous job made him better at Uber	No	Yes, driving-wise
Do you feel Uber limits your chances of full-time work?	No; flexible scheduling makes it manageable	No				No; it's optional time use	Sometimes pays better than full-time jobs	No; did both when employed	This *is* my full-time job	No, it's flexible and optional
What are your long-term career plans?	Maybe another two years; has other plans	Start a business; entrepreneurial spirit		Wants to start a business	Possibly return to previous trade (electrician)	Continue engineering and grow e-commerce ventures	Wants to get into real estate	Wants to travel more in retirement	Building rental property portfolio	Not specified

Table 1 summarizes interviews with ten Uber drivers, showing the diverse ways workers engage with the platform. Some treated it as a full-time job, while others drove only occasionally. Several combined Uber with other employment, while others relied entirely on it for income.

Most drivers saw Uber as a temporary solution rather than a long-term career. Only one planned to keep driving indefinitely. Many described Uber as "a means to an end" that provided short-term financial stability while they pursued other goals, such as starting a business, saving for retirement, or investing in real estate. For some, Uber filled a gap during career transitions or supported entrepreneurial side projects.

Economic inclusion was a key theme. Several drivers highlighted Uber's low entry barrier, noting it provided income when no other work was available. One recent immigrant said Uber helped him "meet daily expenses and spend time with [his] kids."

Flexibility was another consistent theme. Drivers emphasized the value of being able to choose their own hours, balancing jobs, school, or recovery from injury. As one put it, "Uber lets me do it whenever you want, for as long as you want."

Skill development also emerged. Some drivers reported building digital skills, especially comfort with apps and platforms. As one explained, "everything is done through an app these days."

Concerns about precarity were widespread. Drivers frequently mentioned declining pay, unpredictable income, and lack of fairness. "They don't pay as well as they used to," one driver explained, noting that fares had risen but driver earnings had not kept pace. Others described long hours, difficult passengers, or a lack of control over ride destinations.

Taken together, the interviews show Uber's dual role: as a flexible and accessible source of income, and at times a way to build skills or bridge to the future, but also as work that many felt paid less than it should, with the platform capturing more value over time. To better capture the different ways

Uber shaped drivers' lives, the study participants were grouped into categories using a two-by-two framework.

Driver Topology:

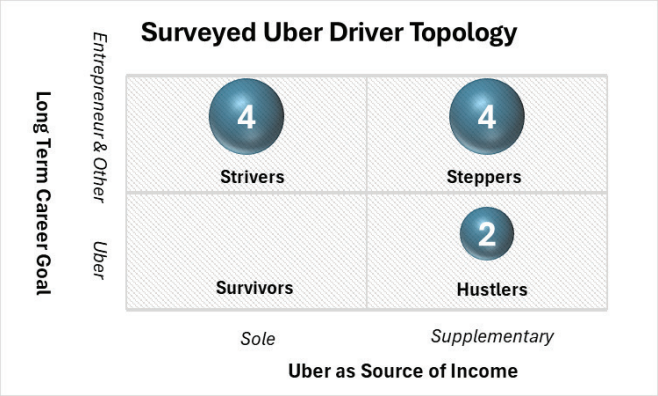


Figure 1: A 2x2 framework of job engagement of surveyed Uber drivers. The categorization illustrates that most surveyed drivers view Uber as a stepping stone rather than a destination, with the majority aiming for long-term goals outside the platform. This supports qualitative feedback from drivers who see Uber as a temporary means to fund entrepreneurial or career transitions.

Drivers were grouped into four categories using a two-by-two framework (Figure 1) that distinguishes between whether Uber is a primary or supplementary income source and whether it is seen as long-term or temporary.

Strivers (Drivers 4, 5, 8, 9): rely solely on Uber but view it as temporary. They are in career transition, waiting for other income sources, or preparing to launch new ventures.

Steppers (Drivers 1, 2, 6, 7): have full-time jobs and use Uber as supplementary income while working toward other goals. For them, Uber provides flexibility to build savings or invest in side projects.

Side Hustlers (Drivers 3, 10): have other jobs and drive casually for extra money, without strong long-term career ambitions.

Survivors: drivers who rely exclusively on Uber and plan to stay long-term. None appeared in this sample, a notable finding suggesting that Uber is rarely seen as a permanent solution.

This typology highlights that most drivers use Uber strategically and temporarily, reinforcing that flexibility—not permanence—is its main value. These categories set up the discussion of broader implications.

Interpretations and Policy Implications:

Taken together, the findings reveal how gig work in Baltimore combines opportunities for inclusion and skill-building with tensions around compensation fairness shaped by the controlling power of the platforms. The low entry barrier allows individuals excluded from traditional labor markets to earn quickly. App-based work helps some drivers build digital literacy and entrepreneurial skills. These themes emerged directly from the driver interviews. At the same time, secondary sources point to broader vulnerabilities that go beyond what this small sample reported, particularly the precarity associated with fluctuating earnings and the absence of employment-based benefits. Situating the drivers' accounts within this wider literature underscores that while gig work can serve as a transitional tool, it often does so under fragile conditions.

The absence of "Survivors" in the driver typology underscores that gig work is rarely viewed as a long-term path and is instead used as a transitional tool. This strategic temporariness highlights both the promise and the limits of the gig economy: it provides flexibility and income in critical moments, but, as secondary sources remind us, it also operates under conditions of precarity and a lack of employment-based benefits. These combined insights point to the need for targeted policy measures, outlined below.

Upward Mobility.

For Strivers and Steppers, targeted support can help turn temporary earnings into long-term opportunities. This includes microloans or grants, digital upskilling, and mentorship networks that connect drivers to professionals and entrepreneurs. Bundling gig work with opportunity-building programs would enable motivated drivers to advance beyond short-term reliance.

Universal Protections.

All drivers, regardless of category, would benefit from portable benefits, minimum earnings guarantee, transparent pay structures, and stronger safety measures. Light-touch regulation should ensure protections without undermining platform viability.

Digital Equity.

Access to affordable internet, smartphones, and digital training is critical in Baltimore, where digital divides remain sharp. Public investment, potentially co-funded by platforms, could expand inclusion and ensure that drivers gain skills that extend beyond gig work. Together, these recommendations frame the broader implications of Baltimore's case, which the conclusion summarizes.

■ Conclusion

This study introduces a simple driver typology—Strivers, Steppers, and Side Hustlers, with no Survivors—that shows how gig workers in Baltimore use Uber as a transitional tool linked to long-term goals like entrepreneurship, savings, or retirement. By connecting these different forms of driver engagement to Baltimore's structural inequalities, including racial unemployment gaps and digital redlining, the analysis demonstrates that platform work is shaped as much by local context as by the platforms themselves. Future research should test this typology in other cities and with larger samples to see how inequality structures gig work differently. For policy, the findings suggest moving beyond one-size-fits-all reforms toward tiered approaches that foster upward mobility, strengthen universal protections, and expand digital equity. Taken together, Baltimore's case illustrates both the limits of gig work as a permanent solution and its potential to support inclusion when paired with context-sensitive support.

DOI: 10.36838/v8i3.93

Acknowledgments

I want to thank Professor Cresa Pugh for being the most incredible mentor throughout this project. From helping me shape the research question to walking me through literature reviews, survey design, and data analysis, she guided every step with care and patience. Her feedback pushed me to think deeper and revise better each time. I learned so much from her, and I truly couldn't have done this without her.

References

1. Nelson, R. K.; Winling, L.; Marciano, R.; Connolly, N.; Ayers, E.; et al. *Mapping Inequality: Redlining in New Deal America [Online]*; Digital Scholarship Lab, University of Richmond: Richmond, VA, 2016–present; https://dsl.richmond.edu/panorama/redlining/
2. Pietila, A. *Not in My Neighborhood: How Bigotry Shaped a Great American City*; Ivan R. Dee: Chicago, 2010. ISBN 978-1-56663-843-2.
3. Assante-Muhammed, D. *The Racial Wealth Divide in Baltimore*; Corporation for Enterprise Development: Washington, DC, 2017. https://civilrights.baltimorecity.gov/sites/default/files/Racial_Wealth_Divide_in_Baltimore_RWDI.pdf.
4. Katz, L. F.; Krueger, A. B. Understanding Trends in Alternative Work Arrangements in the United States. *RSF: The Russell Sage Foundation Journal of the Social Sciences* **2019**, *5* (5), 132–146. https://doi.org/10.7758/RSF.2019.5.5.07
5. De Stefano, V. The rise of the "just-in-time workforce": on-demand work, crowd work and labour protection in the "gig-economy." *SSRN Electron. J.* **2015**. https://doi.org/10.2139/ssrn.2682602.
6. Friedman, G. Workers without employers: shadow corporations and the rise of the gig economy. *Rev. Keynesian Econ*. **2014**, *2* (2), 171–188. https://doi.org/10.4337/roke.2014.02.03.
7. Wolla, S. A. *How does the gig economy support entrepreneurship?* Federal Reserve Bank of St. Louis, Page One Economics, July 26, 2024. http://www.stlouisfed.org/publications/page-one-economics/2024/04/01/how-does-the-gig-economy-support-entrepreneurship.
8. Berg, J. Income security in the on-demand economy: findings and policy lessons from a survey of crowdworkers. *Comp. Labor Law Policy J.* **2016**, *37* (3), 1-26. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2740940.
9. Vallas, S. P.; Schor, J. B. What Do Platforms Do? Understanding the Gig Economy. *Annu. Rev. Sociol.* **2020**, *46*, 273–294. https://doi.org/10.1146/annurev-soc-121919-054857
10. Li, Z.; Hong, Y.; Zhang, Z. The empowering and competition effects of the platform-based sharing economy on the supply and demand sides of the labor market. *J. of Manage. Inf. Syst*. **2021**, *38* (1), 140–165. https://doi.org/10.1080/07421222.2021.1870387.
11. Barrios, J.; Hochberg, Y.; Yi, H. *Launching with a parachute: the gig economy and new business formation*; National Bureau of Economic Research: Cambridge, MA, 2020. https://doi.org/10.3386/w27183.
12. Edmonds, J. D. *Digital redlining: the effects on urban gig workers*. DigitalC. Digitalc.org. https://digitalc.org/blog/the-adverse-effects-of-digital-redlining-on-gig-workers-in-the-urban-core.
13. Behera, B.; Gaur, M. Skill training for the success of the gig economy. *PNR J.* **2022**, *13* (5), 2835–2840. https://www.pnrjournal.com/index.php/home/article/view/4655/5206.
14. Davies, A. C. L. Wages and working time in the 'GIG economy'. *King's Law J.* **2020**, *31* (2), 250–259. https://doi.org/10.1080/09615768.2020.1790820.
15. Muntaner, C. Digital platforms, gig economy, precarious employment, and the invisible hand of social class. Int. J. Health Serv. **2018**, *48* (4), 597–600. https://doi.org/10.1177/0020731418801413.
16. Cherry, M. A.; Aloisi, A. "Dependent contractors" in the gig economy: a comparative approach. *SSRN Electron. J.* **2016**. https://doi.org/10.2139/ssrn.2847869.

Author

Brandon Wu is a student at Gilman School in Baltimore, Maryland, Class of 2027. He has co-founded startup initiatives, leads Gilman's DECA chapter, and serves on the Investment Committee. Outside of school, he enjoys basketball, philosophy, and exploring entrepreneurial and financial ventures.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

The Impact of Image Classification Using Convolutional Neural Network on Wildfire Detection

Shaunak Thamke, Elliot Blin, Kevin Diao

Bronx High School of Science, 75 W 205th St, Bronx, NY 10468; shaunakthamke@gmail.com, elliotblin2509@gmail.com, diaokevin2@gmail.com

ABSTRACT: Wildfires pose significant threats to human populations, ecosystems, and infrastructure worldwide, with early detection remaining a critical challenge. The purpose of this study was to investigate the use of a convolutional neural network (CNN) for wildfire detection through image classification, to optimize factors such as dataset size, number of epochs, and image resolution to enhance accuracy. A CNN model was trained on a wildfire dataset of 2,700 images, categorized into "fire" and "no-fire," utilizing key libraries such as TensorFlow, Pillow, and Sci-Kit Learn for preprocessing, training, and evaluation. The model was iteratively optimized by varying the number of photos, epochs, and image resolutions, with larger datasets and increased epochs significantly improving performance. Results showed that 100x100-pixel images provided the best balance between detail and computational efficiency. These findings highlight the potential of CNNs in wildfire detection, offering a faster and more reliable solution compared to traditional methods. The approach can save lives, protect property, and reduce environmental damage. Future research should focus on enhancing the model's robustness and exploring real-time applications for early wildfire detection using drones or satellites.

KEYWORDS: Earth and Environmental Sciences, Environmental Effects on Ecosystems, Convolutional Neural Networks, Image Classification, Wildfire Detection.

Introduction

In the current state of the world, wildfires are undoubtedly some of the biggest threats to the human population. Wildfires are unpredictable and uncontrolled, and usually start in natural areas, such as forests, grasslands, or prairies. The environment suffers quite a great deal from the impact of wildfire. According to *iii.org*, in 2020 alone, over 10 million acres were burned down due to wildfires. Furthermore, these wildfires leave numerous people without a roof over their heads.

Research on wildfires has been conducted in the past. Research about wildfires has examined the effects of inhaling smoke from wildfires on individuals' health on a more general scale. Earlier research has shown that inhaling such smoke can lead to direct toxicity, oxidative stress, inflammatory reactions, immune dysregulation, genotoxicity, mutations, and other adverse effects.[1] Similarly, other research has been conducted to find the impact of specific wildfire smoke on cardiovascular health. It was found that inhaled wildfire smoke, which contains particulate matter—the primary pollutant in wildfire smoke—can lead to numerous cardiovascular issues. These issues can primarily affect the elderly, pregnant women, young individuals, and the healthy, with symptoms including systemic inflammation and vascular activation.[2] These papers illustrate the destructiveness of wildfires on people. As stated before, the environment also gets damaged by wildfires. Research was done on the impact of wildfire on surface water quality. The study was specifically conducted to examine the differences between pre- and post-wildfire water. They found that within a year, the pH of the water changed, but it returned to normal after a year. The concentration of contaminants in the water also increased significantly after the wildfire. Total Suspended Solids also increased in waters with a flow of 10 m3s–1, while higher flow decreased the total suspended solids, possibly by dilution. This research demonstrates how wildfires can impact water quality, potentially affecting the environment and human health.[3] We can dramatically lower this number by effectively distinguishing real threats of wildfires from non-threatening fires.

Past studies have found the dangers of wildfires to people, their homes, and animals. This research included developing a machine learning model using neural networks that could accurately predict where wildfires could happen. More specifically, this machine learning model can recognize wildfires before they fully develop; then, at such a small scale of a fire, we can quickly put it out. The goal of the research was to see if the CNN model could detect wildfires. The hypothesis was that if certain factors for the CNN model are changed, can the model's accuracy in predicting wildfires from a picture increase? This research could benefit anybody living in wildfire-prone areas, animals in those areas, and nature.

Methods

Libraries needed:

For the project, the Sci-Kit Learn library was utilized. It employed various machine learning tools, including utilities for preprocessing data, splitting datasets, and evaluating models. Using Scikit-learn, data was loaded and prepared to train a CNN model that could make predictions. The model was evaluated for its predictive ability (accuracy). Thus, it was en-

DOI: 10.36838/v8i3.97

sured that data preparation, model training, and evaluation were done most efficiently.

Pillow, a library specializing in image-processing tasks, was also used. With Pillow, important image manipulation features, such as cropping, rotating, and applying filters to images, were possible. This was specifically helpful for the project as it involved many photos.

Specifically, Pillow was also used in the project to resample the picture. Resampling changes the image's size and, therefore, quality. This was helpful for the project, as the size of an image could correlate with the correctness of the model predictions.

TensorFlow was another library used that offered all the tools to define, train, and deploy models. It was specifically helpful in making the CNN model. TensorFlow has been helpful in training neural networks in the past. Training and testing of the neural network have mostly been done using this library.

These libraries were imported at the start of the Python code. Without these libraries, essential steps could not be taken, and the creation of the CNN model was not possible.

Dataset used:

"The Wildfire dataset" was used for the project because of its diversity, scale, and quality. The dataset spanned 2,700 aerial and ground-based images collected from various online sites and databases such as Flickr and Unsplash, capturing a wide range of forest types and geographical accuracy. The dataset is organized into clear 'fire' and 'no-fire' categories to simplify processing and model training. Additionally, the high-resolution images ensured that the detailed data was input into the model.

Figure 1: A picture of a real wildfire in the Wildfire Dataset named "fire" for modeling. Note. From The Wildfire Dataset, by Ismail El Madafri,[4] updated a year ago, Kaggle. https://www.kaggle.com/datasets/elmadafri/the-wildfire-dataset.

Figure 2: A picture of a forest without a wildfire from the Wildfire Dataset named "no-fire" for modeling. Note. From The Wildfire Dataset, by Ismail El Madafri[4], updated a year ago, Kaggle. https://www.kaggle.com/datasets/elmadafri/the-wildfire-dataset.

Importing Images:

Images from the wildfire dataset were stored in two files. The file containing the pictures of real wildfires was named "fire," and the one without was named "no-fire," as shown in (Figure 1) and (Figure 2), respectively; just as examples. These files were input into the code by specifying their paths and assigning a numerical label. The pictures from the "fire" folder were assigned the label 0, while the pictures from the "no-fire" folder were assigned the label 1. Labeling the two files allowed the model to map the input pictures to the correct output. This helped the model understand the difference between the two categories, allowing the model to predict whether a new image contained a wildfire. Photos were rescaled to see if they affected the model's accuracy.

Splitting the data into training and testing sets:

After importing the images, a very important step was to split the data into training and testing sets, as illustrated in Figure 3. Splitting data into training and testing sets is incredibly important for evaluating the model's performance. The training set is used to teach the model to identify patterns that enable classification. The testing set acts as the unseen data to see the model's ability to classify the new images correctly. This can also help identify overfitting, where a model performs well on the training data but not on the new data. The dataset was split 80:20 for training and testing, respectively; ideally, a validation set would be included, but limited data prevented this.

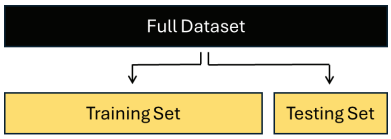


Figure 3: A Model's Division of Data into Training and Testing Data by Sydorenko.[5] This figure shows the standard machine learning practice of splitting a full dataset into training and testing sets. The training set is used to teach the model, while the testing set evaluates its performance on unseen data.

The CNN:

The CNN model used in this research is an artificial intelligence model specializing in processing data with a grid-like topology, such as images. The model is frequently used for tasks like image recognition, object detection, and natural language processing. To achieve this, the CNN model employs a range of convolutional layers, pooling layers, and fully connected layers as shown in Figure 4. The primary function of the convolutional layers is to apply filters to the input of the image to extract its features. In the code, 32 features were detected by a 3x3 filter slid across the image. The output of this process is the production of 32 feature maps, which are slightly smaller than the input images due to the use of 3x3 filters. This is what the first convolutional layer did. A pooling layer was then used to reduce the spatial dimensions of the feature maps. The spatial dimensions were the height and width of a feature map. The output of the pooling layer was that the width and height of the feature map were reduced by half while maintaining the depth of 32. A second convolution

layer was used. For the second convolution layer, 64 different features were detected instead of 32. The size of the filters remained 3x3. The output produced 64 feature maps based on the reduced-size input from the previous pooling layer. These 64 different features were many high-level features than those seen in the first pooling. Another pooling layer was used. The spatial dimensions were further reduced with the use of this pooling layer. The final feature maps were therefore created. The feature maps were three-dimensional (3D). The maps needed to be flattened to make the 3D feature maps into a 1D vector. The CNN contained 128 neurons in each layer. After all of this, the output layer for classification was produced. The model was then compiled. The weights associated with each input were optimized during the training process. The accuracy was also tracked during the training. After the model was compiled, it was ready for training. While training the model, epochs and batch size were identified as the two factors that could affect the accuracy of the model predictions. The epoch was the number of times the model iterated over the entire training dataset for a certain number of times. The batch size was the number of samples into which the dataset was divided. The model updated weights after processing each batch. The weights were then used later when the model was tested on new images for correct classification.

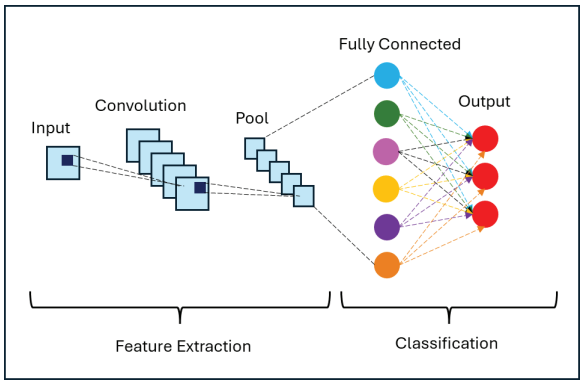


Figure 4: Schematic of CNN architecture (Sthece – ResearchGate), that processes input data through feature extraction layers (convolution and pooling operations) before passing the extracted features to fully connected layers for final classification. The network transforms raw input through hierarchical feature learning, ultimately producing multiple output classifications represented by the red nodes on the right.

Testing the CNN Model:

After being trained, the CNN model was tested on a new data set that it had never seen before. The accurate labels were also stored so that the model could check its accuracy. The model processed the input through its already trained layer to make predictions. The predicted output from the model was compared with the accurate labels. The percentage of classifications was then used to calculate the model's accuracy in detecting correctly. To further evaluate the model's performance, we adjusted various factors in the code during training and testing, including the number of epochs, the number of photos, and the size of the photos.

Results and Discussion

The study's findings demonstrated the potential of CNNs in wildfire detection, as evidenced by varying levels of accuracy based on changing factors, including dataset size, epochs, and image resolution. By optimizing these variables, the CNN model achieved high accuracy in predicting wildfires, as shown in Table 1. The graph in Figure 5 illustrates the relationship between the size of the training dataset (x-axis) and the accuracy (y-axis) of a CNN in identifying wildfires. As the dataset size increased, the model's accuracy improved, demonstrating the importance of larger datasets in enhancing performance in wildfire detection.

Table 1: The average accuracy of each epoch depends on how many photos were in the training and testing data.

	Epoch = 5	Epoch = 10	Epoch = 15	Epoch = 20	Epoch = 25
10 photos	50.00%	50.00%	50.00%	50.00%	50.00%
50 photos	65.00%	60.00%	65.00%	65.00%	70.00%
100 photos	65.00%	67.50%	67.50%	62.50%	67.50%
500 photos	70.50%	71.50%	72.50%	72.00%	73.50%
700 photos	73.50%	72.43%	73.57%	72.36%	74.62%

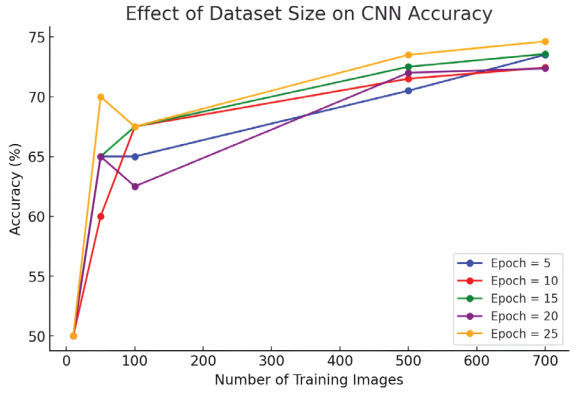


Figure 5: Effect of Dataset Size on CNN Accuracy in Wildfire Identification. Increasing the dataset led to higher model accuracy, highlighting the role of data scale in improving wildfire detection.

Finding various factors that affected the accuracy of the CNN model was very important. The epoch during which the highest accuracy was recorded is noted, as shown in Table 2. The graph in Figure 6 illustrates the highest accuracy achieved by a model at different epochs (x-axis), with accuracy (y-axis) plotted against the number of epochs. It also demonstrates how model performance varies with the number of training epochs. If the model reached 100% multiple times, the epoch (=15), where 100% was first reached, was recorded.

Table 2: Highest accuracy of a certain number of epochs, on detecting a fire or not, using 700 photos for the training and testing data. The epoch numbers shown in brackets in the table below correspond to the epoch in Figure 6 against which the accuracy is noted.

	Epoch = 5	Epoch = 10	Epoch = 15	Epoch = 20	Epoch = 25
700 photos	90.29% (epoch=5)	97.63% (epoch=9)	100.00% (epoch=15)	100.00% (epoch=19)	100.00% (epoch 17)

DOI: 10.36838/v8i3.97

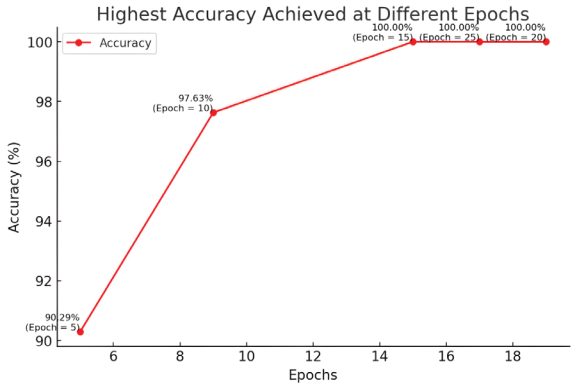


Figure 6: Plots model accuracy against training epochs, showing how performance changes across epochs; if 100% accuracy is reached multiple times (epoch = 15, 20, 25), the first epoch at which it is achieved (e.g., 15) is recorded.

Dataset and Epochs:

The relationship between dataset size and the model's accuracy is shown in Table 1. It was observed that when the model was trained on a larger number of images, such as 500 or 700 photos, the model's accuracy significantly increased. For example, the accuracy increased from an average of 50% with 10 pictures to 74.62% with 700 pictures over 25 epochs. It is suggested, based on these results, that a larger data set provides better training data for the model, enabling the model to generalize better and distinguish between images of fires and non-fires.

The number of epochs also significantly influences the model's performance, with the highest epochs being observed at several intermediate epochs. For example, using 700 photos, accuracy peaked at 74.50% at 5 epochs and slightly fluctuated as epochs increased. Although the accuracy varied depending on the number of epochs, for every row in the table, the accuracy was always higher for the greatest number of epochs used compared to when the fewest number of epochs were used. From this information, it is evident that finding the right balance is crucial, as too few epochs may result in underfitting the model, while too many epochs may lead to overfitting. It is suggested to start with a higher number of epochs, as it usually leads to good accuracy for the model.

Performance at Specific Epochs:

The model achieved the highest accuracy by using 700 photos for training and testing data across different epochs, as shown in Table 2. The model could also reach 100% accuracy as early as epoch 15 and maintain this level throughout the rest of the epoch. The CNN could effectively learn to classify wildfire images accurately, starting from the range around epoch 15. This is extraordinary, as we know the model can reach perfect accuracy, with a sufficient dataset and a suitable number of epochs.

The Impact of Image Resolution:

The model's accuracy was affected by image resolution, as illustrated in Table 3. Images resized to 100x100 pixels achieved the highest accuracy of 73.43% across three trials, outperforming smaller resolutions of 10x10 and even larger resolutions of 250x250, as evidenced in Figure 7. The graph shows the impact of image resolution (x-axis) on the accuracy (y-axis) of a CNN (CNN) across multiple trials. The graph includes data from three trials, demonstrating consistent improvement in accuracy with higher resolutions. These findings indicate that medium-resolution images provide a balance between preserving sufficient detail for feature extraction and minimizing computational demands. Extremely low-resolution images can lose critical features, while extremely high-resolution images can introduce noise or overcomplicate the model. This goes against the common belief that increasing the resolution of a photo always increases the accuracy of a machine-learning model.

Table 3: The accuracy of the model is based on the size to which the images are rescaled. The epoch was set to 10, and 700 photos were used for the training and testing set.

	10x10	50x50	100x100	250x250
Trial 1	71.43%	72.50%	70.29%	72.50%
Trial 2	73.21%	71.36%	76.07%	68.57%
Trial 3	70.71%	71.43%	73.93%	68.21%
Average	71.78%	71.76%	73.43%	69.76%

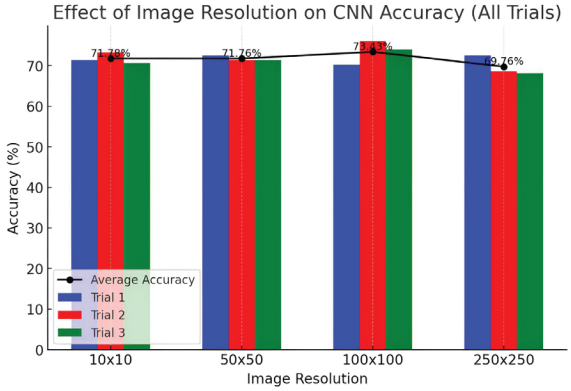


Figure 7: Effect of Image Resolution on CNN Accuracy (All Trials). The results indicate that CNN accuracy is resolution-dependent, with medium-sized images (around 100×100) generally performing best by balancing detail and computation, while very low (10*10) or very high (500*500) resolutions can degrade performance—contradicting the idea that more pixels always help.

Implications for Wildfire Detection:

It is confirmed that machine learning models, such as CNNs, can be highly effective in early wildfire detection. By changing certain factors, such as dataset size, the number of epochs, and image resolution, these models can achieve high levels of accuracy. The early classification of wildfire can save many lives, protect people's property, and reduce environmental damage.

Limitations and Potential Errors:

The project was performed on only one specific dataset. This may not account for all situations in the world, which could cause these results to vary slightly based on other datasets. We could also continue to try to optimize the CNN's structure to converge to an optimal accuracy more quickly.

Future Work:

In the future, more research can be conducted on the model to further maximize its accuracy in predicting whether something is a wildfire. We advise that future research focus on questions such as: Does the model still achieve a high percentage in predicting wildfires from lower-quality images? How well can the model perform under certain weather conditions in the place where the pictures are being taken? What are the common causes of false positives and false negatives in the predictions from the model, and how could this be minimized? Can the model be utilized for real-time detection of wildfires, using video streams?

■ Conclusion

The research shows the effectiveness of the CNN model in accurately detecting wildfires from images, with the optimization of factors such as dataset size, epochs, and image resolution. Unlike traditional methods that rely on delayed reporting to detect wildfires, this approach demonstrates the speed and precision of wildfire detection using a CNN model. The research addresses the research question by affirming that CNN models can serve as an effective tool for wildfire detection when properly trained and optimized. This supports the hypothesis, as changing specific factors can significantly improve the model's accuracy. For example, the experiment demonstrates that using larger datasets and increasing the number of epochs can lead to improved performance. The model can be applied to the real world, using drones and satellites, to enhance the capability of detecting wildfires at an early stage before severe damage occurs. Future advancements can lead to the increased performance of the model. To conclude, the research shows ways that machine learning models, specifically CNN models, can be an effective solution in saving lives and protecting the environment.

■ Acknowledgments

We would like to express our deepest gratitude to our Bronx Science Research teacher, Dr. Vladimir Shapovalov, for his invaluable support, guidance, and encouragement throughout this project. His approval of this research topic, which holds significant implications for Earth and Environmental Sciences as well as the study of Environmental Effects on Ecosystems, has been instrumental in shaping our work.

We are also sincerely thankful to our research mentor, Odysseas Drosis, for his expert guidance in helping us frame our research approach and question, as well as for his assistance in testing and evaluating the accuracy of our model, using Python programming. His insights and mentorship greatly enhanced the quality of this project.

Finally, we extend our heartfelt appreciation to our parents for their constant moral support and encouragement, which motivated us to persevere throughout the course of this research.

■ References

1. Lei, Y., Lei, T. H., Lu, C., Zhang, X., & Wang, F. (2024). Wildfire Smoke: Health Effects, Mechanisms, and Mitigation. Environmental science & technology, 58(48), 21097–21119. https://doi.org/10.1021/acs.est.4c06653
2. Chen, H., Samet, J. M., Bromberg, P. A., & Tong, H. (2021). Cardiovascular health impacts of wildfire smoke exposure. Particle and fibre toxicology, 18(1), 2. https://doi.org/10.1186/s12989-020-00394-8
3. Raoelison, O. D., Valencia, R., Lee, A., Karim, S., Webster, J. P., Poulin, B. A., & Mohanty, S. K. (2023). Wildfire impacts on surface water quality parameters: Cause of data variability and reporting needs. Environmental pollution (Barking, Essex: 1987), 317, 120713. https://doi.org/10.1016/j.envpol.2022.120713
4. El-Madafri I, Peña M, Olmedo-Torre N. The Wildfire Dataset: Enhancing Deep Learning-Based Forest Fire Detection with a Diverse Evolving Open-Source Dataset Focused on Data Representativeness and a Novel Multi-Task Learning Approach. Forests. (2023); 14(9):1697. https://doi.org/10.3390/f14091697
5. Sydorenko, I. (2021). Machine Learning & Training Data. High quality data annotation for Machine Learning. https://labelyourdata.com/articles/machine-learning-and-training-data
6. Jonnalagadda, A. V., & Hashim, H. A. (2024). SegNet: A segmented deep learning-based convolutional neural network approach for drones wildfire detection. arXiv preprint arXiv:2405.00031. https://arxiv.org/abs/2405.00031
7. Maggioros, S., & Tsalkitzis, N. (2024). Wildfire danger prediction optimization with transfer learning. arXiv preprint arXiv:2403.12871. https://arxiv.org/abs/2403.12871
8. Martin, V., Venable, K. B., & Morgan, D. (2024). Development and application of a Sentinel-2 satellite imagery dataset for deep-learning driven forest wildfire detection. arXiv preprint arXiv:2409.16380. https://arxiv.org/abs/2409.16380
9. Phung & Rhee (2019). A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets. Applied Sciences, 9(21), 4500. https://doi.org/10.3390/app9214500
10. Sun, C. (2022). Analyzing multispectral satellite imagery of South American wildfires using deep learning. arXiv preprint arXiv:2201.09671. https://arxiv.org/abs/2201.09671
11. Zheng, Y., Zhang, G., Tan, S., Yang, Z., Wen, D., & Xiao, H. (2023). A forest fire smoke detection model combining convolutional neural network and vision transformer. Frontiers in Forests and Global Change, 6, Article 1136969. https://www.frontiersin.org/articles/10.3389/ffgc.2023.1136969/full

■ Authors

Shaunak Thamke - Shaunak, a rising Bronx Science junior, is a passionate STEM and AI researcher and innovator. He earned top awards at the Terra STEM Fair and STEAMinAI Championship (2025). He leads as Varsity Tennis Captain and Quantum Mechanics Club Co-President. He also volunteers as a certified schoolhouse.world tutor.

Kevin Diao - Kevin, a rising junior and avid tennis player, is a dedicated scholar with a passion for finance and STEM. Earning second place at the Terra STEM Fair (2025), he excels academically, leads as a Varsity Tennis team member, and actively contributes to his school and community.

Elliot Blin - Elliot, a rising junior at Bronx Science, is a multisport athlete, musician, and STEM enthusiast passionate about quantum mechanics. He earned second place at the 2025 Terra STEM Fair, leads as captain and is on 3 varsity teams, and serves as co-president of the Quantum Mechanics Club.

DOI: 10.36838/v8i3.97

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