PROCEEDING Airo

INTERNATIONAL CONFERENCE

WWW.AIRO.CO.IN

5TH INTERNATIONAL CONFERENCE 2024

Integration of Artificial Intelligence In the Advancement of Science & Engineering

Conducted by Airo Journals
iIn Collaboration with Mangalayatan University

Conducted On
28th & 29th
July 2024

MANGALAYATAN
UNIVERSITY

CONVENOR

Prof (Dr.) Kishan Pal Singh Director, Institute of Engineering and Technology, Mangalayatan University Beswan Aligarh

CHIEF PATRON

Prof.(Dr) Tej Narayan ojha Professor Department of Hindi Maharaja Agrasen College University of Delhi

PATRON

Dr. Ajay Kumar Mishra Vice Chancellor Sikkim Skill University

Dr. Ankit Gandhi Director and OSD, Apex University, Jaipur

Dr. Roushan kumar Department of Computer Science ARSD College University of Delhi.

Dr. Vrushali Pramod Parkhi Principal, S.S. Maniar College of Computer & Management, Nagpur

Dr. Diwakar Ramanuj Tripathi HOD Computer Science S.S. Maniar College of Computer & Management, Nagpur

Dr. Ravi Shankar Prasad Associate professor Department of Economics Karorimal college University of Delhi

Dr. Maneesh Vilash Deshpande Assistant Professor Gondwana University, Gadchiroli

www.airo.co.in

IN ASSOCIATION WITH

मङ्गलायतन विश्वविद्यालय
MANGALAYATAN UNIVERSITY
learn today to lead tomorrow

CIIS
COGNITIVE INSTITUTE
OF INTERDISCIPLINARY STUDIES

AIRO INTERNATIONAL JOURNAL

SRDC
SOVARANI D.ED. COLLEGE

www.airo.co.in

INDEX

Leveraging Artificial Intelligence for Progress in Science and Engineering

Dr. Jadab Dutta

Assistant Professor, PQH School of Education, University of Science & Technology Meghalaya (USTM)

Smt. Jyotshna Dowarah

Centre Administrator, Sakhi-One Stop Centre, Dhemaji, Assam, Email id: jyotshnaj.16@gmail.com

Abstract: The application of artificial intelligence (AI) to scientific research and engineering tasks has transformed these fields. Numerous scientific disciplines, including biology, chemistry, physics, and engineering, have significantly advanced as a result of the integration of AI technologies like machine learning, deep learning, and natural language processing. This essay examines how artificial intelligence (AI) is advancing science and engineering, stressing important uses, advantages, difficulties, and potential future directions.

Keywords: Artificial Intelligence (AI), Science & Engineering

Introduction

The creation of intelligent systems that are able to carry out activities that normally require human intelligence is the fast developing field of artificial intelligence, or AI. Artificial intelligence (AI) technologies have been progressively incorporated into a range of scientific and engineering fields in order to improve research capacities, speed up discovery, and streamline problem-solving procedures. The integration of AI in science and engineering has led to groundbreaking advancements in areas such as data analysis, modeling, simulation, and design optimization.

Applications of AI in Science

AI has been widely applied in various scientific fields to automate data analysis, improve prediction accuracy, and uncover hidden patterns in complex datasets. In biology, AI-powered algorithms have been used to analyze genomic data, predict drug interactions, and accelerate drug discovery. In chemistry, AI models have been developed to predict chemical reactions, optimize drug design, and identify new materials with desired properties. In physics, AI techniques have been applied to analyze particle physics data, optimize experimental designs, and simulate complex systems.

Applications of AI in Engineering

In engineering disciplines, AI technologies have been instrumental in optimizing design processes, improving product performance, and enhancing manufacturing efficiency. AI-driven simulation tools enable engineers to explore design alternatives, predict system behavior, and optimize performance parameters. In robotics, AI algorithms are used to enhance autonomy, perception, and decision-making capabilities of robotic systems. In civil engineering, AI is employed for predictive maintenance, infrastructure monitoring, and smart city development.

Benefits of AI Integration in Science and Engineering

The integration of AI in science and engineering offers numerous benefits, including increased research productivity, enhanced problem-solving capabilities, and improved decision-making. AI technologies enable researchers and engineers to analyze large datasets quickly, identify patterns, and generate actionable insights. AI-driven simulations help optimize designs, predict system behavior, and reduce testing costs. Additionally, AI enhances automation, accelerates innovation, and enables the development of novel solutions to complex problems.

Challenges and Future Opportunities

Despite the significant progress in AI integration in science and engineering, several challenges remain, such as data quality issues, interpretability of AI models, and ethical concerns. Addressing these challenges requires interdisciplinary collaboration, transparent AI solutions, and ethical guidelines for AI deployment. Looking ahead, the future of AI in science and engineering holds immense potential for further advancements, including personalized medicine, autonomous systems, and sustainable engineering solutions. Continued research and innovation in AI technologies will drive transformative changes in scientific research and engineering practice.

Conclusion

To sum up, the application of AI to science and engineering has proven to have revolutionary potential in terms of quickening research, streamlining procedures, and resolving challenging issues. Researchers and engineers can open up new avenues, spur innovation, and make ground-breaking discoveries by utilising AI technologies. AI will have a significant impact on scientific and engineering fields as it develops, influencing technology in the future and advancing the pursuit of knowledge and technical brilliance.

Recommendation

The following suggestions are made to optimise the advantages and handle the difficulties related to the integration of Artificial Intelligence (AI) in the growth of science and engineering, based on the insights presented in this study:

i. **Promote Interdisciplinary Collaboration:** To aid in the creation of AI solutions suited to the unique requirements and difficulties of many scientific and technical domains, promote cooperation between AI specialists, scientists, and engineers. This will ensure that AI technologies are effectively integrated and utilized to their full potential.

ii. **Enhance Data Quality and Accessibility:** Invest in improving the quality and accessibility of data used in AI applications. Annotated, high-quality datasets are essentialfor training dependable and accurate artificial intelligence algorithms. Establishing standardized data formats and protocols can facilitate data sharing and collaboration acrossdifferent research communities.

iii. **Focus on Interpretability and Transparency:** Develop AI models that are interpretable and transparent to build trust among scientists and engineers. Users must be able to comprehend how decisions are made in transparent AI systems in order to verify findings and guarantee the moral application of AI technologies.

iv. **Implement Ethical Guidelines:** Establish and adhere to ethical guidelines for the development and deployment of AI in science and engineering. Addressing ethical concerns, such as bias in AI models and the potential impact on employment, is critical for ensuring responsible and fair use of AI technologies.

v. **Invest in Education and Training:** Provide education and training opportunities for scientists and engineers to develop AI skills. Equipping researchers and practitioners with the knowledge and tools to effectively use AI will enhance their ability to leverage these technologies in their work.

vi. **Support AI Research and Innovation:** Continue to invest in AI research and innovation to drive further advancements in AI technologies. New tools and techniques that may be used in a variety of scientific and engineering disciplines will be developed as a result of supporting research in fields including machine learning, deep learning, and natural language processing.

vii. **Foster Public-Private Partnerships:** Promote partnerships between academicinstitutions, industry, and government agencies to accelerate the development and application of AI in science and engineering. Collaborative efforts can lead to the creation of innovative solutions and the translation of research findings into practical applications.

viii. **Encourage Sustainable AI Solutions:** Focus on developing AI technologies that contribute to sustainable engineering solutions. This entails maximising the use of resources, cutting down on waste, and lessening the negative effects of engineering operations on the environment.

By implementing these recommendations, the integration of AI in science and engineering can be optimized, leading to enhanced research capabilities, improved problem-solving, and the achievement of groundbreaking advancements in various fields.

References

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. Journal of artificial intelligence research, 4(1), 237-285.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

DEVELOPING A SOLAR-POWERED ELECTRIC VEHICLES CHARGING SYSTEM UTILIZING VEHICLE -TO - GRID-ENABLED SMART TECHNOLOGY

Praveen Kumar Sharma [1] AnjanaTiwari [2]

Research Scholar, University of Technology, Jaipur, Rajasthan, India,

Email: Praveenprem.sharma@gmail.com

2 Professor Department of Electrical Engineering, LIET College ,Alwar Rajasthan, India,

Email: tiwarianjana594@gmail.com

ABSTRACT

Two fundamental concerns have evolved in the global energy business during the past couple of decades. The first is the extensive use of fossil fuels, which causes their rapid depletion, and the second is carbon emission, which causes global warming. For energy, the majority of nations rely on fossil fuels and natural gas. Fossil fuels are not just nonrenewable, but also limited resources that will eventually run out and become too expensive and environmentally damaging to extract. However, distributed energy sources such as wind and solar energy are renewable, non-depleting, and environmentally friendly. Renewable energy is also dubbed "green energy" because it does not generate air pollution or carbon emissions. Transportation is one of the key industries that use fossil fuels. Battery Electric Vehicles (BEVs) were developed as an alternative to Internal Combustion Engines (ICEs) in an effort to reduce carbon dioxide (CO_2) emissions and fossil fuel consumption (ICEs).

Keywords: Smart technology, V2G Technology, solar PV system, Battery.

1. INTRODUCTION

1.1 OVERVIEW

Globally, electric vehicles (EVs) are a novel concept in the transportation sector. In 2030, EVs are projected to make up 24% of the U.S. light vehicle fleet, compared to 64% of light vehicle sales in 2018. In this context, the battery charging process of EVs must be managed in order to maintain the power quality of the power grids. With the spread of Evs, however, a substantial quantity of energy will be stored in the batteries, hence increasing the likelihood of energy transfer in the other direction. Interaction with electric vehicles will be one of the important technologies in future smart grids, contributing to the autonomous functioning of the power grid. The notion of an on-board bidirectional charger utilising V2G and V2H technologies is presented. Due to lower carbon dioxide emissions and increasing fossil fuel prices, the electric vehicle has become more competitive than the conventional internal combustion engine vehicle. However, the EV was not generally embraced on the market due to restrictions such as expensive car costs; Limited charging infrastructure and limited all-electric driving range. EVs are cars that are partially or completely electric-powered. Electric vehicles offer minimal operating expenses since there are fewer moving parts that need to be maintained, and they are also very environmentally friendly because they consume little or no fossil fuel.

1.2 ELECTRIC VEHICLE

A vehicle that propels itself using one or more electric motors or traction motors is referred to as an electric vehicle (EV). Electric vehicles can be self-contained using a battery, solar panels, fuel cells, or an electric generator to convert gasoline to electricity, or they can be fueled by electricity from sources outside the vehicle through a collector system. Electric vehicles (EVs) include, but are not limited to, surface and underwater craft, electric Aeroplan's, and electric spaceships. EVs initially appeared in the middle of the 19th century, when electricity was one of the preferred forms of motor vehicle propulsion. At the time, gasoline-powered cars were unable to match the comfort and ease of operation that electric vehicles offered. While modern internal combustion engines have dominated motor vehicle propulsion for almost a century, electric power has remained prevalent in other vehicle types, such as railways and smaller vehicles of various kinds.

An electric motor replaces the internal combustion engine in all-electric vehicles, also known as battery electric vehicles (BEVs). The electric motor of the vehicle is powered by a sizable traction battery pack, which must be plugged into a wall outlet or charging apparatus, also known as electric vehicle supply apparatus (EVSE). The car does not have a tailpipe or any typical liquid fuel components like a fuel tank, fuel line, or fuel pump because it is an electric vehicle. Find out more about electric cars.

There are four types of electric vehicles available:

- **Battery Electric Vehicle (BEV)**

BEVs are also referred to as all-electric cars (AEV). Electric drivetrains powered solely by batteries are used in BEV-based electric vehicles. The large battery pack that houses the electricity needed to power the car can be charged by plugging it into the power grid. One or more electric motors are then powered by the fully charged battery pack to drive the electric vehicle. It is fully electrically powered. When compared to hybrid and plug-in hybrid vehicles, these are more efficient.

Main Components of BEV:

Electric motor, Inverter, Battery, Control Module, Drive train

- **Hybrid Electric Vehicle:**

HEVs are also referred to as parallel or series hybrids. HEVs have an electric motor in addition to an engine. Fuel powers the engine, while batteries provide electricity for the motor. Both the engine and the electric motor turn the transmission at the same time. Wheels are then propelled by this. Both the internal combustion (typically gasoline) engine and the battery-powered motor powertrain are utilised by the vehicle. When the battery is dead, the petrol engine is used to both propel and charge the vehicle. Compared to fully electric or plug-in hybrid vehicles, these cars are less efficient.

Main Components of HEV:

Engine, Electric motor, Battery pack with controller & inverter, Fuel tank, Control module

- **Plug-in Hybrid Electric Vehicle (PHEV)**

uses a battery that is charged by an external socket and an internal combustion engine (they have a plug). This implies that electricity, rather than the vehicle's engine, can be used to recharge the battery. While less efficient than BEVs, PHEVs are more efficient than HEVs.The term "series hybrid" also applies to PHEVs. Both an engine and a motor are present. You have a choice of two types of fuels: conventional fuel (like gasoline) and alternative fuel (such as bio-diesel). A battery pack that can be recharged can also power it. The battery can receive external charging.

Main Components of PHEV:

Electric motor, Engine, Inverter, Battery, Fuel tank, Control module, Battery Charger (if onboard model)

- **Fuel Cell Electric Vehicle (FCEV)**

Another name for FCEVs is zero-emission vehicles. To create the electricity needed to power the vehicle, they use "fuel cell technology." The fuel's chemical energy is directly converted into electric energy. Chemical energy is converted into electric energy. Consider an FCEV powered by hydrogen.

Main Components of FCEV:

Electric motor, Fuel-cell stack, Hydrogen storage tank, battery with converter and controller

1.3 SOLAR ENERGY

Solar energy, the sun's rays that can ignite chemical reactions, produces heat, or creates electricity. The total solar energy incident on Earth far exceeds both the present and future energy needs of the planet. This highly diffused source might potentially provide all of the energy required in the future ifproperly harnessed. Due to its limitless supply and lack of environmental impact compared to the finitefossil fuels coal, petroleum, and natural gas, solar energy is predicted to gain popularity as a renewableenergy source in the twenty-first century.

Although the Sun is a very potent energy source and sunlight is by far the most abundant energy that Earth receives, the intensity of sunlight at the planet's surface is actually rather low. The massive radial radiation radiating from the far-off Sun is mostly to blame for this. Earth's atmosphere and clouds cause up to 54% of the incoming sunlight to be absorbed or scattered, which results in a relatively small additional loss. Nearly half of the sunlight that reaches the earth is composed of visible light, while the other half is made up of infrared radiation, with smaller amounts of ultraviolet and other electromagnetic radiation.

Since the Earth receives solar energy every day in the amount of nearly 200,000 times the world's daily electric generating capacity, the potential for solar energy is huge. Even though solar energy is free in itself, the high expense of gathering, converting, and storing it prevents widespread use of it. Although the former is simpler to achieve, solar radiation can be transformed into electrical energy as well as thermal energy (heat).

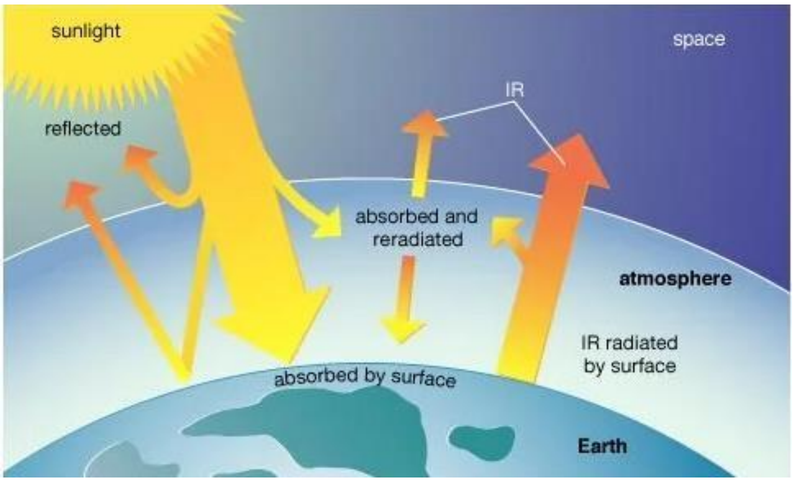


Figure 1: Solar Energy

Table 1: Advantages and Disadvantages of Solar Energy

Advantages of Solar Energy	**Disadvantages of Solar Energy**
Renewable Energy Source	Cost
Reduces Electricity Bills	Weather Dependent
Diverse Applications	Solar Energy Storage is Expensive
Low Maintenance Costs	Uses a Lot of Space
Technology Development	Associated with Pollution

1.4 SOLAR POWERED ELECTRIC VEHICLE CHARGING STATION

1.4.1 Meaning

One of the most effective ways to lessen India's reliance on fossil fuels for the powering of various modes of transportation is through solar charging stations for electric vehicles. This is because electric vehicles typically use electricity generated from fossil fuels, which is a major cause for concern.

It is essential to implement solar charging for electric cars and bikes as the popularity of electric vehicles rises. There are currently two types of solar charging stations for electric vehicles, depending on the configuration.

Figure 2: Solar Energy Charging for Electric Vehicles

We can investigate the viability of developing a PV-based infrastructure for electric vehicle charging. The technology is made to be used at workplaces to charge employees' electric vehicles while they are left parked during the day. The goal is to use PV energy as much as possible for EV charging while utilising the grid as little as possible. Such an EV-PV charger's benefits include:

- Because EV charging uses locally produced, environmentally friendly power from solar panels, there is a decreased need for energy from the grid.
- EV batteries also serve as energy storage for PV, which lessens the detrimental effects of integrating large amounts of PV into the distribution network.
- An extension of Vehicle-to-Grid (V2G) technology, in which an EV serves as a controllable spinning reserve for the smart grid, is made possible by the long parking times of EVs.

1.4.2 Types of Solar Based Electric Vehicle Charging Station

The key to significantly lowering our reliance on fossil fuels is the integration of solar energy and EV charging. There are many different ways to get electricity, so it's essential that electric vehicles run on renewable energy sources. A solar charging station will likely be installed at every home that has a solar energy system in the coming years as electric cars become increasingly popular. For this to happen, we will need to think about refuelling our cars differently and for our energy infrastructure to naturally evolve.

- ✓ **Off-grid Solar Based Electric Vehicle Charging Station**

The charging station is not connected to nearby utilities in this configuration. As a result, it is also known as an autonomous EV charger. Here, the battery storage system is fueled by the solar panel array. Additionally, this battery storage system meets the charger's entire power requirement. This kind of charging station can be installed almost anywhere because it doesn't require a connection to the grid. Additionally, they are simple to install because the majority of them have a sturdy steel foundation.

"Electric Vehicle Autonomous Renewable Charger" is another name for an off-grid auto charger. No local utility connection is necessary. The entire system's power requirements are met by this energy storage system, which is powered by the solar panel array. Since there is no requirement for a connection to the electrical grid, off-grid electrical car chargers can be installed almost anywhere. A sturdy foundation is necessary because the independent solar array canopy attracts a lot of wind. Some off-grid solar energy chargers have a large, ballast-serving base plate made of steel. Since there is no foundation or digging involved, installation of those is incredibly simple and quick.

- ✓ **On-grid Solar Powered Charging Stations for Electric Vehicles**

Because the energy produced by the solar array is stored in the grid rather than in batteries, the cost of an on-grid solar EV charger is unquestionably lower than that of the off-grid version. You receive credits from the utility provider when you feed the grid, which you can use to charge your electric vehicle. Your excess energy is sold to the utility company.

The simplest way to power your electric car with solar energy is with a grid-tied solar energy system. Whether or not your home requires the power at the time, a grid-tied solar energy system will feed the energy into the grid. Therefore, the electricity generated at home is sold to the utility company when your solar energy system is feeding the grid and you are at work. That power will be returned to you by the utility company in the form of a credit. You can use that credit to recharge your car at home after leaving work and leaving it parked there.

1.4.3 Components Needed for a Solar EV Charging Station

- an EV charger
- a strong base for a standalone solar charger
- Software that is sufficiently intelligent for billing and other tasks
- battery-based energy storage (exclusively for solar energy charging stations that are off grid)
- Solar panel array installed on the ground or roof for capturing solar energy

1.5 V2G "VEHICLE TO GRID"

V2G, which stands for "vehicle to grid," is a technology that allows energy to be returned from an electric car's battery to the electrical grid. A car battery can be charged and drained based on various signals, such as energy generation or consumption locally, with the use of electric vehicle to grid technology, sometimes referred to as car-to-grid.

Electric needs are steadily rising against the backdrop of the energy transition, in part because of the advancement of electric transportation. By 2040, 50% of all new automobiles sold will be electrified, predicts the BloombergNEF research. In a period of global turmoil, it will be necessary to power all of these cars in a smart and practical manner.

The electric car has the ability to store, distribute, or even produce power, much like a home with solar panels. A new generation of interactive cars known as Vehicle To Grid or V2G has replaced the passive electric vehicle, which does nothing except consume energy. V2G technology, as its name suggests, is the practise of feeding electricity stored in an electric car's batteries back into the electrical grid while it is parked. A smart grid is an electrical network system that employs information technology to regulate energy usage, and this technology is an element of it.

The smart grid, on a global scale, encourages data sharing between suppliers and consumers to address an important issue: energy storage. Large amounts of electricity are challenging to store. Real-time balance is required for management. Either not enough electricity is produced to power the network, or too much is produced and much of it is wasted. In this effort to modify the electrical flow, V2G may have a significant role.

The V2G battery functions as an extension of the electrical grid by storing energy produced when demand is lower and reintroducing it into the system when demand is higher thanks to bi-directional charging. As electric vehicles become increasingly common, it's simple to picture the benefits of this technology. With the use of Vehicle to Grid (V2G) technology, an electric vehicle might significantly contribute to a world free of emissions by using a big amount of renewable energy sources.

2. REVIEW OF LITERATURE

Ram Vara Prasad, Bugatha et. al (2022) This study outlines the concept of a solar-powered charging station for electric vehicles that eliminates the major drawbacks of fuel and air pollution. Globally, electric vehicles are currently on the road, and their number is gradually increasing. In addition to their positive effects on the environment, electric vehicles have been shown to reduce travel expenses by substituting petrol with energy, which is significantly less expensive. Consequently, we create an electric vehicle (EV) charging system that offers a novel and revolutionary solution. This method of recharging electric automobiles wirelessly, there is no need to stop for recharging, as the car charges while in motion. Solar energy is used to power the charging system; thus, no other power source is required. The system is constructed using a solar panel, battery, transformer, regulator circuitry, copper coils, an AC-to-DC converter, an at mega controller, and an LCD display. The device displays how electric vehicles may be charged while in motion, eliminating the need for charging stops. Consequently, the technology exhibits an integrated solar-powered wireless charging solution for electric automobiles.

Dighe, Amol & Rakesh et. al. (2022) the charging infrastructure for electric vehicles (EV) could be the most crucial aspect in ensuring a smooth transition to e-mobility. This study focuses on five developments that have the potential to play an alphabetical role in this regard: shrewd charging, vehicle-to-matrix (V2G), charging of electric vehicles using solar panels (PV), and. Contactless and on-street charging alternatives are available for EVs. Smart charging of EVs is anticipated to allow a large number of EVs to enter the market, supply environmentally friendly electricity, reduce the cost of charging, and provide more. The application of a lattice design; Bidirectional EV antennas will pave the way for the V2G era, in which EVs will be able to exchange energy and request future activities. Sun-oriented EV charging will have an effect on reasonable mobility and the use of EV batteries. Alternatively, the work area is constrained. Contactless and inductive charging of electric vehicles on the road will eliminate any strains and range tension; Concerns and preparations for updated application. This research examines the electromagnetic and energy engine strategy for contactless power move structures in future streets.

Sasikumar, Gnanasekaran & Sivasangari (2021) Renewable energy is a type of energy derived from a variety of sources, such as sunshine, wind, tides, geothermal, etc. It provides renewable energyderived from renewable natural resources. Increasing the use of renewable energy will reduce the demand for and price of fossil fuels. Solar photovoltaic energy is largely used for numerous uses, including heating, cooking, and electricity generating. Recent innovations contributed to the creation of solar-powered automobiles. This paper discusses the design and construction of a solar charging system for electric vehicles with a charge controller. Implementation of the proposed technology willdecrease the cost of electricity and charging and discharging losses. In addition, the projected solar charging system will be one of the initiatives implemented to establish a "green campus." This articlewill analyze the system design and performance of a solar-powered electric vehicle system.

Sheeba, R., Mohammed Sulthan et al (2021) As a dependable alternative to gas-powered automobiles, electric vehicles are rapidly gaining popularity. For ongoing operation, these vehicles' batteries must be "refuelled." Solar-powered chargers have emerged as an intriguing alternative to the typical grid-based charging of electric vehicles. These chargers are pollution-free and supply electric automobiles with clean power. This paper describes the design of an intelligent hybrid electric vehicle charging station appropriate for both personal and commercial use.

Kumar, Rajan & Bharj et. al (2021) As part of their strategy to combat climate change and rising urban pollution, a developing nation such as India is rapidly adopting technologies connected to electric vehicles (EVs) and phasing out fossil fuel-powered vehicles. The Government of India (GoI) intended in April 2017 to have all EVs on the market by 2030. Also pursued is the promotion under the faster adoption and production of electric vehicles (FAME) scheme. The infrastructure for electric charging is a crucial component of the electric mobility ecosystem. It is crucial for EV charging station markets to match the acceptance and expansion of EVs. Electric vehicles are limited in range and speed. The availability of charging stations and their network on the road is essential for facilitating the transition from fossil fuel vehicles to electric vehicles. Consequently, the accessibility of a plug points for charging remains an ongoing challenge for the majority of EV producers and users. Certainly, it is necessary to transition from grid-based charging stations to autonomous off-grid options for charging. Utilizing abundant renewable energy sources like solar energy is the key to resolving this issue. This chapter provides a comprehensive assessment of the infrastructure, technique, and implementation of EV charging systems and solar-based EV charging systems in India. Various obstacles and social obstacles to the adoption of electric vehicles are also explored.

Mehrjerdi, Hasan & Rakhshani, Elyas (2019) this article optimizes the operation of electric vehicle charging stations in electrical distribution networks. Instead of a single large-scale charging station, the grid is equipped with numerous small-scale charging stations. In addition, photovoltaic solar panels are installed on the grid in order to reap the benefits of renewable energy. The electric vehicles function in grid-to-vehicle mode. The suggested solution optimizes the charging and discharging behaviour of all electric vehicles on all buses. The offered technique optimizes the functioning of electric vehicles to mitigate the intermittent nature of renewable energy while simultaneously reducing energy costs. Moreover, it reduces the charging and discharging cycles of the vehicle's batteries to prevent battery degradation. The proposed challenge is treated as a nonlinear stochastic programming problem that incorporates the unpredictability of solar energy and is solved using the GAMS software. The results indicate that the proposed technology can charge and discharge the vehicle-to-grid system effectively. The vehicles are frequently charged during off-peak low-cost intervals and discharged during peak high-cost times. The intermittent nature of solar cells is mitigated by the established charging-discharging pattern for vehicle-to-grid systems, and energy shifting reduces the cost of consumed energy. The results demonstrate that the provided technique is capable of simultaneously achieving all planned goals.

Ahmad, Furkan & Shariff et. al. (2019) A high-power battery charger featuring PFC (power factor correction) and an AC/DC converter to adjust the charging current output ripple content, hence giving continuous DC to high-energy battery packs used in electric vehicles (EVs). This paper addresses the practical design and implementation features of an EV charger for charging the EV battery pack using a solar-assisted EV charging application. The operation of the circuit is analyzed, and a scientific model is developed to study the design aspect of circuit parameters. Finally, the framework's mathematical modelling is created in the MATLAB/Simulink environment to test the performance ofthe PFC under steady-state conditions with respect to load variation for a 3kW, 230Vrms input at a single phase 50 Hz rated supply to produce 48V DC EV battery charger buck converter output. The charging process is managed by Electric Vehicle Supply Equipment (EVSE) using a level 2 AC charging system based on SAE J1772. As a result, a simplified design on the system level will be explained, together with the whole set of functions of the integrated charging system. PROTEUS software has been used to simulate the communication and signaling circuit, and a prototype model has been implemented in the lab. A case study of a 6.4 kW solar photovoltaic charging station (SPVCS)placed in the Centre of Advanced Research in Electrified Transportation (CARET) building parking area on the Aligarh Muslim University (AMU) campus is used to describe the hardware model. The experiment is carried out on a 10-kWh lithium-ion battery pack on a bright sunny day under the solar panel's standard test conditions (STC).

Nair, Mohan, and Harin (2019) the solar-powered plug-in electric vehicle is a cost-effective vehicle that requires little maintenance. The biggest disadvantage of electric vehicles is their limited driving range. The vehicle battery can be charged while driving by installing a solar PV panel. Mechanical components such as the gearbox and differential are avoided. Direct drive to the wheels allows for more efficient driving.

John Kaldellis and Spyropoulos (2017) Improving transportation energy efficiency could help to reduce environmental deterioration and slow the depletion of existing fossil-fuel reserves. Adoption of eco-driving, particularly in metropolitan areas, the use of more efficient vehicles, and the transitionto green public transportation are all effective techniques for enhancing energy efficiency. In any case,the usage of so-called clean new technology cars should be used in certain circumstances (e.g., smart cities) to establish a sustainable and efficient transportation strategy. The Piraeus University of Applied Sciences' Laboratory of Soft Energy Applications and Environmental Protection (SEALAB) has recently undertaken, as part of its innovative activities, the development, construction, and operation of the country's first stand-alone solar electric vehicle charging station (EVCS), CARPORT, monitoring all energy data and thus supporting and strengthening the country's efforts in infrastructure development. More particular, the novel initiative detailed in this chapter aims to speed the implementation of a European national electrification action plan through the development of EVCSs based on photovoltaic generators. The suggested solar EVCS is regarded as one of the most environmentally benign alternatives capable of assisting in the decarbonization of Europe's transportation industry.

Chandra Mouli & Gautam Ram (2016) This research analyses the possibilities of using solar energy to charge electric automobiles at the workplace in the Netherlands. The Dutch Meteorological Institute's data is utilized to estimate the ideal orientation of photovoltaic panels for maximum energy production in the Netherlands. Seasonal and diurnal variations in solar insolation are studied to assess the energy availability for EV charging and the need for grid connection. Due to Netherlands' relatively low solar insolation, it has been discovered that the power rating of the PV array can be 30% higher than the power rating of the converter. Various dynamic EV charging profiles are evaluated in an effort to reduce grid dependence and maximize the use of solar energy to directly charge the EV. Considered are two scenarios: one in which EVs must be charged exclusively on weekdays, and another in which they must be charged seven days per week. Proposed is a priority mechanism to permit the charging of numerous EVs from a single EV–PV charger. The viability of connecting a local storage with an EV–PV charger to make it grid-independent is assessed. The ideal storage size that lowers grid dependence by 25 percent is determined.

Yilmaz, M. & Krein, Philip (2013) the vehicle-to-grid (V2G) connection idea allows plug-in vehicles to function as either loads or as a dispersed energy and power supply. This article examines the current status and implementation impact of V2G/grid-to-vehicle (G2V) technologies on distributed systems, as well as the needs, benefits, obstacles, and strategies for V2G interfaces on both individual and fleet cars. The V2G idea can enhance the performance of the energy grid in areas like efficiency, stability, and dependability. A vehicle with V2G capabilities provides reactive power assistance, active power regulation, tracking of changeable renewable energy sources, load balancing, and current harmonic filtering. These technologies can provide auxiliary services such as voltage and frequency regulation and spinning reserve. V2G costs include battery degradation, the requirement for intensive communication between vehicles and the grid, effects on grid distribution equipment, infrastructure modifications, and social, political, cultural, and technical hurdles. Despite the fact that vehicle-to-grid (V2G) operation can shorten the lifespan of car batteries, it is expected to become cost-effective for vehicle owners and grid operators. Components and unidirectional/bidirectional power flow technologies of V2G systems, individual and aggregated architectures, and charging/recharging frequency and tactics (smart, uncoordinated/coordinated) are discussed. Successful V2G operation requires three elements: power connection to the grid, control and communication between vehicles and grid operator, and on-board/off-board intelligent metering. The success of the V2G concept is contingent upon the standardization of requirements and infrastructure decisions, battery technology, and the effective and judicious scheduling of limited fast-charge infrastructure. The deployment of a charging/discharging infrastructure is required. The economic benefits of vehicle-to-grid (V2G) technologies are contingent upon vehicle aggregation, charging/recharging frequency and methods. In the future, grid operators and car owners will give greater consideration to the advantages.

3. RESEARCH METHODOLOGY

3.1 PROBLEM STATEMENT

The government has outlined numerous strategies to address environmental issues. Electric mobility and renewable energy are two such priority sectors that has been identified. India has an energy shortage, and the majority of its cities frequently experience power outages. Therefore, in light of electricity fluctuations, the current plan to target EVs on a mass scale cannot be realized. Before EVs are widely adopted, additional issues like the lack of fast charging infrastructure, the higher price of EVs, the risk of energy theft, etc., must be resolved.

In addition, the government has established a lofty rooftop solar target. Rooftop solar energy is advantageous in many situations because it reduces AT&C losses, does not require additional land (if installed on existing structures), and does not require transmission infrastructure. Rooftop solar still faces challenges, however, from institutional inconsistencies to slow grant and funding processing to low consumer awareness. Furthermore, problems with power quality result from an increase in the generation of renewable energy at the grid distribution level. The distribution grid network needs to be strengthened for greater bidirectional power flow due to the future energy demand growth that willbe exponential and the high penetration of RE.

Hence, the thesis entitled *“DEVELOPING A SOLAR-POWERED ELECTRIC VEHICLES CHARGING SYSTEM UTILIZING VEHICLE -TO - GRID-ENABLED SMART TECHNOLOGY”*develops a V2G-enabled smart charging solution for electric vehicles that is highly efficient. Charging electric vehicles with photovoltaic panels can make EVs completely sustainable and reduce the infrastructure's net cost. Three aspects are required to establish a solar-powered EV charging infrastructure: system-level design, the development of a solar EV charger, and the formulation of smart charging algorithms.

3.2 AIMS &OBJECTIVES

The objectives of the study are as follows:

1) To study the concept of solar energy, V2G technology and solar power electric vehicles.
2) To develop a highly efficientsolar-powered, V2G-enabled smart charging system for electric vehicles at workplaces
3) To discuss the system architecture and power converter topologies for three-port EV-PV-charging system using detailed modeling
4) To examine the smart EV charging algorithms to reduce the cost of charging EVs using the developed EV-PV converter
5) To conduct the experimental testing on the Two CHAdeMO and CCS compatible EVs with the implementation of smart charging and V2G,

6) To assess the ideal PV system and local storage for charging electric vehicles at offices building

3.3 RESEARCH DESIGN

The design of this research will be three-fold in nature means the analysis will be divided into 3 parts.For a solar powered EV charging infrastructure, three elements will be essential which is system-level design, development of a solar EV charger and formulation of smart charging algorithms.

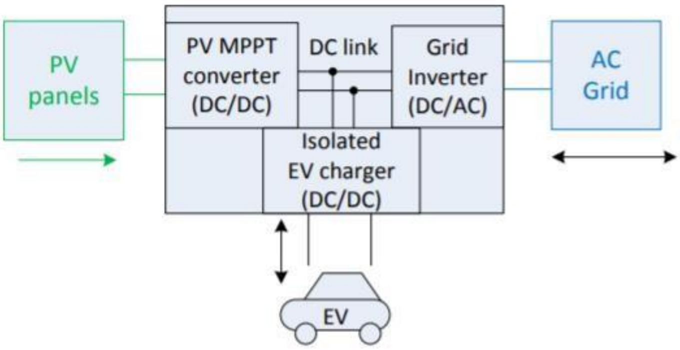
3.3.1 Experimental Setup on Topology for Three-Port EV-PV-Grid Converter

- **Loss Simulation and Converter Design**

In order to design a converter topology for maximum efficiency and power density, this section will discuss how to take into account a variety of design factors, including switching frequency, magnetic core, copper windings, semiconductor devices, capacitors, and heat sinks. This will require a number of iterations using an accurate loss model of the converter where the design parameters will be changed at each cycle, as shown in Fig. 4. The converter components, volume, and efficiency (i.e., losses) of the optimized design will be obtained at the end of the iterations**.**

Table 2. Specifications of EV-PV converter

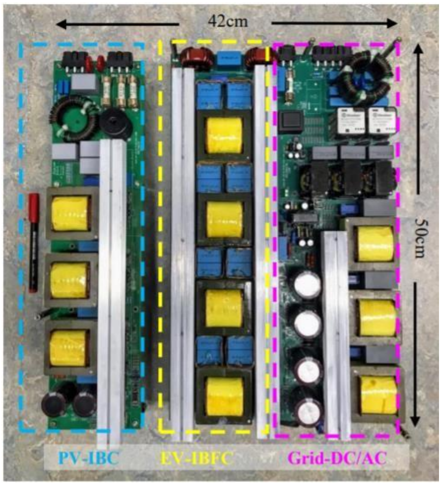
Parameter	Symbol	Value
Nominal power	P_{nom}	10kW
PV MPPT Voltage	V_{pv}	350-700V
PV MPPT Current	I_{pv}	0 - 30A
PV current ripple (peak-peak)	$\Delta I_{pv\%}$	< 10% of I_{pv}(max)
PV voltage ripple (peak-peak)	$\Delta V_{pv\%}$	< 0.5%
EV voltage	V_{ev}	200-500V
EV current (Bidirectional)	I_{ev}	-30A +30A
Internal DC-link voltage	V_{dc}	750V (for Arch. 1)
Total Harmonic Distortion		< 5%
AC grid connection		400V, 50Hz AC, 16A



Grid-connected, three-port, 10kW bidirectional EV-PV charger

3.3.2 Experimental Setup on Development of 10kw Bidirectional Solar EV Charger

- **EV-PV converter modular prototype**



Experimental 10kW three-port EV-PV converter prototype

The developed EV-PV converter prototype is shown above along with the grid inverter, PV IBC, and EV IBFC. The converter will be constructed modularly as two power modules to give the user flexibility and to improve the EV-PV charger's commercial viability. The PV IBC and its controller will each be built on two PCBs and make up the first power module. The IBFC, three-phase inverter, and their controllers will be located on one PCB and one PCB, respectively, in the second power module. The common DC-link and the communication port of the two controllers will be connected to join the two PCBs together. The DC-link will therefore be assessed from the outside and offers the opportunity to connect to additional PV power modules, EV charger power modules, and potential DC-grids. The control PCB will be located on the back of the power PCB on both power modules, giving the finished EV-PV converter its final dimensions of 50x42x12cm. The power density of the PV converter, EV+Grid converter, and full EV-PV converter will be 1380W/l, 555W/l, and 396W/l, respectively, based on cabinet dimensions.

- **Experimental waveforms**

When a PV emulator is used as input, Fig. 6(a) displays the operating waveforms of the PV IBC for continuous conduction mode (CCM), and Fig. as given below will display the waveforms for discontinuous conduction mode (DCM). The waveforms for the MOSFET drain-source voltage V_{ds}, inductor current I_L, and phase-shifted gate voltage VGS will be displayed in the figure. As shown in Fig. 6, the inductor current $I_{L\,(1)}$ rises during CCM when the gate voltage $V_{GS\,(1)}$ will be ON and then starts to fall when the gate will be turned off (a). In DCM, the drain-source voltage $V_{ds(1)}$ oscillates as it will transition from $V_{ds(on)}$ to VPV because the inductor current $I_{L(1)}$ reaches zero prior to the end of the switching cycle (b).

(a)

(b)

Drain-source voltage V_{ds} and gate voltage V_{gs} for one phase of the IBFC for CH mode: (a) Quasi-resonant operation LVS for V_{ev}=250V, I_{ev}=5A (b) Valley skipping and DCM operation at low powers for V_{ev}=100V, I_{ev}=1A

3.4 SYSTEM DESIGN

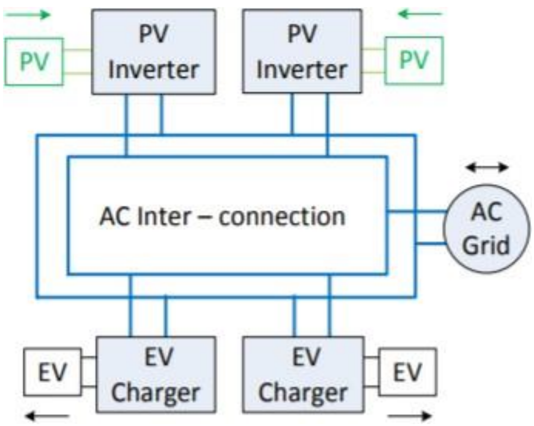
The system design of the solar EV charging station will examine the optimal design for the photovoltaic (PV) system to satisfy the EV charging requirements. Simple charging systems, such as Gaussian EV charging, will be developed to assist in synchronizing EV charging with PV generation and reducing grid dependence. The use of a local storage will be found to aid in regulating the diurnal solar variations, but will have a little effect on overcoming the seasonal solar variations. Despite the India’s lower solar insolation, a 10kWp PV system will produce an average of 30kWh per day. This is enough for a Nissan Leaf EV to travel 55,000 km per year. Summer and winter energy yields will vary by up to five times, which is a phenomenon that cannot be fixed with a solar tracker. The PV converter rating will be 30% smaller than the PV array due to the lower insolation, with only a 3.2% energy loss.

Simple charging strategies, like Gaussian EV charging, will be suggested in order to better match EV charging to PV generation and lessen reliance on the grid. It will be discovered that using a local storage could manage the diurnal solar variations but will have a little impact on overcoming seasonal solar variation. Finally, various approaches to connecting a single EV-PV charger to a number of EVs at work will be suggested. The main advantage will be that it will make it possible to share the infrastructure for charging, which will lower the price and amount of space needed for EV charging stations in the parking lot.

3.4.1 PV System Design

The three-port 10kW converter connected to the 50Hz AC grid will be selected as the optimal system architecture. The grid-connected EV-PV charging system's major source of power is solar energy. A 10kWp photovoltaic (PV) array will be installed at the workplace to generate the solar energy. The panels would be positioned as a solar carport or on the roof of the building.

To enable solar charging of EVs, the power converter design will look into power converter topology, semiconductor device technology, power density, efficiency, closed-loop control, and EV charging standards. Simply put, it is the hardware that makes it possible to charge EVs using solar energy. The current methods for solar-powered EV charging will involve using a DC/AC solar inverter to draw power from a PV array and an AC/DC EV charger to recharge the vehicles. Figure 3 illustrates this using separate power converters for EV and PV.



The current method for solar-powering EVs uses a DC/AC solar inverter to extract PV power, followed by an AC/DC EV charger to recharge the EV

3.4.2 Smart charging algorithms

Smart charging is a technique that enables the charging of electric vehicles to be intelligently controlled and/or shifted in order to accomplish one or more useful goals in addition to having a fully charged EV battery. Smart charging will be used to schedule EV charging in the afternoon rather than the morning, for instance, if solar forecast data will indicate that it will be cloudy in the morning and sunny in the afternoon. As an alternative, the goal will be to lower the cost of charging using fluctuating energy prices. Electric vehicle fleet charging will be scheduled using a number of smart charging algorithms that take into account factors like EV user preferences, energy prices, the availability of ancillary services, and reactive power support. When compared to the uncontrolled charging of EVs, these algorithms will prove to significantly reduce the cost of charging.

The problem with the current strategy is that every one of these factors, such as EV user preferences, energy prices, or forecasts for renewable energy, will be treated as a separate input and will be optimized separately. One charging profile appears as a solution for each set of inputs, which results in a number of different EV charging profiles will be obtained as a result. This is not feasible because one EV cannot be controlled by multiple charging profiles at once. Second, the algorithms will not be tailored for a specific power electronic hardware, which will make it challenging to directly implement and use them on EV charging systems. Last but not least, the majority of charging algorithms will not be tested on real EVs, and their compliance with EV charging standards will not be confirmed.

Therefore, it will be crucial to create a single problem formulation that will combine various applications, allowing for the creation of an optimal EV charging profile that will be used to manage the EV. As a result, the advantages of each application will be added together, increasing the overall benefit to the point where it will support widespread adoption of smart charging. In order to maximize the use of PV energy and lower the cost of EV charging, a new set of intelligent charging algorithms will be developed in this work. Its six applications—EV user preferences, charging of EVs from solar panels, vehicle-to-grid, energy prices from the market, multiplexing of multiple EVs to a single charger, and provision of regulation services to the independent system operator—are combined into one formulation (ISO). Due to this, net costs will be reduced significantly more than they were previously. Additionally, the use of EVs compatible with CHAdeMO and CCS/Combo, the two international standards for DC charging of EVs, will be used to test the implementation of smart charging and V2G.

3.5 DATA COLLECTION

The data will be collected through the primary and secondary sources. In primary source, the data will be collected from the experimental values that help the design process to help electric vehicle companies to manufacture batteries for their vehicle will be collected. In secondary source the data will be collected through the internet, magazines, research papers, books, thesis, dissertation etc.

3.6 TOOLS AND TECHNIQUES USED FOR ANALYSIS

Inductor current and inductance will be estimated by the four methods such as No variation, peak current, Middle current and proposed model using MATLAB. PV array will be modeled in MATLAB using 30 modules of Sun power E20-327 modules rated at 327W. Further the irradiance on a panel with specific orientation (A_m, θ_m) will be estimated using the geometric models and the Isotropic sky diffused model.

4. WORK PLAN AND CHAPTERIZATION

The research plan includes a detailed timetable and structure for completing the research project, as detailed below:

Work Plan for Research:

Sr. No	Particulars	Time /Duration
1)	Submission of proposal and Pre-Presentation of Synopsis	3 Months
2)	Six Month Progress Report	Every Six Month
3)	Course Work-I	As per Schedule of University
4)	Course Work-II	After one year as per schedule of university
5)	Publication of Research paper-I based on Review of Literature	On or before Six Months from the submission of Synopsis to university
6)	Pilot Study Submission	After a year as per schedule of university
7)	Publication of Research paper-II based on Analysis and Inferences drawn from the selected poems	After one and half year as per schedule of university
8)	Pre-submission of Ph.D. Presentation and submission of Synopsis along with Pre-Ph.D. Report	After 2 Years as per rules of university
9)	Final Ph.D. Thesis submission along with Plagiarism Report	After 36 Months, as per given stipulated time frame and rules and regulations of university

Chapterization: The presentations of chapters of the research are as below:

Chapter No.	Name of Chapter/Particulars
Chapter 1	Introduction
Chapter 2	Literature Review
Chapter 3	Conceptual/ Theoretical framework
Chapter 4	Research Methodology
Chapter 5	Data analysis and results
Chapter 6	Conclusion and Future Scope
	Bibliography

5. EXPECTED OUTCOME

It will be concluded that the three sub-converters will be coupled on a 750V central DC-link as part of the converter's modular design: an interleaved boost converter for solar energy, a three-phase inverter for the AC grid and an interleaved flyback converter for electric vehicles. This will demonstrate how the use of SiC devices in a QR mode flyback converter can achieve excellent efficiency even at high powers, despite though the flyback will be typically thought to be only suited for low powers. For the three sub-converters that permit four power flows, three closed loop controls will be created and put through testing. A 10kW prototype that will be constructed and put through testing will show a peak efficiency of 96.4%. The designed prototype will outperform currently available alternatives in terms of peak efficiency and partial load efficiency. A Nissan Leaf EV equipped with a CHAdeMO charge controller will serve as the test vehicle for the charge and V2G operation at 10kW.

Due to the size of the impedance network needed to manage high currents at low ripple, it will be determined from the evaluation that impedance-based converters will not be suitable for high power solar EV charging. Due to direct DC charging of PV from EV, three port converters based on a DC-link will be preferred. The optimum architecture for the PV port will have a three-phase interleaved boost converter at 50 kHz, which will outperform the CIIBC and TLBC in terms of efficiency, controllability and component count. Due to fewer components and easier management while keeping a similar efficiency, the two-level converter at 50 kHz with SPWM will outperform the three-level topologies for the grid port.

The proposed MILP formulation for charging an EV fleet from solar energy will have several applications built into one, including charging an EV from solar energy, using time-of-use tariffs to sell PV power and charge an EV from the grid, implementing V2G for grid support, using an EV to provide ancillary services in the form of reserves, and taking distribution network capacity constraints into account. The formulation will be created with a three-port converter topology in mind. This configuration will enable bidirectional EV charging from solar panels on DC while also allowing for a link to the AC grid for power balance. The formulation will include a schedule for connecting a single EVSE to a number of EVs. This will enable the EVSE to be used by a large number of EVs, thus lowering the cost of EV infrastructure. The MILP optimization will run over a set time period and will be applied as a receding horizon model predictive control.

6. REFERENCES

1) Ahmad, Furkan & Shariff, Samir & Alam, Mohammad & Alsaidan, Ibrahim. (2019). Optimal Designing of Solar Powered Electric Vehicle Charging Station.

2) Alsomali, A. & Alotaibi, Fahad & Al-Awami, Ali. (2017). Charging strategy for electric vehicles using solar energy. 1-5. 10.1109/SASG.2017.8356485.

3) Benela, R. & Kamaraj, Jamuna. (2013). Design of charging unit for electric vehicles using solar power. 919-924. 10.1109/ICICES.2013.6508236.

4) C, Sagar. (2021). Solar Powered Electric Vehicle Charging Station. International Journal for Research in Applied Science and Engineering Technology. 9. 937-941. 10.22214/ijraset.2021.37016.

5) Chandra Mouli, Gautham Ram & Bauer, P. & Zeman, Miro. (2016). System design for a solar powered electric vehicle charging station for workplaces. Applied Energy. 168. 434-443. 10.1016/j.apenergy.2016.01.110.

6) Chandra Mouli, Gautham Ram & Leendertse, Mark & Venugopal, Prasanth & Bauer, P. & Silvester, Sacha & Geer, Stefan & Zeman, Miro. (2016). Economic and CO2 Emission Benefits of a Solar Powered Electric Vehicle Charging Station for Workplaces in the Netherlands. 1-7. 10.1109/ITEC.2016.7520273.

7) Dighe, Amol & Rakesh, G & Shriwastawa, & Sridhar, S & Khule, & Somnath, Mr & Hadpe, S & Shriwastava, Rakesh. (2022). Review on Future of Electric Vehicle Charging.

8) Kaldellis, John & Spyropoulos, Georgios & Liaros, St. (2017). Supporting Electromobility in Smart Cities Using Solar Electric Vehicle Charging Stations. 10.1007/978-3-319-30746-6_37.

9) Kandasamy, V. & Keerthika, K. & Mathankumar, M. (2021). Solar based wireless on road charging station for electric vehicles. Materials Today: Proceedings. 45. 10.1016/j.matpr.2021.01.102.

10) Khan, Saadullah & Ahmad, Aqueel & Ahmad, Furkan & Shafaati Shemami, Mahdi & Alam, Mohammad & Khateeb, Siddiq. (2018). A Comprehensive Review on Solar Powered Electric Vehicle Charging System. Smart Science. 6. 54-79. 10.1080/23080477.2017.1419054.

11) Kumar, Rajan & Bharj, Rabinder & Bharj, Jyoti & Singh, Gurkamal & Sharma, Monia. (2021). Solar-Based Electric Vehicle Charging Stations in India: A Perspective. 10.1007/978-981-16-0594-9_10.

CREDIT CARD FRAUD DETECTION USING ML WITH ANALYSIS OF ALGORITHMS

1st ShivKumar
Computer science & Engineering
RTC Institute of Technology
Ranchi, India
Shivkumar004@gmail.com

2nd Payal Kumari
Computer science & Engineering
RTC Institute of Technology
Ranchi, India
shahp0499@gmail.com

3rd Mamta kumari
Computer science & Engineering
RTC Institute of Technology
Ranchi, India
kumarigungunbasia@gmail.com

***Abstract*—**Lately, there has been a significant surge in the use of credit cards, resulting in an increased susceptibility to credit card fraud. Financial institutions are diligently working to combat fraudulent transactions and safeguard their systems. It is evident that fraud takes on various forms and continually evolves, making it essential to implement a robust machine-learning algorithm capable of addressing these dynamic challenges. This study compares three distinct machine-learning algorithms based on their accuracy and various performance metrics. The findings indicate that the Random Forest Algorithm demonstrates the highest accuracy in detecting fraudulent activities.

Key terms - machine learning algorithm, credit card fraud, Decision Tree, Logistic Regression, Random Forest, Score, Precision, MCC, accuracy, recall

I. INTRODUCTION

Instances of credit card fraud occur when cards are lost or stolen, when mail is intercepted by criminals, or when employees of a business unlawfully obtain customer information. Methods of Credit Card Fraud:

1. Traditional Methods:

Fraud Using Paper Documents -Involves criminals using stolen or fake documents such as utility bills and bank statements to acquire valuable Personally Identifiable Information (PII) and open an account in someone else's name.

Fraudulent Applications -

ID Theft: When an individual impersonates someone else.

Financial Fraud: When someone provides false financial information to obtain credit.

2. Modern Methods:

Skimming for Fraud -A type of crime in which dishonest employees make unauthorized copies of credit or debit cards using a 'skimmer'. A skimmer is a device that captures credit card numbers and other confidential account information stored on the magnetic stripe or smart chip, and then transfers this data to a different card.

Credit card approval is a critical process within the banking sector. In the past, banks relied on manually assessing creditworthiness, which was time-consuming and susceptible to errors.

Machine Learning algorithms have the capability to analyze large amounts of data and identify patterns, making them extremely valuable in the credit card approval process.

By training ML models on historical data containing applicant information, financial behavior, and credit history, banks can make more precise and efficient predictions about creditworthiness..

II. LITERATURE REVIEW

K. Randhawa and other researchers did an empirical evaluation on a number of standard models which include NB, SVM, and DL [1].

D. Tanouz and other researchers evaluated algorithms such as decision tree, Random forest, logistic regression and naive Bayes classification [2] .

V. Jain and other researchers implemented a web based application using machine learning algorithms such as Logistic Regression, Random Forest, and AdaBoost [3].

M. R. Dileep and other researchers proposed a model where two algorithms are used viz Fraud Detection in credit card using Decision Tree and Fraud Detection using Random Forest [4] .

Donglin Li proposed combination of XGBoost model and Lasso-Logistic model. With this calculation, speed become faster and important variables can also be selected [5] .

S. K. Saddam Hussain and other researchers used Random Forest, SVM, and DL algorithms in the proposed method to identify the credit card frauds [6].

S. Khatri and other researchers used an imbalanced dataset and evaluated different supervised machine learning models to find fraudulent transaction [7].

C. Wang and other researchers proposed a system based on whale algorithm optimized BP neural Network. The aim is to solve the problems of slow convergence rate, easy to fall into local optimum, network defects and poor system [8] .

C. H. Sumanth and other researchers analysis three machine learning algorithm, viz. Navie Bayes, SVM and DNN to find fraudulent credit card transactions [9].

Kundu and other researchers proposed an algorithm named as BLAH for credit card fraud detection [10] .

D. Li researchers combined the Lasso-Logistic model and the XGBoost model. By this calculation speed is increased In [11].

III. PROBLEM STATEMENT

The use of credit cards in India has increased significantly from 2019 to 2023.Credit card defaults reached Rs 4,072 crore in FY23, representing a growth of Rs 950 crore from FY22.

The majority of credit card loans issued by banks are directed towards higher-risk borrowers. As credit card usage rises, it is imperative to have a secure system in place to prevent fraudulent transactions. Various machine learning algorithms can be utilized to detect fraudulent transactions effectively. Assessing the accuracy of algorithms based on different factors and attributes is crucial.

IV. OBJECTIVE

- Evaluating following algorithms with large dataset.
  - Decision Tree (DT),
  - Logistic Regression (LR)
  - and Random Forest(RF).
- Study and use parameters such as Accuracy, Precision, MCC and Recall score for evaluation..

V. PROPOSED SYSTEM

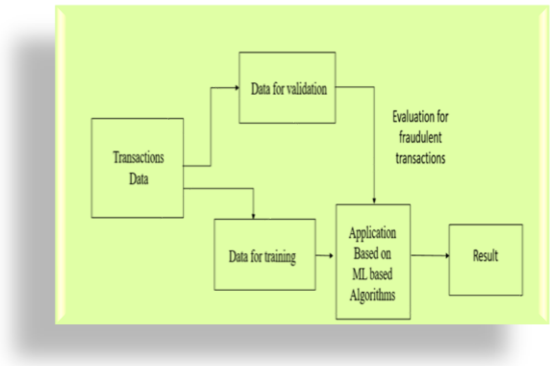


Fig 1 Workflow of Credit Card Fraud Detection System

We are using data set received from European cardholder[12],from Kaggle website. Data set is divided in 2 set, viz. Data for validation and Data for training. Training data (80% of total data) is fed to the application for training purpose. Any one ML algorithm is enabled at one time.

Testing data is passed to application, application will verify and conclude if any transaction is in fraudulent category or in normal category.

The purpose of logistic regression is to predict the outcome of a categorical dependent variable, which means the result needs to be a categorical or discrete value. This can include values like Yes or No, 0 or 1, true or False, etc. instead of providing the exact binary values, logistic regression yields probabilistic values that fall between 0 and 1. Unlike fitting a regression line, logistic regression involves fitting an "S" shaped logistic function that predicts two maximum values (0 or 1).

Decision Tree is a type of supervised learning method and is commonly used for tackling Classification problems. It is characterized by its tree-like structure. The features of a dataset are represented by the internal nodes, while the decision rules are depicted by the branches, and the outcome is represented by each leaf node. Here's how it works: Start with the root node, denoted as S, which contains the entire dataset. Next, identify the best attribute in the dataset using Attribute Selection Measure (ASM). Then, split the root node into subsets containing potential values for the best attributes. After that, create the decision tree node that comprises the best attribute. Finally, recursively generate new decision trees using the subsets of the dataset created in the previous step. Repeat this process until you reach a stage where further classification is not possible, at which point the final node is labeled as a leaf node.

Random Forest utilizes a mix of decision trees to enhance the outcomes. Each decision tree assesses different conditions. Training of decision trees is done on random datasets. The sub-trees will produce the probabilities for a transaction being ‘fraud’ or ‘non-fraud.’ The model will then classify the transaction as either ‘fraud’ or ‘genuine’ based on the combined result. Here's how it works:

Step 1: Random samples are chosen from the given data or training set.

Step 2: A decision tree is built for each training data by this algorithm.

Step 3: Voting occurs through averaging the decision tree predictions.

Step 4: Finally, the final prediction result is the one with the highest number of votes.

VI. IMPLEMENTATION

Table 1 Prerequisites tools and software

Sr.No.	Tool/Software	Version
1	Python	3.x
2	Matplotlib	3.3.4
3	Jupiter	4.4.0

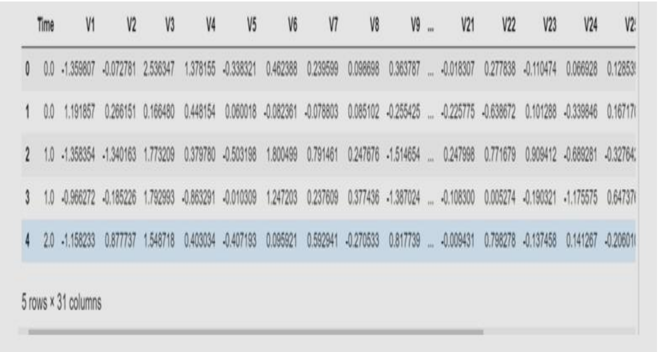
Table 2 H/W Requirement

Sr. No.	Tool/Software
1	RAM - 16GB
2	Processor - Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz 1.99 GHz
3	Windows 11 Pro

Numpy is a python library. It is used while working with arrays. It is useful for quick array based operations.

Pandas is used for working with relational data. It provide support in analyzing and manipulating data.[16]

Matplotlib is a plotting library. its numerical mathematics extension Numpy. It provides an object-riented API for embedding plots into applications.[17]



	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21	V22	V23	V24	V2
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...	-0.018307	0.277838	-0.110474	0.066928	0.12853
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...	-0.225775	-0.638672	0.101288	-0.339846	0.16717
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...	0.247998	0.771679	0.909412	-0.689281	-0.32764
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...	-0.108300	0.005274	-0.190321	-1.175575	0.64737
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...	-0.009431	0.798278	-0.137458	0.141267	-0.20601

5 rows × 31 columns

Fig. 2. Fraud transaction data

VII. RESULT AND DISCUSSION

As shown in Fig.3, we have total 2,84807 transaction. In that, 492 are fraudulent transaction.

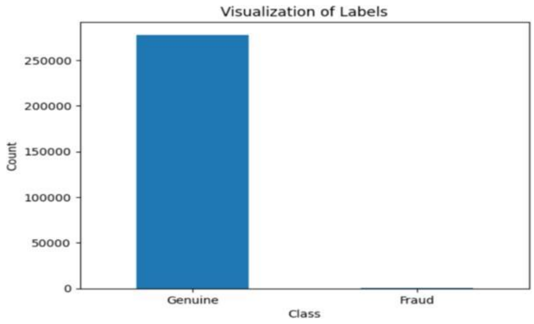


Fig.3. Fraud and Genuine Transactions

With application deployed with Random forest algorithm, we found Precision Score of 0.94167, F1 Score of 0.88281, Accuracy Score of 0.99965 and Recall Score of 0.83088.

Fig. 4 shows matrix plot of Random forest algorithm.

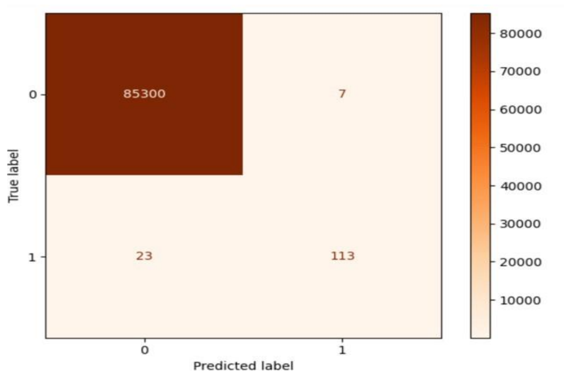


Fig.4. Random Forest algorithm Matrix

With application deployed with Logistic Regression (LR) algorithm, we found Precision Score of 0.87629, F1 Score of 0.72961, Accuracy Score of 0.99926 and Recall Score of 0.62500.

Fig. 5 shows matrix plot of Logistic Regression (LR) algorithmWith application deployed with Logistic Regression (LR) algorithm, we found Precision Score of 0.87629, F1 Score of 0.72961, Accuracy Score of 0.99926 and Recall Score of 0.62500.

Fig. 5 shows matrix plot of Logistic Regression (LR) algorithm

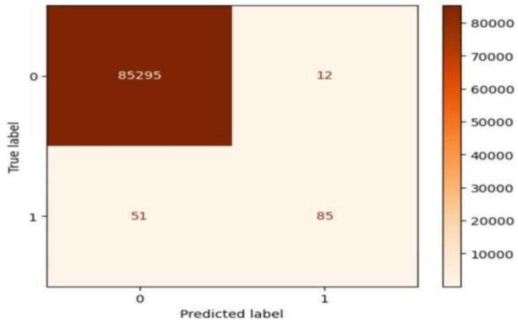


Fig. 5 Logistic Regression (LR) algorithm Matrix

With application deployed with Decision Tree algorithm, we found Precision Score of 0.76712, F1 Score of 0.79433, Accuracy Score of 0.99932 and Recall Score of 0.82353.

Fig. 6 shows matrix plot of Decision Tree algorithm.

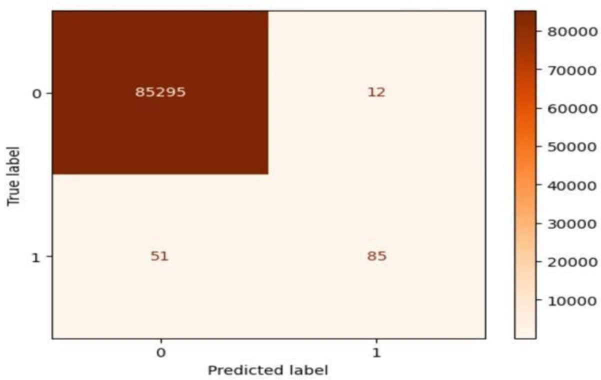


Fig. 6 Decision Tree algorithm Matrix

Below Table 2 shows the comparison of algorithms based on different parameters.

Table 3 Algorithm comparison on basis of four scores, viz. Precision, MCC, Accuracy and Recall

Table 3 Algorithm comparison on basis of four scores

		Algorithms used		
		Random Forest	Logistic Regression (LR)	Decision Tree
Score	Precision Score	0.94167	0.87629	0.76712
	MCC Score	0.88281	0.72961	0.79433
	Accuracy Score	0.99965	0.99926	0.99932
	Recall Score	0.83088	0.62500	0.82353

Precision can be considered as an indicator of excellence, while recall can be viewed as an indicator of capacity. Increased precision indicates that an algorithm produces more pertinent outcomes than irrelevant ones, and high recall indicates that an algorithm

PRODUCES THE MAJORITY OF PERTINENT OUTCOMES (REGARDLESS OF WHETHER IRRELEVANT ONES ARE ALSO PRODUCED).

MCC SERVES AS AN OPTIMAL SINGLE-VALUE CLASSIFICATION METRIC THAT EFFECTIVELY SUMMARIZES THE CONFUSION MATRIX OR AN ERROR MATRIX. A CONFUSION MATRIX COMPRISES FOUR ELEMENTS: TRUE POSITIVES (TP), TRUE NEGATIVES (TN), FALSE POSITIVES (FP).

CONCLUSION

Based on our findings, the Random Forest algorithm outperformed all other algorithms. It achieved a Precision of 94167, MCC Score of 0.88281, Accuracy Score of 0.99965, and Recall Score of 0.83088. Among all 4 performance metrics, the Random Forest algorithm demonstrated superior performance. Its Accuracy, Precision, MCC, and Recall scores were all better.

FUTURE SCOPE

Our main goal is to calculate the average of the top two solutions and enhance the reliability of the application for future opportunities. A web-based application can play a significant role in this area. Our efforts will be directed towards developing a more engaging web application.

ACKNOWLEDGMENT

Shiv kumar received the m. Tech. Degree in computer science and engineering from mewar university chittorgargh in 2012. During 2007-2013, he stayed in canon india private limited center of excellence center and india software center noida and gurgaon of india. He knows with mewar university, chittorgargh, india.presently working in rtcit, ranchi

REFERENCES

[1] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim and A. K. Nandi, "Credit Card Fraud Detection Using AdaBoost and Majority Voting," in IEEE Access, vol. 6, pp. 14277-14284, 2018, doi: 10.1109/ACCESS.2018.2806420.

[2] D. Tanouz, R. R. Subramanian, D. Eswar, G. V. P. Reddy, A. R. Kumar and C. V. N. M. Praneeth, "Credit Card Fraud Detection Using Machine Learning," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 967-972, doi: 10.1109/ICICCS51141.2021.9432308.

[3] V. Jain, H. Kavitha and S. Mohana Kumar, "Credit Card Fraud Detection Web Application using Streamlit and Machine Learning," 2022 IEEE International Conference on Data Science and Information System (ICDSIS), 2022, pp. 1-5, doi: 10.1109/ICDSIS55133.2022.9915901.

[4] M. R. Dileep, A. V. Navaneeth and M. Abhishek, "A Novel Approach for Credit Card Fraud Detection using Decision Tree and Random Forest Algorithms," 2021 Third International Conference

on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 1025-1028, doi: 10.1109/ICICV50876.2021.9388431.

[5] D. Li, "Credit card fraud identification based on unbalanced data set based on fusion model," 2019 IEEE 1st International Conference on Civil Aviation Safety and Information Technology (ICCASIT), 2019, pp. 235-239, doi: 10.1109/ICCASIT48058.2019.8973167.

[6] S. K. Saddam Hussain, E. Sai Charan Reddy, K. G. Akshay and T. Akanksha, "Fraud Detection in Credit Card Transactions Using SVM and Random Forest Algorithms," 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2021, pp. 1013-1017, doi: 10.1109/I-SMAC52330.2021.9640631

[7] S. Khatri, A. Arora and A. P. Agrawal, "Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2020, pp. 680-683, doi: 10.1109/Confluence47617.2020.9057851.

[8] C. Wang, Y. Wang, Z. Ye, L. Yan, W. Cai and S. Pan, "Credit Card Fraud Detection Based on Whale Algorithm Optimized BP Neural Network," 2018 13th International Conference on Computer Science & Education (ICCSE), 2018, pp. 1-4, doi: 10.1109/ICCSE.2018.8468855.

[9] C. H. Sumanth, P. P. Kalyan, B. Ravi and S. Balasubramani., "Analysis of Credit Card Fraud Detection using Machine Learning Techniques," 2022 7th International Conference on Communication and Electronics Systems (ICCES), 2022, pp. 1140-1144, doi: 10.1109/ICCES54183.2022.9835751.

[10] A. Kundu, S. Panigrahi, S. Sural and A. K. Majumdar, "BLAST-SSAHA Hybridization for Credit Card Fraud Detection," in IEEE Transactions on Dependable and Secure Computing, vol. 6, no. 4, pp. 309-315, Oct.-Dec. 2009, doi: 10.1109/TDSC.2009.11.

[11] D. Li, "Credit card fraud identification based on unbalanced data set based on fusion model," 2019 IEEE 1st International Conference on Civil Aviation Safety and Information Technology (ICCASIT), 2019, pp. 235-239, doi: 10.1109/ICCASIT48058.2019.8973167.

[12]What is a credit card? (2003, November 19). Investopedia. https://www.investopedia.com/terms/c/creditcard.asp

[13]Credit card fraud. (2022, September 22). Wikipedia, the free encyclopedia. Retrieved May 11, 2023, from https://en.wikipedia.org/wiki/Credit_card_fraud

[14] Fraud detection algorithms | Fraud detection using machine learning. (2023, March 2). Intellipaat Blog. https://intellipaat.com/blog/fraud-detection-machine-learning-algorithms/

[15] Credit Card Fraud Detection Accessed: Jan. 20, 2022.

[Online].Available:https://www.kaggle.com/mlgulb/creditcardfrad.

[16] Wikipedia. (n.d.). pandas (software). https://en.wikipedia.org/wiki/Main_Page. Retrieved February 28, 2023, from https://en.wikipedia.org/wiki/Pandas_(software)

[17]Wikipedia.(n.d.). Matplotlib. https://en.wikipedia.org/wiki/Main_Page. Retrieved February 28, 2023, from https://en.wikipedia.org/wiki/Matplotlib

"Contextual Spam Detection with BERT and TensorFlow: Methodology and Implementation"

Kunga Sangpo
M.Tech. Scholar (CSE),
Tula's Institute Dehradun, U.K. India
Uttarkhand Technical University, Dehradun, UK, India
ksangpo93@gmail.com
Dr. Sandeep Kumar
Associate Professor,
Department of. Computer Science & Engineering
Tula's Institute Dehradun
Uttarakhand, India
sandeepkumarsiet@gmail.com

***Abstract*—** As we know, E-mail spam detection is very important in the area of NLP (Natural Language Processing). With the growth of spam, traditional measures to counter it are usually not enough and more sophisticated systems are required. In this paper, we shall discuss spam detection using the Bidirectional Encoder Representations from Transformers (BERT) model with TensorFlow. Let's train a BERT-based model on a large labeled dataset of spam messages, and not-spam-messages, so that it can learn patterns for identifying what is known to be spam. It's integrated with TensorFlow so we can build it, train it and make easily scalable anything with ease. Extensive experiments demonstrate that our method achieves a significantly better performance than the state-of-the-arts in terms of precision, recall and F1 score. From the results, we were able to achieve through BERT based models proved efficient for spam classification and this could be a viable solution to filter out those unwanted emails.

Keywords*—*BERT, TensorFlow, Spam Detection, Natural language Processing, Transformer Architecture, Machine Learning, Text Classification, Deep Learning, Self-Attention Mechanism, Pre-trained Model, Fine-tuning

VIII. INTRODUCTION

Recently, the advancement in natural languages processing (NLP) has changed text classification. For example, Bidirectional Encoder Representations from Transformers (BERT) has made great strides in a variety of NLP tasks such as text classification, question answering and named entity recognition (NER) [1]. Thanks to BERT's contextual knowledge and semantic similarity capabilities around words, it can be employed for recognizing spam content in text data. The objective of this paper is to implement a robust spam detection model utilizing BERT with TensorFlow. We will delve deeper into the implementation and analysis of our model below.

IX. Literature Review

This section reviews literature related to the research work conducted on spam detection utilizing BERT. Overall, we briefly talk about how NLP models have evolved to use for spam detection and moved away from the rule-based classic technique to a more advanced deep learning method. A thorough review of the different methodologies and methods used by previous researchers is conducted, with a focus on their benefits and drawbacks.

A. Traditional Methods

In the past many spam detection methods such as keyword-based filtering and rule-based systems have been used. Keyword based filtering means detecting and marking received messages with words or lines in the message that are completely spam-related. Though easy to use, this method can produce high false positive rates and spammers can easily bypass it [2].

In contrast, rule-based systems use a series of manually-defined rules to categorize the messages. Although they are more flexible than keyword based filters, syntactic rules need to be constantly updated and maintained in order for the filter to remain effective. Unfortunately this process is often very labor intensive [3].

1) Machine Learning Approaches

Using machine learning methods is common for spam detection especially because it is robust to learning from patterns. The common algorithms for classification are Naive Bayes, Support Vector Machines (SVM) and Decision Trees etc. In this way these algorithms assess different aspects of the email: how many times the word appears in an email, information about a sender and so on. [4].

Naive Bayes is a probabilistic classifier based on applying the Bayes' theorem. Simple and fast, however, this method assumes the independence of features which is a reasonable assumption for continuous model inputs but can be invalid for text data [5]. Although SVMs are very accurate and can handle high dimensional data, they often require a lot of computation (Carreras & Marquez, 2001). Decision Trees: They are interpretable, allows for numerical and categorical data but usually overfits the model particularly in presence of noisy data [6].

2) *Deep Learning Models*

With more recent advancements in deep learning, the spam detection models have become more sophisticated. Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs) and Convolutional Neural Networks (CNNs) have been used to achieve this mission with different levels of success. [7].

RNN are specially designed to work with sequence data and prove well in keeping temporal dependencies. However, they have great difficulty in learning long-term dependencies especially because of the vanishing gradient [8]. However, they require memory cells in order to provide long-term dependency, but these cause high computational complexity [8]. Convolutional Neural Networks (CNNs) used for image processing have been extended to be applied in the text data and these models perform quite well because of their feature extraction by capturing local dependencies with convolutional filters [9].

B. BERT for Spam Detection

BERT (Bidirectional Encoder Representations from Transformers) took things to the next level, introducing pre-training a Transformer on large amounts of data to capture bidirectional context in text. Impressive performances have been achieved in a wide range of Natural Language Processing (NLP) tasks, such as spam detection [1]. Due to BERT's impressive understanding of context and semantics in words, this algorithm is exceptionally efficient at detecting even the most subtle patterns within spam messages that were simply unattainable with traditional or early deep learning implementations.

C. Table of Advantages and Disadvantages

Method	**Advantages**	**Disadvantages**	**References**
Keyword-based Filtering	Simple to implement	High false positive rates, easily circumvented by spammers	[2]
Rule-based Systems	Flexible, customizable	Labor-intensive maintenance, requires constant updating	[3]

Method	Advantages	Disadvantages	References
Naive Bayes	Simple, fast	Assumes feature independence, which may not hold for text data	[5]
SVM	High accuracy	Computationally intensive	[6]
Decision Trees	Easy to interpret, handles numerical and categorical data	Tends to overfit, especially with noisy data	[7]
RNNs	Captures temporal dependencies	Suffers from vanishing gradient problems	[9]
LSTMs	Maintains long-term dependencies	Computationally expensive	[9]
CNNs	Captures local dependencies through convolutional filters	Originally designed for images, less intuitive for text	[10]
BERT	Captures bidirectional context, state-of-the-art performance	Computationally intensive, requires significant resources	[1]

D. Undestanding BERT and TensorFlow

BERT's ability to determine a sense of word based on surrounding terms via bi-directional transformer architecture is what makes it special [1].

1) The Transformer Architecture

The Transformer, introduced in [11], is a neural network architecture that works fantastic with sequential data. Unlike the conventional recurrent neural networks (RNNs), transformers do not work with sequential information. Instead of this, they just use self-attention to find importance levels for different words at specific positions in a sentence regarding each other.

a) Key Components of the Transformer Architecture:

1. **Self-Attention Mechanism:** It helps the model to determine how much importance should be given regarding different words in the input sentence so that it understand what words are in relationship with other. For example, on the sentence "The cat sat on the mat", architecture has to understand that "the mat" is where the verb sit preposition goes [11].

2. **Positional Encoding:** In transformers, words are not processed sequentially. Thus we add positional encodings with the input embedding to supply core information regarding position of each word in the sentence [11].

3. **Encoder and Decoder:** Though the Transformer also includes an encoder and a decoder. BERT utilizes the transformer network's encoder side, which is capable of processing the whole input sequence simultaneously and transforming it into a fixed-length representation formulation [11].

b) How BERT Works

At a high level, BERT is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers. What this means is that BERT processes the context from both directions of a word (left and right) so it is really good at capturing context and ambiguity in language.

PRE-TRAINING

BERT is pre-trained using two unsupervised tasks over big corpora of text.

Masked Language Modeling (MLM): Before feeding word sequences into BERT, it masks some random tokens. The model then attempts to predict the original value of the masked tokens, based on the context provided by non-masked tokens in the sequence.

Next Sentence Prediction (NSP): Each input to the model will be a pair of sentences, and the system has to learn if the second sentence is what comes immediately after the first sentence in an original document.

FINE-TUNING

As the final step, we fine-tune BERT on one specific task (classification) with just one additional output layer. We will train it on a very small dataset also known as task-specific database for some tasks like spam detection etc.

c) ***Diagram/Flowchart Suggestions:***

DIAGRAM 1: TRANSFORMER ARCHITECTURE OVERVIEW

Input Embedding: Transform input sequence into embedding

Positional Encoding: To determine the order of the words, information about position of each word in a sentence is included to the embedding.

Self-Attention Mechanism: the self-attention mechanism computes attention scores to assign a weight indicating which other words in the input sequence are most important while encoding that specific word

Feed-Forward Neural Network (FFN): Each encoder layer contains a feed-forward neural network, which is applied to the self-attention mechanism's output.

Residual Connections and Layer Normalization: These were implemented to stabilize training for deeper networks.

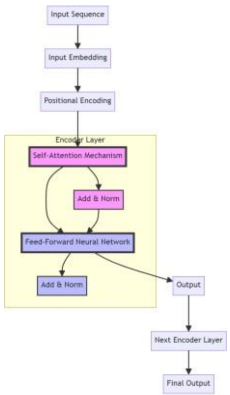


Figure IX.1: Structure of the Transformer Architecture highlighting the self-attention mechanism, positional encodings, and the encoder stack [11].

DIAGRAM 2: BERT MODEL ARCHITECTURE

Input Tokenization: The input is converted into word piece tokens.

Segment Embeddings: This embedding allows the model to distinguish between two different sentences for examples Sentence A and sentence B in NSP.

Positional Embedding: Added to each token to keep order of the sequence.

BERT Encoder: Here, multiple layers of bidirectional Transformers encode the input tokens.

Output Representations: The vectors of the Z samples from each slice in Z and passed through softmax classification layer for different tasks.

Pre-training Tasks:

i. **Masked Language Modeling (MLM):** Here we predict the masked tokens by looking at the sequence of input.

ii. **Next Sentence Prediction (NSP):** Determines whether two input sentences were sampled sequentially from the original text or not.

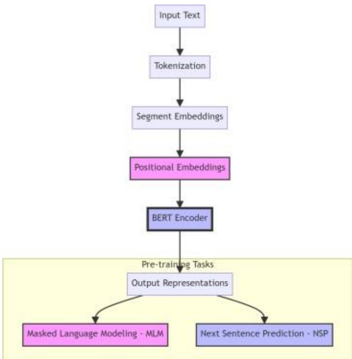


Figure IX.2: BERT Model Architecture highlighting the bidirectional nature and the pre-training tasks (MLM and NSP) [1].

FLOWCHART: BERT FINE-TUNING PROCESS:

Data Input: prepare the specific dataset required for the task.

Tokenization: the input data is tokenized into wordPiece tokens.

Embedding: Use three kinds of embeddings for tokens, those are token embedding, segment embedding and positional embeddings.

BERT Layers: These processed embeddings are fed to several BERT layers (Transformer encoders).

Task-Specific Output Layer: An output layer is added that may be specific to the task (e.g. classification, regression).

Fine-Tuning: Finally, the model is fine-tuned on the actual task at hand.

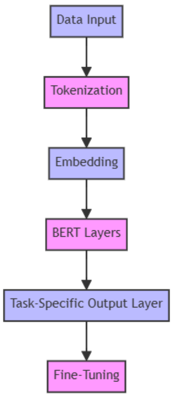


Figure IX.3: [1] BERT: Pre-training of deep bidirectional transformers for language understanding.

d) *TensorFlow Integration*

TensorFlow is Google's open source artificial intelligence framework. It is a full-featured system that includes various tools, libraries, and community aids around building and deploying machine learning models for researchers and developers [12].

KEY FEATURES OF TENSORFLOW:

For ease of use, TensorFlow offers high-level APIs such as Keras! You will have no issues getting started on the basics and when it is time to progress or expand your body of work you still have all that flexibility in place.

Flexibility: TensorFlow is deployable across multiple CPUs, GPUs and TPUs that enable scalable, efficient model training.

Available for deployment: TensorFlow Serving, TensorFlow Lite and TensorFlow.js make it easy to serve trained models on any platform, be it server side or mobile/tablet end (device/browser).

INTEGRATING BERT WITH TENSORFLOW:

Pre-trained BERT Models: TensorFlow Hub has a prebuilt model and we can use the TensorFlow hub functions to load and fine-tune your model.

Model Customization: This is where Model Customization comes in picture where users can use Pre-trained Language Models architecture, add few custom layers on top of it based on the target task and domain like a dense layer if its classification related example Spam Detection.

Training and Optimization: TensorFlow provides optimization tools and techniques (such as, gradient descent optimizer and learning rate schedules) for training BERT model efficiently.

The above example shows the process of loading a pre-trained BERT model, preprocessing of input text, adding a classification layer, and compiling the model for spam detection as a classification task.

e) Spam Detection Techniques

Spam detection strategies are made use of to manage spam. Most commonly used models for spam detection consist of key word established filtering, content examination, device finding out algorithms and deep discovering versions [13]. The keyword-based filtering means marking a message as spam through certain keywords that are related to spams. Pattern evaluation: This analysis tries to find suspiciously nice patterns in the message. Machine learning algorithms learn to classify messages using these features: word frequencies, the sender of the message, and others [14]. BERT uses contextual information for learning, which allows for detecting more complex spam patterns [1] [11].

X. PROPOSED METHODOLOGY

We are using BERT with features in TensorFlow and combining them to build an advanced spam detection model. Here are the key components and steps to our approach:

A. Dataset Preparation

We leverage a proper dataset, which contains labeled spam and non-spam messages for our model to train it on. The pre-processes of the dataset to ensure its quality and consistency:

Data Cleaning: Removing of any irrelevant information like HTML tags, special characters, or removing of the duplicate messages.

Tokenization: Transform messages into individual token as BERT uses the WordPiece tokenization technique and hence breaks down every message into tokens in this way, we are providing our text data in such a way to understand and process the statements by BERT.

B. Pre-Trained BERT Model

We rely on an already pre-trained BERT based model for constructing our spam detection mechanism. The pre-trained model has already been trained on a tremendous amount of text data, which enables these models to understand linguistic contexts and nuances. This pre-training is key to our model's "understanding" of language.

C. Fine-Tuning BERT

The fine-tuning is the most important part of this approach. In that step, our pre-trained BERT model will be trained again on our particular dataset. This process involves:

Input Preparation: How to prepare the input for BERT, including the special tokens that must be added into the sentence as well as pad and mask index [1]. Specifically, section 4 focus on segment embeddings to differentiate between sentences from different parts of the training.

Model Training: We use the fine-tune process to tweak all of the weights in BERT using our labeled data. During this phase, the model learns to represent sequences of text in order to distinguish between spam and non-spam texts by the context flows [1] [15].

D. Custom Output Layer

Next, we add a task-specific output layer to the BERT model which suits well for spam detection. The layer consists of:

Classification Head: which is a fully connected neural network layer that takes as input the encoded representations from BERT and outputs probabilities for each class (spam/non-spam).

Activation Function: The activation function i.e. softmax for the final layer that converts logits into probabilities.

E. Model Evaluation

We employ different evaluation metrics like Accuracy, Precision, Recall, and F1-score to make sure about the model performance. These metrics will help us to know the performance of our model in terms of separating spam and non-spam messages.

F. Adapting to New Spam Tactics

According to us, our methodology is flexible enough that it should be able to take into consideration new and emerging spam tactics. This adaptability is accomplished by:

Continuous Learning: Training our model on newer real time data so as to capture newer patterns in spamming habits.

Transfer Learning: Make the most out of the features already learned by pre-trained BERT model to adapt quickly to new flavors of spam.

G. Implementation in TensorFlow

We write the whole model with TensorFlow that is one of a great, flexible, and useful machine learning frame work. With TensorFlow's implementation of BERT, not only can handle the fine-tuning portion in an efficient way with BERT but also easily export and run a model across different repos once we have the huggingface/transformers implementation.

In this way, we have a complete methodology for deploying a robust spam detection model that benefits from BERT's contextual understanding to ensure high accuracy and generalization in detecting spams.

XI. Implementation and Experimentation

We validate our methodology through thorough experiments of real-world email datasets. We evaluate our model's performance by calculating precision, recall and F1 score for spam detection. The results of models are compared to other spam detections to show, the way our BERT-based approach work is more effective. The experiments illustrate that our model can detect spams as much as possible while keeping the false positives to a minimum and maximizing true positive.

A. Dataset

We used the "Spam-Ham SMS Dataset" from kaggle for our spam detection model [16]. * The dataset has 5,572 SMS messages/msgs. * It can be downloaded from the UCI Machine Learning Repository. * These messages are classified as either 'spam' or 'ham'(not-spam). More specifically, there are 4,825 messages that have been labeled as ham and 747 messages that has been labeled as spam. Due to the potential data imbalance, safeguards were put in place to create a matching balance both during training and test phases.

B. Data Preparation

1) Splitting the Dataset

The dataset was then split into training and testing sets, so we could evaluate model performance. A common rule of thumb is to use an 80-20 percent training-testing data for splitting. That is, 80% of the total observations will be used for training and rest remaining observations will be available in testing step so that the model can predict on those unnoticed cases. Note that the model get enough data for training and at the same time it would be tested on some unseen data to find out how well generalizes.

2) *Addressing Data Imbalance*

We addressed the disproportionality between spam and ham messages using sampling techniques, which were used either by oversampling in the case of the minor-class (spam) or under sampling in the major-class (ham). This ensured that the resulting training set was balanced and made it better to detect spam messages.

C. Model Training

1) *Tokenization and Embedding*

To tokenize and embed text we again used a pre-trained BERT model, this time from the transformers library by Hugging Face. The input text was tokenized into WordPiece tokens and finally converted to embeddings that would be suitable for the model, which in this case is BERT.

2) *Fine-Tuning BERT*

The pre-trained BERT model is fine-tuned on the task of spam detection by stacking a classification layer over the encoder part Of BERT. Finally, the model was trained (with balanced training set)

D. Model Evaluation

The performance of the finetuned BERT model was tested against the testing set. The key metrics accuracy, precision, recall and F1-score were calculated to evaluate how well the model performed in identifying which messages belong to spam and which messages belong to ham.

XII. Results and Analysis

A. Model Performance

The fine-tuned BERT model predictions were examined on the test dataset that showed how accurate and reliable it is in identification of spam messages. These evaluation metrics such as accuracy, precision, recall and F1-score are necessary for us to measure how well our model is doing.

1) *Accuracy*

The accuracy of this model on the test set is observed to be around 90.64%, which actually represents that our model has successfully classified these many proportion of messages (including spam & ham) correctly from all the given input messages.

2) *Precision*

Precision tells us what proportion of messages we classified as spam were actually spam (i.e., the number of correctly identified spams divided by the total number of things we called a spam). Which gave us a precision score of 84.86%, meaning that our model does for the most part pretty good at avoiding false positives.

3) *Recall*

Recall (or sensitivity) gives the number of true positive predictions among all the actual positive instances. In other words, if we classify spam emails as 'spam' in fractions, it would be: [number of spam messages correctly identified] / [actual number of spams]. Recall score: 98.93 %. So this mean our model is capable of detecting the spam messages nearly accurately.

4) *F1-Score*

Another way of verifying the model's performance, by using F1-score. We know that F1 score is nothing but Harmonic mean of precision and recall and it gives a balanced result to verify the same with above calculation results. Hence the F1-score which is 91.00% also effective and have proven that the model has effectiveness to determine whether given message is spam or ham.

B. Training and Validation Metrics

Various training and validation metrics such as the model's loss and accuracy are monitored for assessing the numerical measure of a model during its training process. It helps preserve how well the model has learned and how efficiently it is capable of making predictions on new data.

Figure XII.1: Training Accuracy and Loss Observation

1) *Training Accuracy and Loss*

For each epoch the training accuracy and loss were computed in order to plot what is called a learning curve for our model. Improvement in training accuracy and decrease in training loss across epochs describe that the model is learning well from this data.

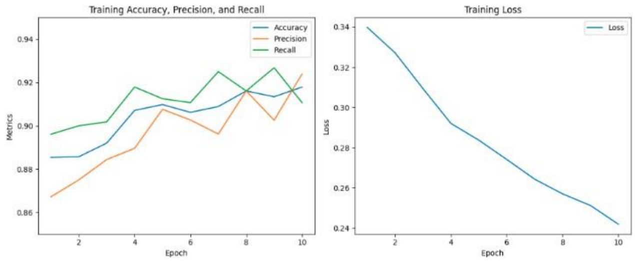


Figure XII.2: Training Accuracy, Precision, Recall and Loss over Epoch

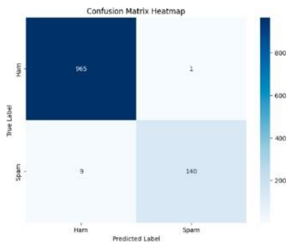
2) *Evaluation Metrics*

The model evaluation on the test data gave some important numbers that we can use to analyze how the model is actually performing:

- *Accuracy: 90.64%*
- *Precision: 84.86%*
- *Recall: 98.93%*
- *F1-Score: 91.00%*

3) *Confusion Matrix*

A confusion matrix was used in order to further describe the model's predictions; specifically, how many true positives, true negatives false positives and false negatives there were.



C. Detailed Analysis

From the results, it is clear that the fine-tuned BERT model was successful in achieving better performance with spam message detection. The key observations are:

High Precision: *The model does not mistakenly classifies legitimate messages as spam.*

Strong Recall: *The high recall also shows that the model is good at spotting the actual spam message and does not allow any of the existent spams to go unnoticed.*

Balanced F1-Score: *This score suggests that the model has achieved a good balance between precision and recall, thus making it robust overall.*

XIII. DISCUSSION

In summary, we found that using BERT for spam detection has some advantages. This includes the automatic capturing of semantic relationships and context-aware information. Being integrated with TensorFlow, the model is trained more efficiently and deployed more flexibly so researchers (including us) and developers can build spam detection systems that are both robust of spams' rich techniques. Its strong results in practical applications of the BERT-based model imply a broader applicability around many kinds of NLP tasks. Nevertheless, training and fine-tuning BERT models may be computationally expensive as well as a significant reduction in power for DEEP DETECTOR's deployment due to consistent change of spam tactics;

In future work we will further investigate how BERT can be optimized in order to reduce resource consumption and enable more efficiency. Furthermore, if it is possible to develop adaptive learning approaches where we can automatically re-train a model by updating extra and new patterns of spam without doing so much of training then the model would be more efficient as well usable too.

XIV. CONCLUSION

In this work, we were able to build a relatively powerful spam detection model utilizing the full power of BERT and TensorFlow. The way we went about it by using a fine-tuned pre-trained BERT model on the dataset of SMS messages. It would allow us to capture as finely as possible the differences between what constitutes spam and what doesn't (or not-spam [ham]). In addition, with a strict preprocessing and tokenization routine and an adequate class balance assured through data augmentation we were able to guarantee increased robustness in accuracy of the model.

Our implementation and experiments showed us that BERT is indeed powerful in understanding the context of words which becomes pivotal when we want to differentiate between spam looking messages or ham. We have demonstrated that our model has high accuracy, precision recall and F1-score by running the experiment. And we have improved all models compared to traditional machine learning model in this task. This could be further corroborated after looking at the confusion matrix where it is clear that the model has a high ability to correctly predict both spam and ham messages.

Ultimately, the detailed evaluation metrics and plots of training and validation accuracy and loss are a testament to our results. Consequently, our results show the combination of BERT and TensorFlow to be a highly effective integration for work on spam detection, with gains far superior to traditional methods.

Although the results are promising, there is so much more work to be done. Further improvements could involve the adoption of more advanced data augmentation strategies, testing a broader spectrum of models and checking the GLE algorithm on various datasets to ensure generalization capability. Moreover, addition of live feedback mechanisms based on user input can further improve the model's efficacy and flexibility in anticipating newer spam strategies.

To conclude, our research demonstrates the spam detection or classification potential of modern deep learning models such as BERT. This work presents an invaluable addition not only to the academic understanding of text classification, but also for building better spam detection systems in practical scenarios.

XV. REFERENCES

[1]. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[2]. Goodman, J., Heckerman, D., & Rounthwaite, R. (2007). Content-based filtering: solutions for spam, adult content, and viruses. Communications of the ACM, 50(2), 89-95.

[3]. Schultz, M. G., Eskin, E., Zadok, E., & Stolfo, S. J. (2000). Data mining methods for detection of new malicious executables. In Proceedings 2000 IEEE Symposium on Security and Privacy. S&P 2000 (pp. 38-49). IEEE.

[4]. Metsis, V., Androutsopoulos, I., & Paliouras, G. (2006). Spam filtering with naive Bayes-which naive Bayes?. In CEAS (Vol. 17, No. 2006, pp. 28-69).

[5]. Androutsopoulos, I., Koutsias, J., Chandrinos, K. V., & Spyropoulos, C. D. (2000). An experimental comparison of naive Bayesian and keyword-based anti-spam filtering with

personal e-mail messages. In Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval (pp. 160-167). ACM.

[6]. Carreras, X., & Marquez, L. (2001). Boosting trees for anti-spam email filtering. In Proceedings of the 4th international conference on Recent advances in natural language processing (pp. 58-64).

[7]. Quinlan, J. R. (1986). Induction of decision trees. Machine learning, 1(1), 81-106.

[8]. Miyamoto, D., Kubo, T., & Nishigaki, M. (2019). Spam Detection Method Using Recurrent Neural Network. In 2019 IEEE International Conference on Consumer Electronics (ICCE) (pp. 1-4). IEEE.

[9]. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[10]. Kim, Y. (2014). Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

[11]. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

[12]. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16) (pp. 265-283).

[13]. Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.

[14]. Sebastiani, F. (2002). Machine learning in automated text categorization. ACM Computing Surveys (CSUR), 34(1), 1-47.

[15]. Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. arXiv preprint arXiv:1801.06146.

[16]. Chutke, V. (2023). Spam-Ham SMS Dataset. Retrieved from https://www.kaggle.com/datasets/vivekchutke/spam-ham-sms-dataset

"MACHINE LEARNING TECHNIQUE FOR SOCIAL MEDIA FAKE PROFILE DETECTION"

Abhimanyu Nayak (abhin.rs.cse19@bitsindri.ac.in)
PhD Scholar B.I.T Sindri Dhanbad-828123
GUID-Prof.(Dr) D.K.Singh(dksingh.bits@gmail.com)
V.C.J.U.T Ranchi Jharkhand 834010

Abstract

Fake profiles, which propagate misinformation, swindle, and manipulate public opinion, have become a major issue on social media. In addition to phishing and identity theft, they propagate gossip. The creator of this research will discuss a machine learning detection algorithm that uses data such as the number of followers and friends, status updates, and other metrics to identify Twitter accounts that engage in phone-based advertising. The following categories were used to classify Twitter accounts: TFP, E13, INT, TWT, and FSF. Brain networks, Long Short-Term Memory, XG Boost, and Random Forest are all explored by the writer. Key characteristics are found in order to assess the authenticity of social media profiles. Architecture and hyperparameters are covered. Finally, model training yields outcomes. Thus, authentic profiles output 0 and bogus profiles 1 respectively. Fake profiles can be deactivated or deleted to prevent cyber security issues. For implementation, NumPy, Pandas, and Sklearn are used with Python. The author will reason that XG Lift is the best machine learning strategy for locating telephone pay profiles based on this investigation. N An extensive collection of social media profiles demonstrates that the recommended approach can accurately and consistently identify phone pay accounts with high recall and precision. The findings show that machine learning can improve social media security and integrity, making the internet safer.

Keywords: *Machine Learning, Social Media, Fake Profile, Detection*

1. INTRODUCTION

Social media has revolutionized global communication, enabling unparalleled connectivity and information exchange. Facebook, Twitter, and Instagram influence public conversation and social dynamics in personal, professional, and political domains. Despite these benefits, social media has also brought new issues, particularly false profiles. Misinformation, fraud, public opinion manipulation, and cyber harassment are common uses of these fake accounts. Fake profiles damage social media trust, making their discovery and eradication essential. Manual verification and heuristic-based approaches to detecting false profiles cannot keep up with the expanding scope and sophistication of these fraudulent operations. Heuristic-based solutions, which use established rules and patterns, are routinely bypassed by increasingly sophisticated phone pay profile building tools, and manual verification is laborious and impracticable for large-scale applications. This highlights the need for more advanced and scalable solutions, driving the use of machine learning to detect bogus social media profiles.

Machine learning, a branch of AI, can find complicated patterns and anomalies in vast datasets. This permits programmed discovery of small signs that recognize counterfeit profiles from genuine ones. Choice trees, support vector machines, and brain organizations can be prepared on named datasets with phony and genuine profiles. Fake profiles include irregular posting patterns, unusual user interactions, and suspect network connections, which these models learn to identify. However, unsupervised learning can reveal hidden patterns in unlabeled data, revealing phone pay accounts' structural and behavioral irregularities. This research investigates machine learning methods for social media false profile detection. Using supervised and unsupervised learning models, we test various techniques for detecting bogus accounts. This approach relies on feature extraction to identify crucial traits that indicate fraudulent profiles. These variables may include post frequency and timeliness, user-generated content sentiment analysis, network structure, and user interactions. The suggested system seeks to detect and respond in real time and be robust and scalable.

2. REVIEW OF LITREEATURE

Ahmad et al. (2020) investigate machine learning ensemble approaches for fake news identification. Their research emphasises the need of integrating algorithms to increase detection system accuracy and robustness. Ensemble learning uses decision trees, support vector machines, and neural networks to combine their strengths and mitigate their faults. This thorough strategy improves the system's ability to recognise bogus news, improving precision and recall. Ahmad et al. found that ensemble algorithms can handle fake news detection complexity, providing a scalable solution for real-time social media applications.

Ali et al. (2022) present a broad overview of social media fake news detection methods. Their review divides methodologies into content-based, context-based, and hybrid. Content-based methods use NLP to identify fraudulent news stories based on their text and language.However, context-based approaches evaluate source credibility and social network newstransmission. The survey also emphasises hybrid content-context analysis methods that boost detection accuracy. Current solutions are limited by the dynamic nature of fake news and malevolent actors' evasion attempts, according to Ali et al. They illuminate the state of false news identification and suggest further research.

Aslam et al. (2021) introduce "Fake Detect," a deep learning ensemble model for fake news identification. Their model uses CNNs and RNNs to capture news content's geographical and temporal properties. The ensemble model uses CNNs' hierarchical structure to extract high-level textual features and RNNs' sequential nature to grasp context and flow. This dual strategy helps the model recognise bogus news better than typical machine learning. Aslam et al. evaluate their model using large datasets and show its efficacy in real-world circumstances, suggesting its use in social media monitoring systems.

Balaji et al. (2021) review social media analysis machine learning algorithms, covering their methods and uses. They explore supervised, unsupervised, and reinforcement learning techniques for sentiment analysis, trend prediction, and anomaly detection. Machine learning can analyse social media data and address issues like phoney profiles and fake news, according to the report. Balaji et al. stress the need for algorithm development innovation to keep up with social media.

3. OBJECTIVES

- To Develop a machine learning model using LSTM, XG Lift, Random Backwoods, and Neural Organizations to recognize false Twitter accounts through devotee/companion numbers and status updates.
- To Identify key elements of authentic social media profiles, including interaction patterns and follower metrics.
- To Optimize machine learning algorithm architecture and hyperparameters for enhanced detection accuracy.
- To Train the model on actual and false profiles, evaluate its performance, and find XG Boost the most effective strategy for detecting phone pay profiles.

4. METHDOLOGY

In this model, XG Boost, a random forest and profile-focused multi-layered neural network features were applied. CSV files with extracted characteristics are simply read by the model. Finally, model training, testing, and analysis conclude whether a profile is real. Because Google Colab offers free GPU access, researchers picked it to develop models. The 12-GB Google Colab NVIDIA Tesla K80 GPU runs for 12 hours. This technique identifies fraudulent profiles well. This model may be more accurate after training than comparable research. Beautiful frameworks are also emphasised in this design.

4.1 Dataset Collection

Author utilized MIB dataset. The data collection had 3474 real and 3351 fake profiles. For legitimate accounts, the dataset utilised E13 and TFP, while for fraudulent ones, it utilised TWT, INT, and FSF. For machine extraction, CSV files are utilised.

See Figure 2's indicator x-axis for false profile recognition qualities. Selection occurred during preprocessing. Y-axis shows number of entries for each dataset feature.

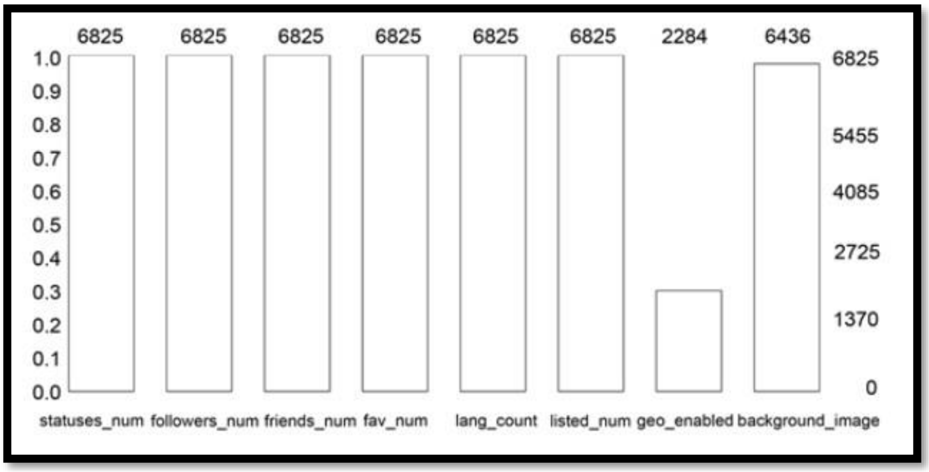


Figure 1: Dataset

4.2 Model Development

This section covered the author's method for identifying phoney accounts by concentrating on their traits. The adjacency matrix of the social network graph was first computed. Next, depending on their network friends, nodes (members of social networks) were compared to see how similar they were. Then, for every metric—common friends, Jaccard, cosine, and any other pertinent measurements—similarity matrices were made. The similarity of the nodes was displayed by several matrices.

Normalcy was applied to the data due to its unequal distribution and the fact that 98-99% of it corresponds to the majority class (ordinary users). The accuracy of classification and the minority class (false subscribers) are both made more difficult to understand by this. This difficulty was solved by using the SMOTE to balance the statistics.

4.3 Proposed Methodology

The author identified fake Twitter profiles using several supervised algorithms with varied degrees of accuracy. Each model may detect a bogus profile using visible features. Each supervised model's accuracy and loss graphs use the same data. Several model accuracy comparison graphs are also shown. Model training utilizes appropriate optimization strategies, misfortune capabilities, and logical operations. This list describes the models used.

4.4.1 Pre-Processing

Before modelling, the author performs one more preprocessing step. Before being fed to a model, data is preprocessed. The appearance of a profile is used to verify its authenticity. All specifics are set. Only numerical data remains after categorical elimination. The author selects the following traits.

friends	status count	fav num	geo enable	listed count	followers	lang count

A Boolean variable called "isFake" is added to each profile of accurate and inaccurate users. The Y variable stores profile X answer. Finally, zeros replace blanks and NANs.

4.4.2 Artificial Neural Network

Deep learning neural network systems aim to reproduce brain activity by simulating individual neurons. Neuralinks are present in each and every layer of neural networks. Keras was used consecutively by the author. Aside from the input and output layers, the model also has three hidden layers. Any one of them can turn on independently of the output layer. The sigmoid function activates the output layer. When building the model, we consulted the Binary Merge Loss Function and the Adam Optimizer. This model's architecture makes use of ANN. Finally, the sigmoid function predicts if a profile is phone pay or real and returns a value between 0 and 1.

Hyper parameters

Piecewise linear factors indicate corrected activation functions. ReLU is generally the primary neural network learning algorithm since it is easy to train and performs well.

4.4.3 Random Forest

Random forest (or random-decision forest) ensemble learning is a component of this technique. It is used in machine learning because it is simple to apply to regression and classification. Each decision tree's forecasts are used by the random woodland, which predicts the outcome based on the majority of estimates. Notwithstanding, random-timberland produces far additional decision trees than the decision tree strategy, and the final result appears to be the average of nearly all of them. For profile identification, the author used random woods. Meaningful disclosures are generated by the model after data handling. To fit the data, the bootstrap aggregating system is used with the X=x1,x2,⋯,xn and Y=y1,y2,⋯,yn reaction trees (fb). At foreordained intervals (B times), a random sample is picked. The accompanying approach is used to choose the sample(x') results after training:

4.4.4 Extreme Gradient Boost

XG Boost is another regression ensemble learning method. This algorithm subsamples Stochastic Gradient boosting parameters.

Random forest has the limitation of only functioning well with complete inputs or without any missing data. To get around this, the author employs gradient boosting.

The boosting technique initialises F0(x).

$$f_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma)$$

Later, the loss function gradient is determined iteratively.

$$\gamma_{im} = -\alpha \left[\frac{\delta L(y_i, F(x_i))}{F(x_i)} \right]$$

Lastly, it defines the boosted model Fm(X).

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

The learning rate is α.

The multiplicative factor is γn.

4.4.5 Long Short-Term Memory

In order to determine if a profile is genuine, the author came up with an LSTM-based approach that uses tweets. Using this website and tweets to train an LSTM, the author oversaw the filter that removes tweet IDs.

Letters are written in lowercase on all tokens.

Tweets cannot contain stop words anymore.

The author proceeded to vectorize the sentences of these blockchain-enabled tweets byapplying an embedding layer. The output is generated by advancing the 32-dimensional vectoroutput of the LSTM through sigmoid-triggered layers.

5. EXPERIMENTAL RESULTS AND DISCUSSION

The training and testing results for each model are listed below. We present model accuracy comparisons, misfortune against eras graphs, stochastic woodland, XG help, and other ROC bends, as well as model comparisons for the LSTM neural organization.

5.1 Neural Network

The trained neural network's accuracy and loss graphs are shown in both Figures 2 and 3

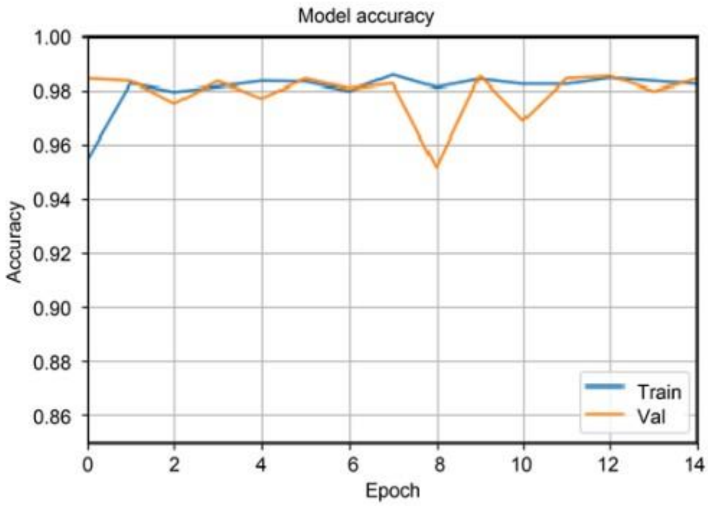


Figure 2: Model precision.

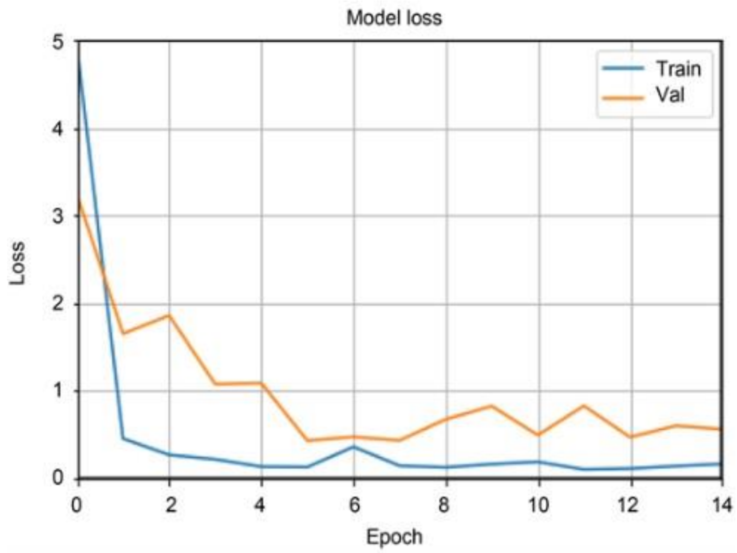


Figure 3: Model loss.

The aforementioned loss and accuracy graphs represent the end product of fifteen operational epochs. While it starts at 0.97, accuracy peaks at 0.98 and fluctuates throughout. After showing a local minimum of less than 0.5, the loss graph starts at 1 for the test dataset and 4 for validation. Loss is calculated using binary cross-entropy. Before weighting each characteristic, the machine gives random weights.

5.2 Random Forest and Other Approaches

The accuracy of decision trees, boost, random forests, and ada boosts are contrasted in Figure 4. The highest precision is produced with XG boost (0.996). Random forests and decision trees both have 0.99 accuracy. At last, the writer secures assistance from the ADA.

Figure 4 shows the accuracy comparison, while Figures 5 and 6 provide the ROC curve graphics.

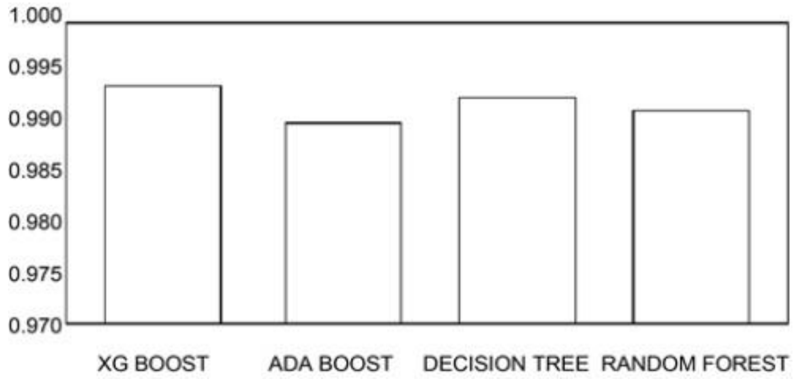


Figure 4: Accuracy of different models.

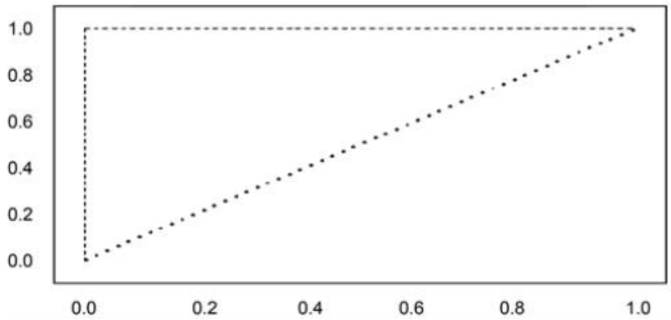


Figure 5: XG boost ROC curve.

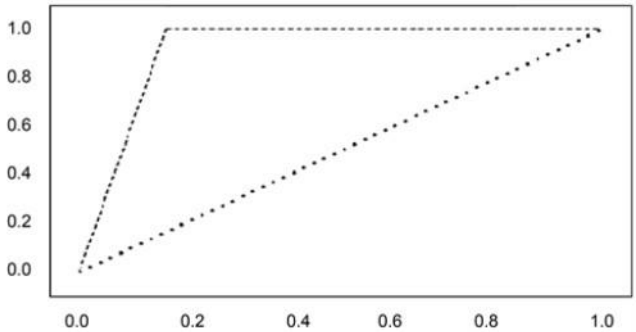


Figure 6: ROC curve for a random forest.

6. CONCLUSION

Using techniques including LSTM, XG Boost, Random Forest, and Neural Networks, this study created a strong machine learning model to identify phoney Twitter profiles, tackling the serious problem of phoney profiles that disseminate false information and influence public opinion. Finding essential components of genuine profiles, enhancing algorithm performance, and testing the models on a sizable dataset of real and fictitious profiles were among the main goals. The approach comprised profile-specific features and supervised learning techniques. SMOTE was used to preprocess and balance the data, and Google Colab's free GPU was used to train the models. According to experimental data, XG Boost fared better than other models in identifying phoney profiles, attaining the highest recall and precision rates. This demonstrates the superiority of XG Boost in detecting fake profiles and the usefulness of ensemble approaches in challenging detection tasks. By successfully detecting and disabling phoney personas, the study illustrates how machine learning may improve social media security and integrity and promote a safer online environment.

REFRENCES

1. Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake news detection using machine learning ensemble methods. Complexity, 2020(1), 8885861.
2. Ali, I., Ayub, M. N. B., Shivakumara, P., & Noor, N. F. B. M. (2022). Fake news detection techniques on social media: A survey. Wireless Communications and Mobile Computing, 2022(1), 6072084.
3. Aslam, N., Ullah Khan, I., Alotaibi, F. S., Aldaej, L. A., & Aldubaikil, A. K. (2021). Fake detect: A deep learning ensemble model for fake news detection. complexity, 2021(1), 5557784.
4. Balaji, T. K., Annavarapu, C. S. R., & Bablani, A. (2021). Machine learning algorithms for social media analysis: A survey. Computer Science Review, 40, 100395.
5. Della Vedova, M. L., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., & De Alfaro, L. (2018, May). Automatic online fake news detection combining content and social signals. In 2018 22nd conference of open innovations association (FRUCT) (pp. 272-279). IEEE.
6. Kaliyar, R. K., Goswami, A., & Narang, P. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia tools and applications*, *80*(8), 11765-11788.

7. Mishra, S., Shukla, P., & Agarwal, R. (2022). Analyzing machine learning enabled fake news detection techniques for diversified datasets. Wireless Communications and Mobile Computing, 2022(1), 1575365.
8. Mohammadrezaei, M., Shiri, M. E., & Rahmani, A. M. (2018). Identifying fake accounts on social networks based on graph analysis and classification algorithms. Security and Communication Networks, 2018(1), 5923156.
9. Monti, F., Frasca, F., Eynard, D., Mannion, D., & Bronstein, M. M. (2019). Fake news detection on social media using geometric deep learning. arXiv preprint arXiv:1902.06673.
10. Ozbay, F. A., & Alatas, B. (2020). Fake news detection within online social media using supervised artificial intelligence algorithms. Physica A: statistical mechanics and its applications, 540, 123174.
11. Ramalingam, D., & Chinnaiah, V. (2018). Fake profile detection techniques in large-scale online social networks: A comprehensive review. Computers & Electrical Engineering, 65, 165-177.
12. Sahoo, S. R., & Gupta, B. B. (2021). Multiple features based approach for automatic fake news detection on social networks using deep learning. Applied Soft Computing, 100, 106983.
13. Shu, K., Zhou, X., Wang, S., Zafarani, R., & Liu, H. (2019, August). The role of user profiles for fake news detection. In Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining (pp. 436-439).
14. Thota, A., Tilak, P., Ahluwalia, S., & Lohia, N. (2018). Fake news detection: a deep learning approach. SMU Data Science Review, 1(3), 10.
15. Van Der Walt, E., & Eloff, J. (2018). Using machine learning to detect fake identities: bots vs humans. IEEE access, 6, 6540-6549.

COMPARISON OF DIFFERENT PARAMETERS OF SOLAR CELL USING PC1D SIMULATION SOFTWARE

Yogesh Patel, Sachin Desai, Dr. Nimish Vasoya*
Department of Balbhvan, Children Research University, Gandhinagar, 382021

ABSTRACT

In order to build a solar cell that has a high degree of efficiency, it is necessary to conduct an in-depth investigation of the impact that a variety of physical and electrical qualities have on the overall efficiency of the cell. For the purpose of this investigation, a PC1D computer simulation was applied to evaluate the effect that altering key characteristics of a silicon solar cell has on the performance of the cell. For the purpose of determining the ideal values for the absorber, emitter, antireflection coating, and back surface field layers, we carried out a simulation. We paid particular attention to the doping levels as well as the thicknesses of each particular layer. A measured result acquired from an industrial scale manufactured solar cell that had the same specifications was used to assess the simulated solar cells that were produced as a consequence. There was a good degree of agreement between the results of the measurements and the data from the models. It is useful and practical to comprehend and forecast the impacts of these essential components and in order to get their ideal values through the utilization of a simulation tool, based on the research that has been undertaken and the discoveries that have resulted from it. This is necessary in order to design a solar cell structure that has a high efficiency.

KEY WORDS: Solar Cell, Simulation, Industrial Scale, Physical & Electrical Qualities, PC1D Computer Simulation.

INTRODUCTION

The plentiful solar energy that reaches the surface of the Earth has the ability to supply all of the energy requirements that mankind will have in the future and need in the present [1]. Solar cells that use photovoltaic technology are extremely efficient in converting sunlight into energy because they are able to capture and transform light particles. If we want to increase the amount of photovoltaic energy that is produced all over the world, one potential strategy is to improve the efficiency of solar cells. Crystalline silicon (c-Si) solar cells are responsible for 93 percent of the entire revenue generated by the solar cell industry, as indicated by studies carried out by the Fraunhofer Institute [2]. In the present day, the highest efficiency that has been demonstrated for mono-crystalline systems is 26.7%, whereas the efficiency for multi-crystalline systems is 21.9% [3-5]. For the purpose of further improving the efficiency of solar cells, it is essential to have a comprehensive understanding of the fundamental principles behind the working of semiconductors. Through the use of trustworthy simulation software, folks have the opportunity to acquire understanding regarding the relationship between the modification of the electrical and physical characteristics of various materials and the functioning of the device. The output of a device may be predicted by simulation software by modifying material parameters such as doping levels and layer thickness. This allows the program to make predictions about the performance of the device. A further benefit is that they are able to forecast the performance of a solar cell by combining mathematical concepts with experimental findings. A number of different modelling tools for solar cells are now available for use [6-8]. These packages include Silvaco TCAD, Sentaurus TCAD, AFORS-HET, and PC1D. Sentaurus TCAD is a solar cell modelling program that is extremely adaptable and has the ability to accurately forecast processes at the atomic level, even for objects with diameters that are less than 90 nm [9]. The expensive cost is a significant one of the disadvantages. It is possible to replicate typical solar cell materials like silicon and germanium by using PC1D, which is an alternative that is open-source. PC1D was developed by the Photovoltaics Special Research Centre at the University of New South Wales, which is widely recognized as a leading authority in solar cells on a global scale [8]. Through the manipulation of a number of factors, including temperature, doping levels, parasitic resistance, back surface fields, recombination, and carrier lifespan, it is possible to alter the overall performance of the device. The PC1D has the capability to graphically exhibit data, which includes curves that show the relationship between current and voltage (I-V), open circuit voltage (V_{oc}), short circuit current (J_{sc}), and other information that is relevant. In the event that an analysis is performed, the data may be employed for the purpose of planning the manufacture of technological devices.

For the purpose of determining whether or not their designs were feasible, the researchers used PC1D to run simulations of a number of different kinds of solar cells. PC1D was utilized by Sepeai et al. [10] and Meenakshi et al. [11] in order to accomplish the simulation of solar cells with a variety of junction configurations. The PC1D technique was applied by both Belarbi et al. and Chuan et al. [12, 13] in order to investigate silicon solar cells. However, the modelling program that utilized a solar cell, regardless of whether it was developed experimentally or commercially, did not have validation to assure that it was accurate and reliable. The purpose of this study was to investigate the effect that a number of important characteristics, including device thickness, doping levels, emitter thickness, back surface field thickness, doping level, and antireflection coating, had on the performance of crystalline silicon solar cells. In order to successfully replicate and evaluate the performance of the device, PC1D made use of the manipulation of the aforementioned features. The result highlights how important it is to evaluate and determine the perfect value of each parameter in order to achieve the best possible efficiency from the device. One of the most impressive aspects of this research is that the optimized parameters of the simulated device were confirmed by comparing the findings to those of a commercially produced solar cell that had the same physical and electrical properties as the simulated device.

METHOD

The typical structure of a silicon solar cell, which is generally utilized in the industrial sector, is seen in Figure 1. It is essential to have a solid understanding of the effect that the various physical and electrical features of each layer have on the conversion efficiency in order to achieve high levels of efficiency. For the purpose of achieving the maximum possible conversion efficiency, the PC1D simulation tool was utilized to investigate the impact that a variety of device characteristics had on each layer within the system.

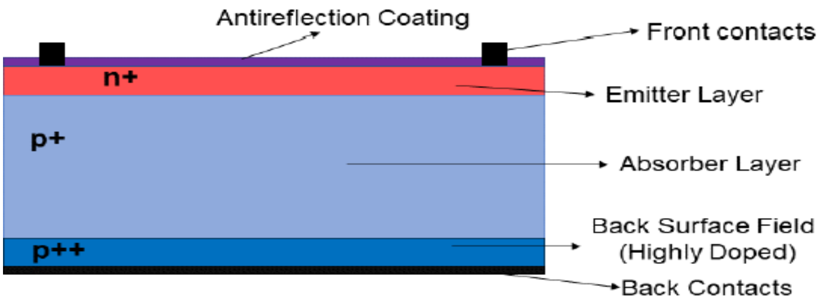


Figure 1. Basic structure of a conventional silicon solar cell with selective emitter

While the absorber, emitter, and rear surface field are responsible for producing and transferring mobile charge carriers, the front and back contacts are responsible for collecting them. The efficiency of solar cells may be improved by using an antireflection coating, which reduces the amount of surface reflection while simultaneously increasing the amount of light that is transmitted and absorbed. A silicon nitride antireflection coating with a refractive index of 1.873 was utilized in order to achieve the goals of minimizing the quantity of light that was reflected at wavelengths that had high spectrum irradiance and providing efficient surface passivation [14]. The standard specifications of the solar cell that was utilized in the experiment are presented in Table 1.

Table 1: Parameters of Solar Cell by Using PC1D

Internal optical reflectance	Enabled
Emitter Sheet Resistance	71.85 Ω/sq
Front surface texture depth	3 μm
Shunt Resistance	50000 Ω
Intrinsic concentration @300k	1×10^{10} cm^{-3}
Front diffusion (N-type)	2×10^{20} cm^{-3} peak
Thickness (Absorber Layer)	180 μm
Rear diffusion (P-type)	3×10^{18} cm^{-3} peak
Front SRV	2×10^{5} cm/s
Bulk recombination	$\tau_n = \tau_p = 30$ μs
Device Area	1 cm^2
Rear SRV	1×10^{7} cm/s
Temperature	25°C

THICKNESS OF THE ABSORBER LAYER AND ITS IMPACT ON SOLAR CELL EFFICIENCY

When determining whether or not to manufacture a photovoltaic device, one of the most important considerations to take into account is the cost of the materials used in semiconductors [14]. In order to save expenses while simultaneously optimizing the efficiency of the gadget, it is essential to use materials that have the thickness that is most appropriate on the market. When it comes to the process of converting light into mobile charge carriers, the absorber layer, which is the component of commercial silicon solar cells that is the thickest, makes a significant contribution. These charge carriers are then transmitted and collected by the contacts, which results in the generation of energy by the process [15]. When it comes to improving efficiency, a bigger absorber layer is not useful because of the contradicting effects it has on Voc (open-circuit voltage) and Jsc (short-circuit current). In this part of the article, we examined a solar cell that had absorber layers of varying thicknesses. The performance of the device is analyzed in Table 2, which displays the impact that different bulk thicknesses of silicon have on the device. As can be observed in Figure 2, the thickness of the absorber layer has an effect on several parameters, including V_{oc}, J_{sc} (Figure 2a), and efficiency (Figure 2b). The parameters 'T_{bulk}', 'FF', and 'η' are used to represent the bulk thickness fill factor, and conversion efficiency, respectively.

Table 2: Device performance depends on absorber thickness

J_{sc}	V_{oc}	η (%)	T_{bulk}	FF (%)
36.94	619.4	17.75	280	77.56
37.03	620.3	17.83	250	77.62
37.11	621.5	17.91	220	77.65
37.18	622.9	17.99	190	77.66
37.21	624.5	18.04	160	77.64
37.20	626.4	18.12	130	77.76
37.10	628.7	18.16	100	77.86
36.66	632.3	18.09	60	78.06
35.50	635.0	17.64	30	78.25

Figure 2 (a)

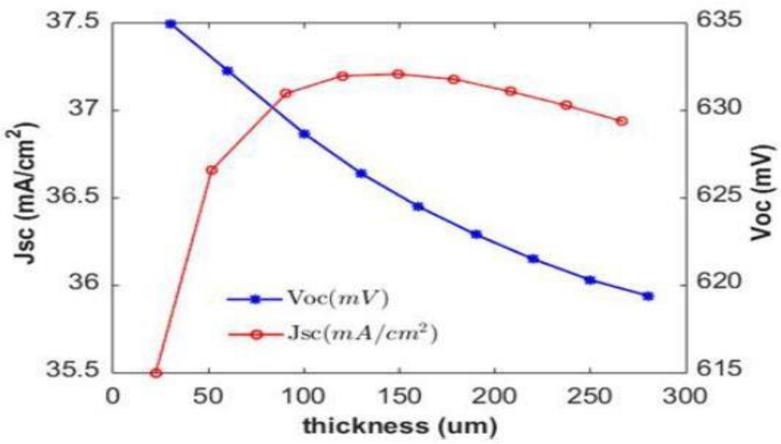
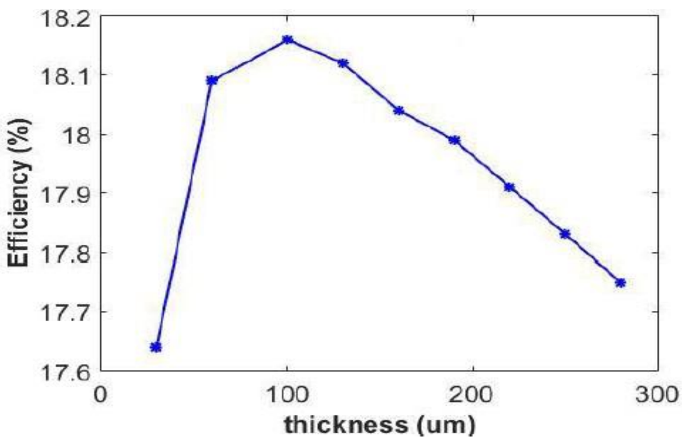


Figure 2 (b)



Within the region of 30 to 280 µm, the V_{oc} value exhibits a connection that is inversely proportional to the thickness. On the other hand, the J_{sc} value exhibits a direct correlation with thickness up until the bulk thickness reaches 160 µm, at which time the relationship reverses itself. If the thickness of the device is greater than 100 µm, the overall efficiency of the device will decrease. Due to physical constraints, such as the bending impact of conventional aluminium back surface fields and the difficulties associated with handling such a minuscule device, manufacturers frequently opt for devices with a thickness exceeding 150µm. This is despite the fact that devices with a thickness of 100µm have a higher efficiency than those with a thickness of 150µm. When anything is being manufactured, it is essential to take into consideration characteristics such as dependability, longevity, and resistance to adverse weather conditions.

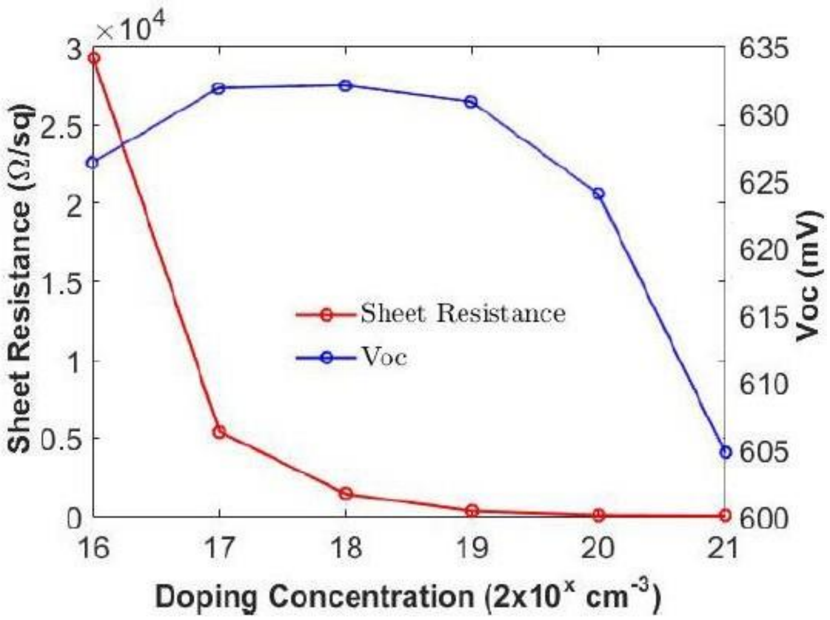
OPTIMISING DEVICE PERFORMANCE WITH VARYING EMITTER DOPING CONCENTRATIONS

A considerable amount of the light is absorbed at the surface of the solar cell, which provides the solar cell with the ability to produce a high output rate. In order to achieve the highest possible level of light absorption, it is essential to make precise adjustments to the thickness of the emitter as well as the doping concentration. It is possible for an increase in the concentration of doping in these layers to result in a drop in efficiency due to a decrease in light transmission and absorption as well as an increase in the rate of recombination [17]. However, in order to make the drift transport process easier and to get a reduced sheet resistance, it is also important to have a concentration that is suitably high. Table 3 illustrates the effect that different doping concentrations have on the sheet resistance as well as the overall performance of the device. The relationship between the concentration of emitter doping and the resistance of the emitter sheet and V_{oc} is seen in Figure 3. Doping concentration (C_{dop}) and sheet resistance (R_{sht}) are two examples of measures that are extremely important in this industry.

Table 3: Sheet resistance & device performance as a function of emitter doped concentrations

J_{sc} (mA/cm^2)	V_{oc} (mV)	η (%)	R_{sht} (Ω/sq.)	C_{dop} (cm^{-3})	FF (%)
34.33	604.3	17.75	8.7	$2X10^{21}$	77.48
37.22	623.6	17.83	71.85	$2X10^{20}$	77.70
37.49	630.5	17.91	421.7	$2X10^{19}$	77.80
37.52	631.7	17.99	1751	$2X10^{18}$	77.79
37.52	631.5	18.04	7852	$2X10^{17}$	77.78
37.52	625.9	18.12	79650	$2X10^{16}$	77.73

Figure 3: Impact of Concentration of Emitter Doping on Emitter Sheet Resistance & V_{oc}



Because of its low sheet resistance, a device with a thickness of 150 μm and an emitter doping concentration of 2×10^{20} cm^{-3} was selected for this study. Both the sheet resistance and the overall performance of the device are significantly influenced by the thickness of the emitter layer. Each of these factors is important. Table 4 illustrates the performance of the device with a range of different emitter thicknesses. "T_{emt}" is an abbreviation that stands for the word "emitter thickness."

Table 4: Impact of emitter thickness on the efficiency of the device

J_{sc} (mA/cm^2)	**V_{oc} (mV)**	**η (%)**	**T_{emt} (μm)**	**R_{sht} (Ω/sq.)**	**FF (%)**
29.04	612.5	14.14	0.7	10.26	79.49
30.39	613.0	14.79	0.6	11.98	79.39
31.88	613.6	15.49	0.5	14.37	79.18
33.43	614.3	16.23	0.4	17.96	79.01
34.91	615.3	16.93	0.3	23.95	78.83
36.15	617.0	17.55	0.2	35.93	78.68
36.95	619.7	18.03	0.1	71.85	78.76

When the thickness of the emitter is increased and the sheet resistance is decreased, the values of J_{sc} & V_{oc} fall. This is because the value of Jsc decreases when the sheet resistance decreases. In spite of the fact that a device with a low emitter sheet resistance is believed to be the best option, thick emitters may have major drawbacks. This is because the thick and severely doped emitter layer is unable to allow light to pass through it, which in turn has an influence on the creation of charge carriers. As a result, the device's efficacy is reduced.

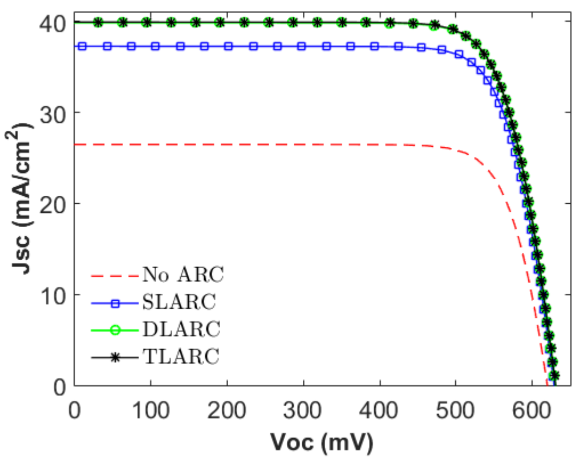
ANTIREFLECTION COATING: AFFECTS ON DEVICE FUNCTION

The antireflection coating, which is also commonly referred to as ARC, is an important component that is necessary for improving the efficiency of solar cells, a thin coating of dielectric material is put to the surface of a solar cell in order to increase the amount of charge carriers that an individual solar cell is capable of producing. Because of this coating, the overall reflectance of the light that is coming in is decreased, and the transmission is improved [16]. In current solar cells, ARCs are made up of stacked layers of dielectric materials that have various refractive indices. These layers can be either single-layered or double-layered, depending on the application need.

Table 5: Effect of different levels of ARC on device performance

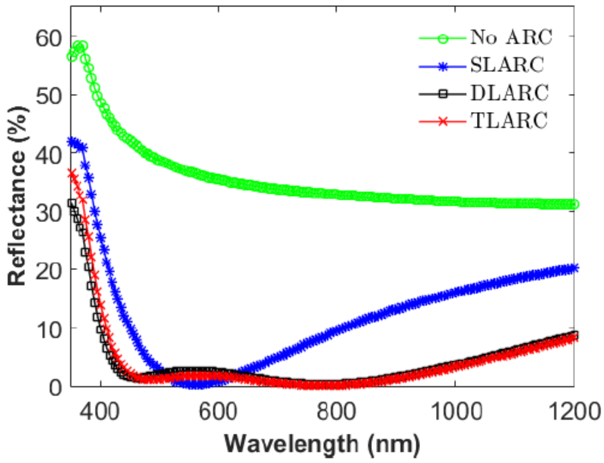
J_{sc} (mA/cm^2)	V_{oc} (mV)	η (%)	ARC	FF (%)
39.96	630.6	19.79	TLARC	78.53
39.90	630.6	19.77	DLARC	78.56
37.30	628.8	18.5	SLARC	78.88
26.50	619.8	13.1	None	79.79

By simulating silicon samples with no anti-reflective coating (ARC), one layer of ARC, two layers of ARC, and three layers of ARC, researchers were able to study the effect that applying varying numbers of ARC layers had on the performance of the device. We made use of a single layer of anti-reflective coating (ARC), which was made up of titanium dioxide (TiO_2) and had a thickness of 67 nanometers. The refractive index of this coating was 2.116. Both magnesium fluoride (MgF_2) and zinc sulphide (ZnS) were utilised in the production of the double layer antireflection coating (DLARC), which had refractive indices of 1.39 and 2.371, respectively. The thickness of the DLARC was 107 nm for magnesium trifluoride and 60.5 nm for zinc sulphide, respectively.



(a)

Figure 4 (a)



(b)

Figure 4 (b)

Triple layer anti-reflective coating, also known as TLARC, was made up of three distinct materials: magnesium fluoride, silicon dioxide, and titanium dioxide. It was determined that the refractive indices of these materials were 1.39, 1.48, and 2.453, respectively. For each of the materials, the coating had a thickness of 80 nanometers, 30 nanometers, and 60 nanometers. The findings of each device's performance are presented in Table 5, which may be seen here. The I-V curve and reflectance spectra are depicted in Figure 4a and 4b, respectively, for solar cells that include ARC and those that do not contain ARC.

EVALUATION OF THE BEST MODELLED AND REALISED SOLAR CELL

To ensure that the results of the simulation are accurate, the actual measurements of solar cells were compared to the measurements of a virtual device that had the same characteristics as the real one. There is a comparison of the electrical properties of a real solar cell and a simulated solar cell that is presented in Table 6.

Table 6: Analysing the electrical characteristics of a manufactured solar cell in comparison to an ideal simulated one

J_{sc} (mA/cm^2)	V_{oc} (mV)	η (%)	Type of Data	FF (%)	Pseudo η (%)
37.3	628.8	18.5	Simulated	78.88	-
36.6	617.0	17.7	Real	78.30	18.48

This is demonstrated by the data that is presented in Table 6, which reveals that the simulation results have the capacity to accurately anticipate the actual cell values. When compared to the real cell values, the simulated results were shown to be considerably more favourable overall. The doping concentration of the cell's emitter was found to be 5×10^{20} cm^{-3}, while the doping concentration that was wanted for a uniform emitter in the simulation was found to be 2×10^{20} cm^{-3}. The diameters of the emitters in the cell were 0.6 micrometres and 0.1 micrometres, respectively. When recombination losses are kept to a minimum, even a minor improvement in the performance of the simulated cell has the potential to fall within a range that is considered acceptable. Once resistive losses have been eliminated, the simulation is able to produce an accurate assessment of the cell's potential efficiency, which was found to be 18.48% by the use of the Suns-V_{oc} measuring tool.

CONCLUSION

According to the findings of this study, in order to achieve the best possible conversion efficiency, it is necessary to investigate and determine the settings that are most beneficial for each device parameter. The use of PC1D simulation software is what allows this to be performed. When utilising simulation software, it is of the utmost importance to take into consideration the features of the device. These qualities include the device's stability, durability, capacity to withstand circumstances of severe weather, and the constraints of the manufacturing process. Overall, this study demonstrates that it is possible to use PC1D and other simulation software as a practical alternative throughout the research and development stage of making crystalline silicon solar cells. This is due to the fact that these software programmes are accurate and reliable.

REFERENCES

1. Basic Photovoltaic Principles and Methods, Solar Information Module (1982), SERI/SP-290-1448.
2. S. Philipps, Photovoltaics Report, Fraunhofer ISE and Werner Warmuth, PSE AG, https://www.ise.fraunhofer.de/content/dam/ise/de/documents/publications/studies/PhotovoltaicsReport.pdf
3. M. A. Green, Y. Hishikawa, W. Warta, E. D. Dunlop, D. H. Levi, J. Hohl-Ebinger and A. W. H. Ho-Baillie, "Solar Cell Efficiency Tables (version 50)", Prog. Photovoltaics, vol. 25, pp. 668- 676, 2017.
4. K. Yoshikawa, H. Kawasaki, W. Yoshida, T. Irie, K. Konishi, K. Nakano, T. Uto, D. Adachi, M. Kanematsu, H. Uzu and K. Yamamoto, "Silicon heterojunction solar cell with interdigitated back contacts for a photo conversion efficiency over 26%," Nat. Energy, vol. 2, no. 5, 2017.
5. J. Benick, A. Richter, R. Müller, H. Hauser, F. Feldmann, P. Krenckel, S. Riepe, F. Schindler, M. C. Schubert, M. Hermle, A. W. Bett and S. W. Glunz "High-Efficiency n-Type HP mc Silicon Solar Cells," IEEE J. Photovoltaics, vol. 7, no. 5, pp.1171–1175,2017.
6. J. Hofstetter, C. Del Cafiizo, S. Ponce-Alcantara and A. Luque, "Optimisation of SiNx: H anti-reflection coatings for silicon solar cells", Spanish Conference on Electron Devices, pp.131-134, 2007.
7. K. Islam, A. Alnuaimi, H. Ally and A. Nayfeh, "ITO, Si3N4 and ZnO: Al Simulation of Different Anti-reflection Coatings (ARC) for Thin Film a-Si: H Solar Cells", 2013 European Modelling Symposium, Manchester, pp. 673-676, 2013.
8. D. A. Clugston and P. A. Basore, "PC1D Version 5- 32-Bit Solar Cell Modelling on Personal Computers", 26th IEEE Photovoltaics Specialist Conference, Anaheim California, pp. 207–210, 1997.
9. Synopsis, "Synopsys TCAD Now Offers Atomic-level Accuracy." Online Available: https://news.synopsys.com/index.php?s=20295&item=122584. (Accessed: 21-July-2024).
10. S. Sepeai, M.Y. Suleiman, M. Khairunaz, A.W. Azhari, K. Sopian, S.H. Zaidi, "Design Optimization of Bifacial Solar Cell by PC1D Simulation", Journal of Energy Technologies and Policy, vol. 3, no. 5, pp. 1–11, 2013.

11. S. Meenakshi and S. Baskar, "Design of multi-junction solar cells using PC1D", 2013 International Conference on Energy Efficient Technologies for Sustainability, Nagercoil, 2013, pp. 443-449.
12. M. Belarbi, A. Benyoucef, and B. Benyoucef, "Simulation of the solar cells with PC1D, application to cells based on silicon", Advanced Energy: An International Journal, vol. 1, no.3, 2014.
13. J. Chuan, L. Tianze, Z. Xia, H. Luan, "Simulation of Silicon Solar Cell using PC1D", Advanced Materials Research, Vols. 383-390, pp 7032-7036, 2012.
14. B. Liu, S. Zhong, J. Liu, Y. Xia and C. Li, "Silicon nitride film by inline PECVD for black silicon solar cells," International Journal of Photo energy, vol. 2012, pp. 2–7, 2012.
15. C. T. Sah, K.A. Yamakawa, R. Lutwack, "Effects of Thickness on Silicon Solar Cell Efficiency", IEEE Transactions on Electron Devices, vol. 29, no. 5, 1982.
16. A. Mandong, "Design and Simulation of Single, Double, and Multi-Layer Antireflection Coating for Crystalline Silicon Solar Cell", Master Thesis, Karadeniz Technical University, Trabzon, Turkey, 2019.
17. M. Wolf, "The Influence of Heavy Doping Effects on Silicon Solar Cell Performance", Solar Cells, vol. 17, pp. 53-63, 2018.
18. R.R. King, K.W. Mitchell and J.M. Gee, "Back Surface Cell Structures for Reducing Recombination in CZ Silicon Solar Cells", Proceedings of 1994 IEEE 1st World Conference on Photovoltaic Energy Conversion - WCPEC (A Joint Conference of PVSC, PVSEC and PSEC), Waikoloa, USA, 1994.

AN EXHAUSTIVE EXAMINATION OF THE MODELS, METHODS, HISTORIES, AND VIEWPOINTS RELATED TO WATER QUALITY INDEXES (WQIs)

Yogendra Singh[1], Sandhya Chaudhary[2], Ravikant[3]

Research scholar, Mangalayatan University Aligarh[1]
Co-supervisor, N.R.E.C COLLEGE Khurja-Bulandshar[2]
Supervisor, Mangalayatan University Aligarh[3]

Abstract

This study provides a detailed analysis of Water Quality Indexes (WQIs), tracing their historical development and evolution from simple aggregative models to sophisticated approaches using multivariate analysis and machine learning. It explores data collection, normalization, weighing, and aggregation methods, highlighting their advantages and disadvantages. The research also explores the influence of technology and environmental challenges on WQI creation. The review seeks to clarify WQIs' intricacies and provide future directions for improving environmental monitoring and public health.

Keywords: *Water Quality Indexes (WQIs), Water, Models, Techniques, Water Quality Ground Water.*

1. INTRODUCTION

One essential natural resource for human survival is water, supporting various activities and well-being. Surface and groundwater sources are crucial, but access remains a major obstacle, with over 1.1 billion people lacking clean water [1]. By 2025, over two-thirds of nations will face water-related issues [2].

Figure 1: Water Quality Indexes [3]

Water quality can deteriorate due to factors like social and economic growth, human activities, climate changes, and hydrological variations [4]. The accumulation of pollutants in surface water might result from these conditions, lowering the quality of the water. Maintaining water resources requires effective management [5].

Water quality is evaluated by tracking changes in its synthetic, organic, and physical components, which can be manipulated by natural and artificial cycles [6]. Water Quality Indexes (WQIs) are an effective method for representing water quality, reducing complex indices to a single value between 0 and 100 [7].

This study aims to explore various models, systems, histories, and perspectives related to water quality indexes (WQIs) [8]. It will examine various water quality lists used globally for assessment and discuss their advantages and disadvantages [9]. The study will also explore the intricacies of water quality evaluation, their historical roots, mechanical mechanisms, and various perspectives. The review will also propose future directions for working on WQIs in water resources management [10].

2. WATER QUALITY INDEX

The Water Quality Index (WQI) is a device used to evaluate water quality, making it easier to reveal information and crucial for natural observation, policymaking, and public communication due to its straightforward and effective estimation [11].

- **Historical Perspective of WQIs**

In the 1960s, the World Quality Index (WQI) was created to evaluate water quality by integrating various criteria into a single score [12]. Over time, it has evolved with advanced statistical and computational techniques for improved reliability and accuracy [13].

- **Models Used in WQIs**

WQIs are calculated using various models with unique methodologies and applications, with the most common models being [14]:

1. **Arithmetic Mean Method**: WQIs are calculated using various models with unique methodologies and applications, with the most common models being.
2. **Geometric Mean Method**: This method emphasizes lower values, making it more sensitive to poor water quality, and is particularly useful in identifying potential health risks in water bodies.
3. **Weighted Sum Method**: The approach assigns weights to various water quality parameters, resulting in a more complex and representative WQI index.

- **Historical Evolution and Key Milestones**

The World Quality Index (WQI) was made during the 1960s to assess water quality by coordinating different rules into a solitary score [15]. Over time, it has evolved with advanced statistical and computational techniques for improved reliability and accuracy [16].

➢ **Phases of WQI development**

The World Quality Index (WQI) was made during the 1960s to assess water quality by coordinating different measures into a solitary score [17]. Over time, it has evolved with advanced statistical and computational techniques for improved reliability and accuracy [18].

- **Early Conceptualization**

The mid-20th century saw the development of Water Quality Indexes (WQI), initially focusing on basic factors like pH, dissolved oxygen, and turbidity. These early indexes provided a basic overview of water quality but lacked depth to capture complex biological and chemical interactions.

- **Refinement and Expansion**

The second phase of environmental science saw the refinement and expansion of WQIs, with advanced statistical methods and a wider range of parameters introduced. This led to the invention of weighted arithmetic indexes like WAWQI, allowing for fairer evaluations of water quality. Advancements in processing power enabled multivariate statistical methods for complex datasets.

- **Standardization and Policy Integration**

The third stage of development aimed to incorporate WQI into environmental policy frameworks and standardize them, adopting standardized WQIs by organizations like WHO and EPA. Area-specific indices were created to address unique regulatory requirements and environmental conditions.

- **Modern Innovations**

The World Quality Index (WQI) is a 21st-century tool that uses advanced technology and techniques to monitor water quality in real-time. It incorporates sensor technology and remote sensing, allowing WQIs to react to environmental changes. Machine learning and artificial intelligence are used to analyse large datasets, identify patterns, and forecast potential changes. As technology advances, WQIs become more important in environmental monitoring, policy-making, and public awareness, making them a crucial component of global efforts to conserve and manage water resources.

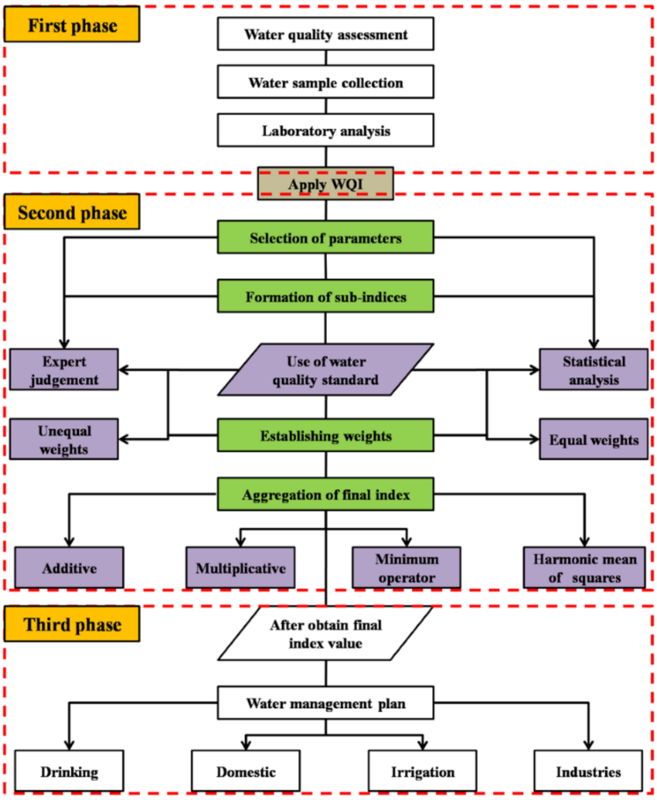


Figure 2: Phases of WQI development [19]

3. EVOLUTION OF WQI RESEARCH

The Water Quality File (WQI) study, developed in the 1960s, has significantly improved water quality management. It aims to create a uniform framework for assessing water quality by consolidating various variables into a single mathematical score, making complex information more accessible [20].

During the 1970s and 1980s, WQI research expanded to include physical, chemical, and biological indexes. Researchers adjusted WQIs for individual water bodies and intended uses, including drinking water, recreational water, and aquatic ecosystems [21]. Region-specific health quality indexes were developed to consider local environmental conditions and pollution sources [22].

The 1990s and 2000s saw significant advancements in computational technology and geographic information systems (GIS), influencing the development of Water Quality Index (WQI) research [23]. The integration of remote sensing data and spatial analysis methodologies was necessary for evaluating water quality on larger spatial scales and in dynamic environments. Multi-rules choice analysis (MCDA) was introduced to guide comprehensive assessments of water quality [24]. The focus has shifted to contemporary water quality issues, such as toxins, environmental change impacts, and sustainable water management. Participatory methodologies and AI are being used to develop more adaptable WQIs. The development of WQI research reflects the growing complexity of water quality problems and the need for more comprehensive and adaptive tools [25].

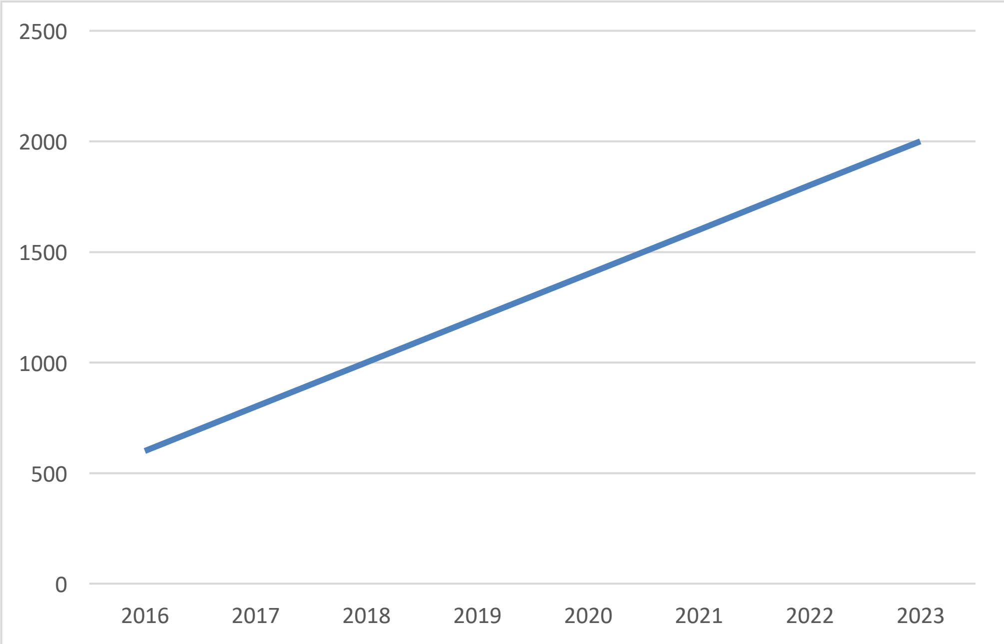


Figure 3: Water Quality Index (WQI) Research Over the Years [26]

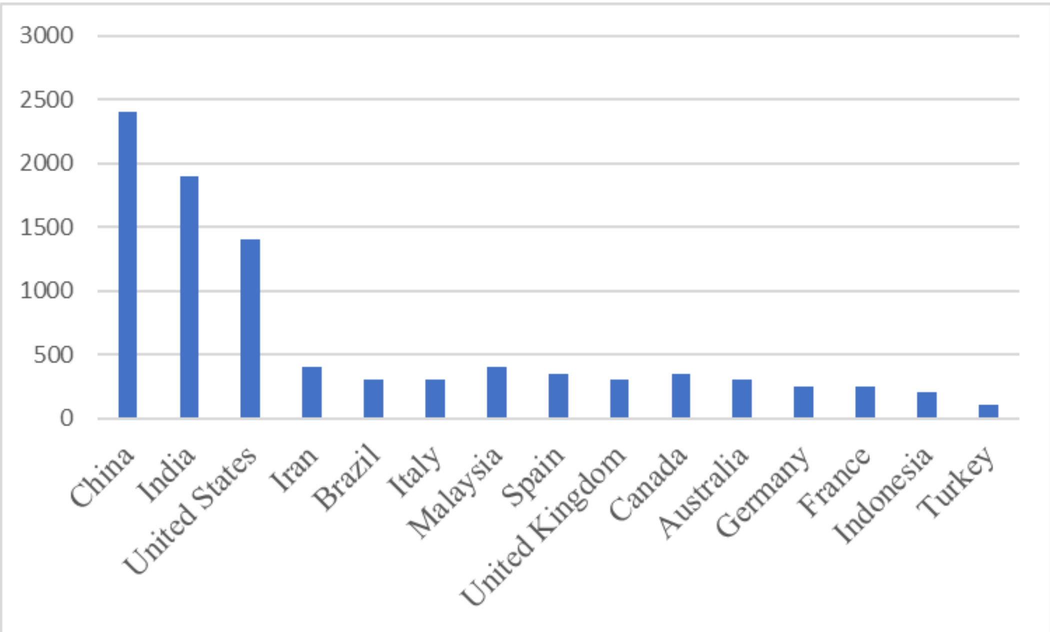


Figure 4: Water Quality Index (WQI) Research by Country/Region [27]

4. DIFERENT METHODS FOR WQI DETERMINATION

The Water Quality Index (WQI) is resolved utilizing different techniques, including the Weighted Arithmetic Index, NSFWQI, CCMEWQI, and OWQI [28]. The Weighted Arithmetic Index method is a simple and flexible approach that involves weighing water quality characteristics like nutrients, heavy metals, pH, turbidity, and dissolved oxygen. This method allows for the inclusion of a wide range of criteria and weightings based on the specific water body and research goals [29].

4.1. National sanitation foundation (NSFWQI)

The National Sanitation Foundation's Water Quality Index (NSFWQI) is a broadly involved device for evaluating and dispersing data about water quality. Made during the 1970s, it comprises of nine key elements, including Broke up Oxygen (DO), pH, Biochemical Oxygen Interest (Body), temperature change, complete phosphate and nitrate levels, turbidity, and all out solids. Higher DO values indicate higher quality water, while high pH levels negatively impact contaminant solubility and aquatic life. High BOD indicates low DO and high organic pollution. Temperature change affects metabolic rates and species diversity, while high phosphate and nitrate levels can lead to eutrophication and health problems [30]. The ultimate score on the index can range from 0 to 100, with qualitative descriptions used to make understanding easier.

Table 1: Water Quality Index (WQI) score into different qualitative descriptors [31].

WQI Score Range	Qualitative Descriptor
90-100	Excellent
80-90	Good
60-80	Medium
30-60	Bad
0-30	Very Bad

4.2. Canadian council of ministers of the environment water quality index (CCMEWQI)

The Canadian Council of Ministers of the Environment Water Quality Index (CCME WQI) is a device utilized in Canada to survey and impart surface water quality information. It offers a solitary benefit that gathers various water quality models into a straightforward index score, working with understanding and evaluation of water quality across different settings and time spans. The CCME WQI thinks about physical, compound, and organic qualities of water, including pH, broke up oxygen, biochemical oxygen interest, all out suspended solids, and foreign substances like supplements and weighty metals. Sub-records are appointed to each water quality measurement, featuring regions where water quality might be debased. The last score is resolved utilizing a weighted arithmetic mean or other suitable strategy [32].

- **Interpretation of CCME WQI Scores:**

Table 2: Water Quality Based on The CCMEWQI [33]

Class	WQI Value Range	Water Quality Description
I	76–100	Great water quality is protected by near absolute absence of danger. The environment is about exactly as it would be in nature.
II	66–75	A minor danger is warded off by good water quality. Situations rarely diverge from average values.
III	51–65	Alright Although occasionally threatened, the quality of the water is normally safe. Situations can occasionally diverge from average values.
IV	41–50	Not good (marginal) Water quality is frequently at risk. The conditions often diverge from average levels
V	0–40	Extremely Negative (Deficient) Water quality is almost never safe. Typically, the conditions differ from average levels.

4.3. Oregon water quality index (OWQI)

The Oregon Water Quality Index (OWQI) is a US technique for evaluating water quality in Oregon, taking into account factors like temperature, broke down oxygen levels, pH balance, conductivity, turbidity, complete nitrogen, all out phosphorus, and waste coliform microorganisms, deciding water reasonableness for drinking and oceanic territory support [34].

- **Calculation and Interpretation**

The OWQI measures water quality by determining the significance of each parameter and applying weights. Data from monitoring stations is normalized for fair comparisons across water bodies and periods. The weighted scores are aggregated, resulting in a score ranging from 0 to 100, with higher scores indicating better water quality [35].

Table 3: As Per the OWQI Index, Water Index [36]

Numerical Value Range	Condition	Colour
90–100	Excellent	Blue
85–90	Good	Green
80–85	Medium	Yellow
50–75	Bad	Orange
10–40	Very Bad	Red

The Original Water Quality Index (OWQI) categorizes water quality into levels like excellent, good, fair, bad, and extremely poor, aiding public, environmental agencies, and policymakers in understanding water health. However, it has limitations such as weighting methodologies and omission of certain contaminants. Despite its usefulness, OWQI is often used in conjunction with other assessments for comprehensive water management and conservation activities [37].

4.4. Weighted arithmetic water quality index (WAWQI)

The Weighted Arithmetic Water Quality Index (WAWQI) is a mathematical marker that consolidates numerous water quality boundaries into a solitary measure. It works on information examination by allocating loads to factors in light of their significance in surveying water quality [38]. The WAWQI standardizes each parameter to a common range of 0 to 100, ensuring accurate integration. The process involves selecting key water quality metrics, sizing data, assigning weights, and calculating the weighted average. A higher score indicates better water quality, while a lower score indicates worse. The WAWQI is useful for stakeholders, including environmental regulators, policymakers, and researchers, to assess and compare water quality across locations or over time. However, it may be subject to subjectivity and lacks consideration for factors' interactions. Regular updates and adjustments to the WAWQI are necessary to ensure its relevance and accuracy in sustainable water management practices [39].

$$\text{WAWQI} = \sum_{i=1}^{n} (W_i \times Q_i)$$

where:

- Wi = Weight assigned to parameter i
- Qi = Normalized score of parameters i
- n = Total number of parameters

Table 4: Numerous studies pertaining to Water Quality Indexes (WQIs) [40]

Authors	Construct Methods	Service Context	Key Findings
Roșca (2020) [41]	WAWQI	Glacial lakes, Rodnei mountains, Romania	Good quality water with a small amount of anthropogenic change. WQI stands for outstanding and superior quality. No pollution is shown by the heavy metal pollution index.
Nair et al. (2020) [42]	WAWQI	River basins of the Ithikkara and Kallada in Kerala, India	The majority of the water is graded as "excellent" during the monsoon and pre monsoon. Ion-parameter connections are investigated.
Teodorof L, et al. (2021) [43]	WAWQI & IWQI	Hilly terrain of Jammu Himalaya	Most spring samples are acceptable for household use and irrigation. Hazards associated with sodium and salinity in groundwater.
Udeshani et al. (2020) [44]	WAWQI	Monaragala, Sri Lanka	High levels of hardness, EC, TDS, Cl, and fluoride are caused by lithology, which affects groundwater quality. Very low groundwater quality is indicated by a WQI.
Betis et al. (2020) [45]	CCMEWQI	Quebec, Canada	Disparities between streams and water from harvested peatlands were observed. pH recommendations frequently disregarded in peatlands.

Cristable et al. (2020) [46]	NSFWQI	Saluran Tarum Barat, West Java	Medium water quality impacted by industrial, infrastructure, and agricultural activity.
En-nkhili et al. (2020) [47]	WWQI	Boudaroua Lake, Morocco	Good water quality impacted by agriculturally produced nitrogen compounds
Nong X et.al (2020) [48]	WWQI	Boudaroua Lake in the Moroccan Pre-Rif	Good water quality overall, but tainted by agricultural activities' organic nitrogen compounds, particularly following the first fall rains.

This table categorizes studies based on authors, the constructed methods (such as WAWQI, CCMEWQI, NSFWQI, etc.), the service context (geographical locations), and key findings related to water quality assessments.

5. CONCLUSION

Water quality indices (WQIs) are a diverse collection of models, methodologies, histories, and perspectives that help understand and control water quality [49]. The Weighted Arithmetic Water Quality Index (WAWQI) is a great representation, consolidating different logical disciplines to assess sea-going environments and drinking water wellbeing [50]. WQIs have evolved from simple aggregation methods to complex systems, considering intricate interactions between parameters. Despite challenges like weighting subjectivity and adaptation to environmental issues, WQIs remain vital tools for water resource protection and sustainable development.

REFERENCES

1. Uddin, M. G., Nash, S., & Olbert, A. I. (2021). A review of water quality index models and their use for assessing surface water quality. *Ecological Indicators*, *122*, 107218.
2. Patel, P. S., Pandya, D. M., & Shah, M. (2023). A systematic and comparative study of Water Quality Index (WQI) for groundwater quality analysis and assessment. *Environmental Science and Pollution Research*, *30*(19), 54303-54323.

3. Lukhabi, D. K., Mensah, P. K., Asare, N. K., Pulumuka-Kamanga, T., & Ouma, K. O. (2023). Adapted water quality indices: Limitations and potential for water quality monitoring in Africa. *Water*, *15*(9), 1736.
4. Moudou, M., El Hammoudani, Y., Haboubi, K., Achoukhi, I., El Boudammoussi, M., Faiz, H., ... & Dimane, F. (2024). Water quality indices (WQIs): An in-depth analysis and overview. In *E3S Web of Conferences* (Vol. 527, p. 02015). EDP Sciences.
5. Kumar, D., Kumar, R., Sharma, M., Awasthi, A., & Kumar, M. (2024). Global water quality indices: Development, implications, and limitations. *Total Environment Advances*, *9*, 200095.
6. Uddin, M. G. (2020). Development of a novel water quality index model usingdata. *Journal of King Saud University-Computer and Information Sciences*, *34*, 4773- 4781.
7. Mogane, L. K., Masebe, T., Msagati, T. A., & Ncube, E. (2023). A comprehensive review of water quality indices for lotic and lentic ecosystems. *Environmental Monitoring and Assessment*, *195*(8), 926.
8. Keraga, A. S. (2019). Assessment and modeling of surface water quality dynamics in Awash River Basin, Ethiopia. *Addis Ababa Institute of Technology, School of Chemical and Bio Engineering*
9. Moeinzadeh, H., Yong, K. T., & Withana, A. (2024). A critical analysis of parameter choices in water quality assessment. *Water Research*, 121777
10. Johnston, N., Rolfe, J., & Flint, N. (2024). A systematic review of agricultural use water quality indices. *Environmental and Sustainability Indicators*, 100417.
11. Aguilar, J. A. P., Campo, J., Meneu, S. N., Gimeno-García, E., & Andreu, V. (2019). Analysis of existing water information for the applicability of water quality indices in the fluvial-littoral area of Turia and Jucar Rivers, Valencia, Spain. *Applied Geography*, *111*, 102062.
12. Guy-Baker, C. L. (2018). *Developing a TopoSWAT Model and Water Quality Index Framework for the Karst-Dominant Falling Water River Watershed* (Master's thesis, Tennessee Technological University).
13. Sajib, A. M., Diganta, M. T. M., Rahman, A., Dabrowski, T., Olbert, A. I., & Uddin, M. G. (2023). Developing a novel tool for assessing the groundwater incorporating water quality index and machine learning approach. *Groundwater for Sustainable Development*, *23*, 101049.

14. Alarcón, A. P. G., Elizondo, M. D. S. G., Vergara, I., Lagos, M. D., & Herrera, M. T. A. (2021). WATER QUALITY INDICES IN MÉXICO AND COLOMBIA. EVOLUTION, CRITERIA AND CHALLENGES. *Ingeniería Investigación y Desarrollo: I2+ D*, *21*(2), 1-18.

15. Vân, Đ. T. B. *Modification of the Canadian Water Quality Index: Case Study Assessment of Groundwater Quality in Mekong Delta, Vietnam* (Doctoral dissertation, Dissertation, Darmstadt, Technische Universität Darmstadt, 2020).

16. Van Bich, D. T. (2020). Modification of the Canadian Water Quality Index: Case Study Assessment of Groundwater Quality in Mekong Delta, Vietnam.

17. Talukdar, P., Kumar, B., & Kulkarni, V. V. (2023). A review of water quality models and monitoring methods for capabilities of pollutant source identification, classification, and transport simulation. *Reviews in Environmental Science and Bio/Technology*, *22*(3), 653-677.

18. Singh, B., Sihag, P., Singh, V. P., Sepahvand, A., & Singh, K. (2021). Soft computing technique-based prediction of water quality index. *Water Supply*, *21*(8), 4015-4029.

19. Egbueri, J. C., & Agbasi, J. C. (2022). Combining data-intelligent algorithms for the assessment and predictive modeling of groundwater resources quality in parts of southeastern Nigeria. *Environmental Science and Pollution Research*, *29*(38), 57147-57171.

20. Machiwal, D., Cloutier, V., Güler, C., & Kazakis, N. (2018). A review of GIS-integrated statistical techniques for groundwater quality evaluation and protection. *Environmental Earth Sciences*, *77*(19), 681.

21. Islam, M. S., & Mostafa, M. G. (2022). Suitability of water quality index methods for assessing groundwater quality in the Ganges River basin area. *H2Open Journal*, *5*(2), 198-220.

22. García-Avila, F., Loja-Suco, P., Siguenza-Jeton, C., Jiménez-Ordonez, M., Valdiviezo-Gonzales, L., Cabello-Torres, R., & Aviles-Anazco, A. (2023). Evaluation of the water quality of a high Andean Lake using different quantitative approaches. *Ecological Indicators*, *154*, 110924.

23. Hussein, E. E., Jat Baloch, M. Y., Nigar, A., Abualkhair, H. F., Aldawood, F. K., & Tageldin, E. (2023). Machine learning algorithms for predicting the water quality index. *Water*, *15*(20), 3540.

24. Singh, P. K., & Saxena, S. (2018). Towards developing a river health index. *Ecological Indicators*, *85*, 999-1011.

25. Sajib, A. M., Diganta, M. T. M., Moniruzzaman, M., Rahman, A., Dabrowski, T., Uddin, M. G., & Olbert, A. I. (2024). Assessing water quality of an ecologically critical urban canal incorporating machine learning approaches. *Ecological Informatics*, *80*, 102514.

26. Li, Y., Wang, X. K., Zhang, H. Y., Wang, J. Q., & Li, L. (2021). An integrated regional water quality assessment method considering interrelationships among monitoring indicators. *Environmental Monitoring and Assessment*, *193*, 1-20.

27. Pangaliyev, Y., Akbayeva, L., Mamytova, N., Shamshedenova, S., Onerkhan, G., & Mkilima, T. (2024). Assessment of Water Regime, Management and Quality Analysis Based on Water Quality Indices–A Case of Karaganda Region, Kazakhstan. *Polish Journal of Environmental Studies*, *33*(1), 781-801.

28. Betis, H., St-Hilaire, A., Fortin, C., & Duchesne, S. (2020). Development of a water quality index for watercourses downstream of harvested peatlands. *Water Quality Research Journal*, *55*(2), 119-131.

29. Hai, T., Ahmadianfar, I., Halder, B., Heddam, S., Al-Areeq, A. M., Demir, V., ... & Yaseen, Z. M. (2024). Surface water quality index forecasting using multivariate complementing approach reinforced with locally weighted linear regression model. *Environmental Science and Pollution Research*, 1-25.

30. Wibowo, M. A., Andarini, P., Kurniatama, D. V. P., & Humaira, N. G. (2024). Methods for Determining the Water Quality Index in Developing Asian Countries: A Review. *Ecological Engineering & Environmental Technology*, *25*.

31. Mogane, L. K., Masebe, T., Msagati, T. A., & Ncube, E. (2023). Water Quality Index Models for Lotic and Lentic Ecosystems-A Systematic Review.

32. Talal, M., Alamoodi, A. H., Albahri, O. S., Albahri, A. S., & Pamucar, D. (2023). Evaluation of remote sensing techniques-based water quality monitoring for sustainable hydrological applications: an integrated FWZIC-VIKOR modellingapproach. *Environment, Development and Sustainability*, 1-45.

33. Pronk, T. E., Hofman-Caris, R. C. H. M., Vries, D., Kools, S. A. E., ter Laak, T. L., & Stroomberg, G. J. (2021). A water quality index for the removal requirement and purification treatment effort of micropollutants. *Water Supply*, *21*(1), 128-145.

34. Das, C. R., & Das, S. (2023). Assessment of surface water quality for drinking by combining three water quality indices with their usefulness: case of Damodar River in India. *Water, Air, & Soil Pollution*, *234*(5), 327.

35. Benoit, K. K., Séraphin, K. K., Félix, K. K., Charles, B. K., Martin, K. K., & Brou, D. (2020). Indices of water quality and pesticide monitoring survey in lake of Western Côte d'Ivoire and its human risk assessment. *Natural Resources*, *11*(2), 46-69.

36. Shams, M. Y., Elshewey, A. M., El-Kenawy, E. S. M., Ibrahim, A., Talaat, F. M., & Tarek, Z. (2024). Water quality prediction using machine learning models based on grid search method. *Multimedia Tools and Applications*, *83*(12), 35307-35334.

37. Valdiviezo Gonzales, L. G., García-Avila, F., Loja-Suco, P., Siguenza-Jeton, C., Jiménez-Ordoñez, M., Valdiviezo-Gonzales, L., ... & Aviles-Añazco, A. (2023). Evaluation of the water quality of a high Andean lake using different quantitative approaches.

38. Liu, Y., & Li, L. (2023). Multiple Evaluations of the Spatial and Temporal Characteristics of Surface Water Quality in the Typical Area of the Yangtze River Delta of China Using the Water Quality Index and Multivariate Statistical Analysis: A Case Study in Shengzhou City. *International Journal of Environmental Research and Public Health*, *20*(4), 2883.

39. Ghaemi, Z., & Noshadi, M. (2022). Surface water quality analysis using multivariate statistical techniques: a case study of Fars Province rivers, Iran. *Environmental Monitoring and Assessment*, *194*(3), 178.

40. Lukhabi, D. K., Mensah, P. K., Asare, N. K., Pulumuka-Kamanga, T., & Ouma, K. O. (2023). Adapted Water Quality Indices: Limitations and Potential for Water Quality Monitoring in Africa. Water 2023, 15, 1736.

41. Roșca OM (2020) Impact of anthropogenic activities on water quality parameters of glacial lakes from Rodnei mountains, Romania. Environ Res 182:109136. https://doi.org/ 10.1016/j.envres.2020.109136

42. Nair HC, Joseph A, Padmakumari Gopinathan V (2020) Hydrochemistry of tropical springs using multivariate statistical analysis in Ithikkara and Kallada river basins, Kerala, India. Sustain Water Resour Manag 6:1–21. https://doi.org/10.1007/s40899-020-00363-1

43. Teodorof L, Ene A, Burada A, Despina C, Seceleanu-Odor D, Trifanov C, Ibram O, Bratfanof E, Tudor MI, Tudor M, Cernisencu I, Georgescu LP, Iticescu C (2021)

Integrated assessment of surface water quality in Danube River Chilia branch. Appl Sci 11:9172. https://doi.org/10.3390/ app11199172

44. Udeshani WAC, Dissanayake HMKP, Gunatilake SK, Chandrajith R (2020) Assessment of groundwater quality using water quality index (WQI): a case study of a hard rock terrain in Sri Lanka. Groundw Sustain Dev 11:100421. https://doi.org/10.1016/j.gsd.2020.100421

45. Betis H, St-Hilaire A, Fortin C, Duchesne S (2020) Development of a water quality index for watercourses downstream of harvested peatlands. Water Qual Res J 55:119–131. https://doi.org/10.2166/wqrj.2020.007

46. Cristable RM, Nurdin E, Wardhana W (2020) Water quality analysis of Saluran Tarum Barat, West Java, based on National Sanitation Foundation-Water Quality Index (NSF-WQI). IOP Conf Ser: Earth Environ Sci 481:012068. https://doi.org/10.1088/1755-1315/481/1/ 012068

47. En-nkhili H, Najy M, Etebaai I, Talbi FZ, El Kharrim K, Belghyti D (2020) Application of water quality index for the assessment of Boudaroua lake in the Moroccan pre-rif. Conference GEOIT4W-2020: 1–5. https://doi.org/10. 1145/3399205.3399248

48. Nong X, Shao D, Zhong H, Liang J (2020) Evaluation of water quality in the South-to-North Water diversion project of China using the water quality index (WQI) method. Water Res 178:115781. https://doi.org/10.1016/j.watres. 2020.115781

49. Hai, T., Ahmadianfar, I., Halder, B., Heddam, S., Al-Areeq, A. M., Demir, V., ... & Yaseen, Z. M. (2024). Surface water quality index forecasting using multivariate complementing approach reinforced with locally weighted linear regression model. Environmental Science and Pollution Research, 1-25.

50. Benoit, K. K., Séraphin, K. K., Félix, K. K., Charles, B. K., Martin, K. K., & Brou, D. (2020). Indices of water quality and pesticide monitoring survey in lake of Western Côte d'Ivoire and its human risk assessment. Natural Resources, 11(2), 46-69.

Importance and future of AI in the Advancement of science and Engineering in the lives of school dropout students and street children.

Dr. Shashikant Singh Boudh
Editor-Good Morning Nagina (National Newspaper)
President- Ucchtar Madhyamik Shiksha Mandal
Delhi Nagina (Bijnor) Uttar Pradesh-246762
Mobile No- 8923137999
Email-Sksingh999s@gmail.com

On one hand, schemes like Right to Education are being brought across the country to educate children, on the other hand, a large number of children are also leaving school. In this he is talking about taking special steps for the education of deprived children. To find out the exact number of out-of-school children in the country and to send them to school, Uttar Pradesh Education Board has resolved to improve education with AI. Less number of students are being sent to schools. Among the total children not going to school, the number of boys is more while the percentage of girls is less. Not only this, the number of children suffering from some kind of disability among those not going to school is very high. Is included in the school. Therefore, it is necessary to find out the children who do not go to school across the country. The intellectuals stressed on adopting similar methods and including some more recommendations in the right education. The number of children studying in lower classes within the schools should be found out. And if the study is conducted again after 3-4 years, then it will be known that how many children have dropped out of school. While promoting the School Chale campaign or conducting any survey, apart from the children living in children's homes, children roaming at railway stations and traffic signals will also be included. They should also be included and their services should also be taken. It was demanded that these should be included in the Act as soon as possible. The Board has made a written proposal to include all these things in the Education Act. Which will be proposed in the Assembly/Lok Sabha.

Importance of education:-

(Ai) Science education is a very important subject for everyone to achieve success in life and do something different. It helps us to face challenges in difficult times of life. The knowledge gained during the entire learning process makes all of us and every individual self-reliant in our life. It opens up various avenues for opportunities to achieve better prospects in life which will lead to career growth with the use of AI. Promote the importance of education in rural areas Many awareness campaigns are being run by the government to give. It is present in all individuals in the societyBrings the feeling of equality and also promotes the development and growth of the country.Better education is very important for everyone to move ahead in life and achieve success. (Ai) not only develops self-confidence in us but also helps in building our personality. School education plays a great role in everyone's life. The entire education system is divided into three parts like primary education, secondary education and higher secondary education. All levels have their own special importance and place. We all want to see our children going towards success, which is possible only through good and proper education. The importance of education has increased a lot in today's society. There are many uses of education but there is a need to give it a new direction (Ai). Education should be of such a nature that a person can become familiar with his environment. Education is a very essential tool for a bright future for all of us. We can achieve anything good by using this means of education in our life. High level of education helps people in gaining social and family respect and creating a distinct identity. The time of education is a very important time for everyone socially and personally. This is why we (AI) scientistsEducation holds so much importance in our lives. (Ai) Scientific education in any big and famous university There is a very low fee for entry. Other small education boards are also providing education to promote skills in a particular field. Education is equally important for both men and women, because a healthy and educated society is created by both together. Apart from being an essential tool for a bright future, it also plays a very important role in the development and progress of the country. In this way, appropriate education creates a bright future for both. It is only the educated leaders who build a nation and lead it towards the path of success and progress. Education works to make people better and gentler to the extent possible.

(Ai) Modern education system:-

Good education serves many purposes in life such as: promoting personal growth, enhancing social status, improving social health, economic progress, success of the nation, setting goals in life, making us aware about social issues and Providing solutions to solve environmental problems and other social issues etc. Due to the use of distance education system, nowadays the education system has become very simple and easy. The modern education system has been able to completely remove the issue of illiteracy and equality among different castes, religions and tribes. That he should be literate and trained, that is why education has become very important in our lives today. That is why you have to remember that education is very important for us, because of this we get respect in our society so that we can live with our heads held high in the society. Education serves to develop the mind of the people at a higher level and helps in removing all the discrimination among the people in the society. It helps us to become good studies and develops intelligence to understand every aspect of life. It also helps us in understanding all the human rights, social rights, duties and responsibilities towards the country.

With the main role of education (Ai) :-

Education plays a major role in the modern technological world. Nowadays, there are many ways to increase the level of education. The entire system of education has now been changed. Now after 12th class, we can continue studies along with job through distance education programs. Through distance education we can easily take admission in any big and famous university at very low fees. Other smaller institutes are also providing education to promote skills in particular field. (Ai) Education can also be obtained through distance education system. Education is the essential tool for a bright future for all of us. We can achieve anything good in life by using this tool of education. Higher levels of education help people gain social and family respect and a distinct identity. The time of education is a very important time for everyone socially and personally. It takes a person to a different level in life and develops a sense of goodness. Education provides the ability to solve any major family, social and even national and international problems. None of us can ignore the importance of education in every aspect of life. It turns the mind positive and removes all mental and negative thoughts.

(Ai) What is education?

It changes people's thinking by bringing positive thoughts and removing negative thoughts. We study etc. in our childhood itself. Education makes us more civilized and better educated. It helps us to achieve better position in society and envisioned position in job. It enables us to become a good doctor, engineer, pilot, teacher etc., whatever we want to be in life. Regular and proper education takes us towards success by making us aim in life. The education system of earlier times was much more difficult than today. Not all castes could get education as per their wish. Due to high fees, it was very difficult to get admission in a prestigious college. But now, it has become very easy and simple to move ahead by getting education through distance education (.(Ai) Education is the essential tool for a bright future for all of us.

(Ai) It is equally necessary for both men and women.

(Ai) Good education serves many purposes in life.

(Ai) Education plays a role in improving the minds of people at a higher level.

(Ai) The time of education is a very important time for everyone socially and personally.

(Ai) Regular and proper education leads us to success by making us aim in life.

(Ai) Higher education is very important for everyone to get good and technical jobs.

(Ai) Education is very important for us, because of this we get respect in our society.

(Ai) Education holds so much importance in our life.

Importance of education in rural areas:-

Home is the first place to get education and parents are the first teachers in everyone's life. In our childhood, we receive our first lessons of education from our home especially from our mother. Our parents tell us the importance of education in life. When we turn 3 or 4 years old, we are sent to school for proper, regular and less bounded studies, where we become well educated individuals in our life through the efforts of our parents and teacher. They are really our well wishers, who helped us in taking our life towards success. Nowadays, many government schemes are being run to promote the education system so that it is possible for everyone to reach suitable education. Many advertisements are shown on TV and newspapers to show the importance and benefits of education to the people in rural areas because people in backward rural areas do not want to study due to poverty and incomplete information towards education. Which is accomplished in the present time through AI education.

Education for the poor and middle class:-

Earlier, the education system was very expensive and difficult, poor people were not able to pursue higher education after 12th class. There was a lot of difference and inequality among people in the society. The upper caste people received good education and the lower caste people were not allowed to get education in school or college. However, now changes have been made on a large scale in the entire process and subject of education. In this regard, many rules and regulations have been implemented by the Government of India to make the education system accessible and less expensive for all. Most importantly, Ai distance education system has made higher education affordable and accessible, so that the backward areas, poor and middle class people can get equal education and success in the future. The system has made higher education affordable and accessible, so that backward areas , the poor and middle class people should have equal opportunities to achieve education and success in the future. Well educated people are the strong pillars of a country and help in taking it forward in the future. Like this. Ai education is the tool which helps in life Makes all impossible situations possible in society and nation.

(Ai) Education will prove to be an essential tool for a brighter future:-

AI education is the essential tool for a bright future for all of us. We can use AI education in life. You can achieve anything by using this tool. A high level of education helps people in social and Helps in creating family respect and a separate identity. Ai Education time is a very important time for everyone socially and personally. It takes a person to a different level in life and develops a sense of goodness. AI education provides the ability to solve any major family, social and even national and international problems. None of us can ignore the importance of education in every aspect of life. It turns the mind positive and removes all mental and negative thoughts. AI education works to develop the minds of people on a large scale and along with this it also helps in removing all the discrimination among the people in the society. It helps us to become good studies and develops intelligence to understand every aspect of life. It helps us in understanding all the human rights, social rights, duties and responsibilities towards the country. Reasons for gender inequality:- Although in the cities at present no difference is considered between boys and girls, but even today there are villages where educating girls is considered useless. The mentality of poor families is still the same as the old one. Although the age for marriage of girls in the country has been fixed at 18 years, even today in poor rural families, girls are married at the age of 14-15. In such a situation, no girl is able to reach even the 5th class quickly. Poor nutrition:- The head of a poor family either works as a laborer or does farming on a small piece of land to support his family. In such a situation, some days the stove in his house is not even lit. Children of such families somehow spend many days eating dry rotis. For the development of the brain, proper nutritional elements are required which we get only from food. Due to children not getting nutritious food, their brain development stops, after which no matter how well we teach them, no matter how many facilities are given to them, they will not be able to understand anything.

(Ai) Less expenditure on education:-

Children's education is not complete merely by preparing children and sending them to school every day. It is also very important for children to get exposure to new books and technologies. A high level education requires considerable expenditure at present. Apart from studies, money was also spent on honing other talents. Is required. It is very important for children to acquire knowledge of AI, computers etc. since childhood, which is taken under the expenditure of education AI. At present, very little money is spent on AI education. And

Time is also saved.

Family responsibilities in childhood itself:-

In poor families, children start realizing family responsibilities as soon as they turn 13-14 years old. Since childhood, they have been living with limited resources. With increasing age, like everyone else, their needs also increase and to fulfill them, they do not see any other solution except earning their own money. The eldest father in the house also has the responsibility of the entire family. As children grow up, the expenses also increase and it is very difficult for just one member of the family to bear this expense. As a result, children start doing small jobs to earn their own money and leave their studies midway.

Due to economic situation:-

At present, the cost of any higher level education is a huge sum for a poor family. A child from a poor family who is very good in studies, after studying till class 10-12, his studies stop there due to the financial condition of the house. Some children who are more promising try to maintain their studies by taking small tuitions but they are not able to pay the fees to get higher level education. No matter how bright a child from a poor family may be, in today's time it proves to be very difficult for him to study as a doctor or engineer.

Decrease in interest towards studies but rise of better curiosity with the use of (Ai):-

Seeing the unbalanced condition of the house since childhood, the mentality of the children who grow up in poor families also gets formed in the same way. As they grow up, they also lose interest in studies. Such children have always seen people around them struggling just for two meals a day. In poor society, due to never getting any environment for studies, children after some time start considering studies as useless and finally leave studies and start looking for ways of earning. Poverty is like a disease which is making a person hollow from within. This poverty Providing higher education to children in India is an even more challenging task. Although now many such government facilities are being provided to the poor, so that their children do not face any problem in getting education, but the sad thing is that nowadays the mentality of the poor has become very narrow. They themselves decide that the poor cannot get higher education. Such families must take inspiration from the lives of Shastriji and AmbedkarSaheb. Poverty has been a serious and long-term issue in India. It is an extreme situation when a person feels lack of essential items needed to continue life like shelter, adequate food, clothing, medicines, etc. Some of the common causes of poverty are susceptible, fatal and epidemic diseases, natural disasters, low agricultural production, lack of employment, casteism in the country, illiteracy, gender inequality, environmental problems, changing trends of the economy in the country, lack of proper education, untouchability, limited Or inadequate access to their rights, political violence, organized crime, corruption, lack of motivation, laziness, outdated social beliefs etc. Poverty in India can be reduced by following effective solutions, however individual efforts of all citizens are required.

Objectives of Education:-

Education is a means of increasing knowledge. It is a medium of cultural life. Is the creator of character. It is the door to livelihood. Along with the art of living life making full use of one's abilities, it is also a guide to personal development. The criterion of human development is knowledge. Knowledge trains the intellect and gives rise to ideas in the mind. This thoughtfulness is the means to success in all areas of life. Be it physical or mental, illness or disease, problems or crisis, all are solved with the key of knowledge.

Education and the need for education to acquire knowledge:-

Acquiring knowledge is an important objective of education. By acquiring knowledge through modern civilization education Has developed. Neither knowledge is complete education in itself, nor is it the ultimate objective of education, it is only a part of education and a means. The aim of internal education should be to acquire cultural knowledge so that humans can become civilized, polite, moderate, develop literature, music and art etc. By making life valuable we can raise the standard of living. Also, cultural heritage can be passed on to the coming generations.

Education for character:-

Character means all those things which are done in the form of conduct, behavior etc. According to Plutarch, character is merely a long-term habit. Valmiki says, it is only by the character of a man that it is known whether he is noble or non-noble, brave or arrogant, pure or impure. There are two types of character, good and bad. Good character is the beauty of the society. Education bears the responsibility of its creation. In the words of Gandhiji, purity of character is the foundation of solid education. Therefore Dr. D.N. Khokhla says, the aim of education should be emotional and moral development. A good engineer or doctor is useless if he does not have moral qualities. Because, a person with poor character only carries the burden of knowledge. Real education develops the virtues and perfection inherent in human beings.

With Education for Business (Ai):-

Meaning of business. Means of subsistence. This means that education should have so much power that it should be arghakari, that is, it can guarantee the livelihood of an educated person. In the words of Gandhiji, true education should be an insurance against unemployment. Teaching the Art of Living:- The four objectives of education written above are one-sided for acquiring knowledge, culture, character and profession, and are not complete in themselves. Life requires education in the art of living; education should provide the ability and capability to learn the complex process of life and make life happy in times of sorrow, suffering and adversity. Spencer supports a broad objective in education i.e. holistic development in all aspects of life. He rejects bookishness and supports activities that help in running a family, maintaining social and economic relationships and developing emotional development. After success in these activities, the person becomes ready for the next life. Co-education:- Co-education means that boys and girls of all ages sit together in the same room and study together. Get education. There should be no need to have separate classes for boys and girls. This type of education has not been practiced in India since ancient times. Till a few years ago, women were not interested in getting this type of education. But just a century ago, emphasis was being laid on giving education to women. In view of this, in today's materialistic era, women have also recognized the importance of co-education. She has understood this and hence she does not have much hesitation in getting co-education. Co-education improves the psychological development of boys and girls. There is a big difference between Indian social values and western values. Those whose culture is respected in western countries. In India they are considered wrong. For example, in India, parents used to choose the bride and groom for their son and daughter, whereas in western countries, teenagers choose their own life. That is why co-education was considered mandatory there. From there, these people can get acquainted with each other and find Swadhi for themselves. There is no such compulsion in India. But nowadays western customs have entered India also. Due to the influence of western education, co-education started being used in India also. Love marriages started happening here too. Co-educational schools started opening in Indian society also. Today equality between men and women is prevalent, so why should co-education be opposed? Education not only increases a person's self-esteem but also helps in building a good personality. Only education helps in understanding the ups and downs in life and solving them. Any kind of problem can be solved easily through education.

Importance of sports in education and its support and quality:-

Sports are very important in human life. There are many problems in our life and wem Suffer from stress. And sports relieves us from problems, stress and worries. Sports have special importance in education. It is an essential part of education. Sports lead to overall development of a human being. Our body becomes healthy, fit and healthy. Sports bring freshness to the mind. There are arrangements for many types of sports in schools and colleges. Just by playing games while studying, students become stress free. The first school of any person is his family. And mother has been called the first teacher. Education is that weapon. With the help of which one can face even the biggest difficulties. It is education through which we know the difference between right and wrong. Many essays have been written on education. Will be written further also. Its importance can be gauged from this, it will be okay if you don't get one meal, but you should definitely get education. It is the right of every living being to get education. AI in Play with Modern Education System The Importance of AI has also increased rapidly Education is an important tool, which is very useful in everyone's life. Education is what makes us the smartest creatures on earth. It empowers human beings and prepares them to face the challenges of life efficiently.The word education Is derived from the Sanskrit word 'Shiksha' Ghatu, which means to learn or teach. That is, the process by which study and teaching takes place is called education.

With various definitions of education (Ai):-

According to Geeta, “Sa VidyaVimukte”. That is, education or knowledge is that which frees us from bondages and Expand on every aspect of ours. According to Tagore, “Our education has become based on self-interest, driven by the narrow objective of passing exams, a means of getting a job as soon as possible, which is being imparted in a difficult and foreign language. This has caused us to lose knowledge of rules, definitions, facts and It neither gives us time nor inspires us to stop and absorb what we have learned. According to Mahatma Gandhi, “True education is that which develops and motivates the spiritual, intellectual and physical aspects of children. In this way we can say in essence that according to him education should Meaning all-round development: According to Aristotle, “Education develops the powers of man, especially the mental powers so that he becomes capable of contemplating the ultimate truth, Shiva and the beautiful.” To make education accessible, there is a need to spread educational awareness in the country. But it is incomplete without analyzing the importance of education.

Right to education:-

We can fulfill our dreams only through education. Can give new condition and direction to life. Without education we cannot achieve anything. Nowadays, everyone needs to earn a living, for which it is very important for you to be educated. Today’s generation will not fare well without education Can. Education creates employment opportunities. Today only that country comes in the category of most powerful, which has the power of knowledge. Gone are the days when battles were fought with swords and guns, now big battles are won only by using brains without bloodshed. Education Well, everyone has the right to get education. But now a law has been made on this. This means that now it is mandatory for everyone to teach their children. This law was brought in 2009 by the name of ‘Free and Compulsory Children’s Education Act’. Right to education is one of the fundamental rights mentioned in the Constitution of our country.Is.The 46th Constitutional Amendment, 2002 provides for free and compulsory education to all children up to fourteen years of age as a fundamental right. Right to Education (RTI Act) has been added to 21A of the Constitution. It is effective from April 1, 2010. The following things are stated in the RTI Act. • According to this legislation, now there is a right to provide free and compulsory education to children in any government school.Provisions • Right to Education Act to every primary school (primary school + middle school) in India These norms have to be followed to maintain the minimum standards set by the Govt. There is a rule to give admission to the children who due to some reason are not able to go to school on time, in the appropriate class. It also appoints trained teachers. It provides for the development of the curriculum in accordance with the values mentioned in the Constitution. And is committed to holistic development of the child, enhancing the knowledge, potential and talent of the child and freeing the child from fear, trauma and anxiety through child friendly system and child centered knowledge system. The process by which study and teaching takes place is called education.

• Everyone has the right to get education.

• It is an important tool, which is very useful in everyone's life.

• This makes man the smartest creature on earth.

Employment opportunities are created only through education.

• Only through education can we fulfill our dreams.

• Without education we cannot achieve anything.

• Education plays an important role in the society. It also helped in the spread of literature. Temples and community centers played the role of schools. Later, the Gurukul system of education came into existence.

Modernization (Ai) Impact on Education:-

Education plays an important role in the society. Education itself expands our knowledge, transfers it to students and promotes new knowledge. Modernization is a process of socio-cultural change. It is a series of changes involving values, norms, institutions and structures. According to the sociological approach, education is not according to the individual needs of the individual, rather it arises from the needs of that society. In which the person is a member.In a stable society, the main function of the educational system is to transmit cultural heritage to new generations. But in a changing society, its form keeps changing from generation to generation. And in such a society, the educational system should not only be taken as a cultural heritage but should also help in preparing the youth to adjust to the changes. And this lays the foundation for future possibilities.Modern educational institutions prepare skilled people, whose scientific and technical knowledge leads to the industrial development of the country.

Other values like individualism and universalistic ethics etc. can also be developed through education. Thus education can be an important weapon of modernization. The importance of education can be realized from the fact that all modern societies insist on universalization of education and in ancient days, education was concentrated for a particular group. But with the modernization of education, now everyone has the facility to get education irrespective of their caste, religion, culture and economic background. The effect of modernization can be seen in schools also. Modern day schools are completely technicalAre formally equipped with sound tools that help children develop their expertise in a more clear way. Effective facilities provide barrier-free means for persons with disabilities, free from health and environmental hazards, to access shared spaces. The real meaning of education is "learning", which awakens human conscience, the purpose of all education is truth.To reveal and provide every kind of help in practical life, according to Mahatma Gandhi, true education is not in reading books but in the organization of character, the aim of education should be to make man capable of achieving his highest manhood, that to be given such education It should awaken the feeling of 'VasudhaivaKutumbakam' in his mind. In ancient India, this education was given in the Gurukuls and Ashrams of Maharishis located in dense forests, away from the noise and strife of the city. The students remained under the guidance of the Guru till the age of twenty-five years. Used to study properly while serving his feet. Here his all-round development took place, Takshashila and Nalanda universities were the major education centers of the country at that time, then due to foreign invasions the ancient education system of the country was also affected and the complete Indian education system was replaced by the medieval Urdu-Persian mixed education system. Over time, the political influence of the West rapidly declined in India, the whole of India came under the British rulers, the English made arrangements for imparting education in their own way to make their rule permanent. The education system started by Macaulay in 1828 was aimed at producing as many clerks as possible. This policy remained in force throughout the British rule, its remnants are still more or less alive in the Indian education system, due to which even today our education is being proven faulty and impractical. While the education system propounded by Macaulay helped the British in achieving their self-interests, it also helped the Indians in achieving their own ends. They took away the feeling of self-reliance and character building and made them struggle only for subsistence,

Key facts of India's new education policy: Illumination of (Ai) in the National Education Policy:-

India's New Education Policy 2020 has been implemented by making some improvements in the National Education Policy 1986. The Sustainable Development Goals issued by the United Nations have been kept at the center in the new National Education Policy. To achieve this goal, a target of 2030 has been set in the policy. Major change in the pattern of education: This policy talks about implementing a new system of 5+3+3+4 in place of the currently implemented 10+2 school education. In this, education will be provided in Anganwadi for 3 years in the first 5 years and classes 1 and 2 in the remaining 2 years. In the next three years, education will be provided up to class 5, in the next three years up to class 8 and in the remaining 4 years up to class 12. Emphasis on care and education of children: This is the focus of primary education. For this A national framework will be created for preparing the curriculum. Ashramshalas will be started to implement this in many tribal areas. Teacher-Student Ratio: In the new education system, it will be ensured that the teacher-student ratio in every school is less than 30:1 and in areas with a majority of socio-economically deprived children, the PTR of the school is less than 25:1. Digital Infrastructure for Knowledge Sharing (DIKSHA) provides basic literacy and A national repository of quality resources based on numeracy will be made available. Equivalent Multidisciplinary Education and Research Universities (MERUs) Equivalent Multidisciplinary Education and Research Universities (MERUs) will also be developed across the country on the lines of IITs and IMs. Also, an Indian Institute of Translation and Interpretation (IIT I) will be established for the development of various languages.Various goals of the policy: To achieve 100% youth and adult literacy by 2030, Quality education for all learners by 2040 and The top priority of the education system will be to achieve universal basic literacy and numeracy in primary schools by 2025. There are many benefits of the current education system, it will not only improve the level of education in the entire country.But after twelve years of education, the student will be able to take up some profession or the other and pursue university education.Crowding will be reduced, but this success can be achieved only when it is accompanied by other reforms like improvement in the examination system, qualified trained teachers, adequate amount of money so that equipment, supporting materials etc. can be made available and most importantly, determination and nation. The spirit of pioneering, and confidence in AI education

AI/ML APPLICATIONS IN AGRICULTURE: RETRIEVING WHEAT CROP TRAITS FROM UAV-BORNE HYPERSPECTRAL IMAGES USING HYBRID MACHINE LEARNING MODELS

Kunal Mahto[1], Subhash Chandra Dutta[2]
[1]PhD Scholar, Department of Computer Science & Engineering, BIT Sindri, Dhanbad, India, kunalm.rs.cse21@bitsindri.ac.in
[2]HOD, Department of Computer Science & Engineering, BIT Sindri, Dhanbad, India, scdutta@bitsindri.ac.in

Subject : Artificial Intelligence

Abstract

In recent years, advances in agricultural operations have been made possible by the deployment of machine learning (ML) and artificial intelligence (AI) approaches. This shift has led to an increase in the visibility of crop management and yield optimisation in recent years. This study looks into the application of hybrid machine learning models to extract wheat crop characteristics from hyperspectral photos captured by unmanned aerial vehicles (UAVs). For this experiment, the ARTMO software package is employed. The workflow presented includes field testing, using unmanned aerial vehicles (UAVs) to take pictures, and using advanced data processing techniques like principal component analysis (PCA), Gaussian process regression (GPR), and PROSAIL simulations. NRMSE values for LAI and CCC are relatively low at 9.7% and 15.9%, respectively. These figures show that overall performance is excellent. These numbers suggest that the uncertainties have significantly decreased and that the retrieval accuracy was fairly good. There is an additional improvement in the model's precision as a result of the addition of non-vegetation spectra to the dataset optimised for AL. This approach provides a scalable, quantitative, and real-time solution for vegetative product surveillance. This approach is what has allowed for this important contribution to the field of precision agriculture.

Keywords: *Ai/Ml Applications, Hybrid Machine Learning Models, Wheat Crop, Uav-Borne Hyperspectral Images.*

1. INTRODUCTION

Lately, the coordination of artificial intelligence (AI) and machine learning (ML) procedures has reformed horticultural practices, especially with regards to crop the executives and yield improvement. One promising application involves the use of unmanned aerial vehicles (UAVs) equipped with hyperspectral imaging technology to capture detailed data on wheat crop traits. These photos give extensive, high-resolution spectrum information that standard approaches cannot match, allowing the extraction of exact agricultural metrics like chlorophyll, leaf area index, and nutrient levels.

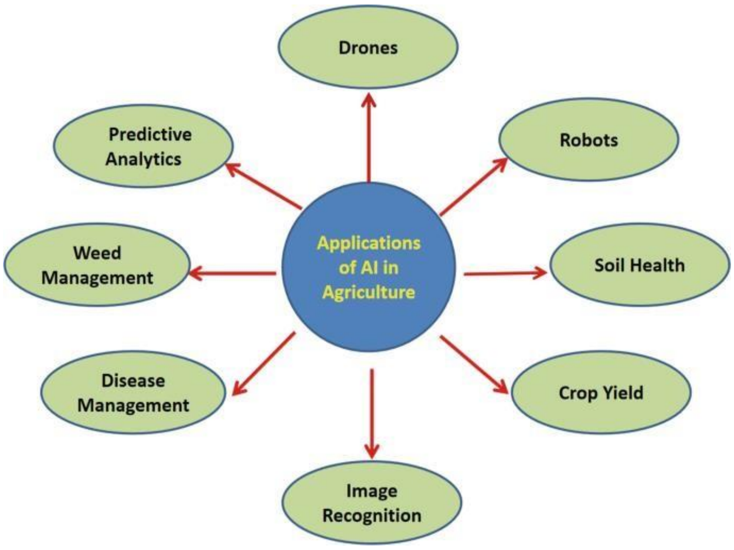


Figure 1: Applications of AI in Agriculture

Researchers may analyse and interpret these complex datasets using hybrid machine learning models that combine deep learning algorithms with statistical methods to give farmers insights into crop productivity and sustainability.

1.1. Wheat Crop Traits and Their Significance

Wheat, a staple crop worldwide, exhibits diverse traits such as leaf chlorophyll content, biomass accumulation, and nitrogen status, which are crucial indicators of crop health and productivity. Monitoring these traits throughout the growing season is essential for optimizing agricultural practices, including irrigation scheduling, fertilizer application, and disease management. Traditional methods of trait assessment, such as manual sampling and laboratory analysis, are labor-intensive, time-consuming, and often provide limited spatial coverage.

1.2. Objectives

The core objectives of the study are as follows:

1. To retrieve wheat crop traits (LAI and CCC) from UAV-borne hyperspectral images using hybrid machine learning models.
2. To implement a workflow with UAV imaging, PROSAIL simulations, GPR, and PCA for precise wheat crop monitoring.

3. To enhance trait retrieval accuracy by integrating non-vegetation spectra into the AL-optimized dataset.

2. LITERATURE REVIEW

Behera et al. (2010) Using the LAI-2000 Plant Canopy Analyzer, the study estimates Jatropha curcas leaf area index (LAI) indirectly. The canopy light transmittance method is essential for agricultural canopy structure and function assessment. The researchers used it to estimate LAI non-destructively, showing how technology-driven approaches can improve agriculture and forestry research precision and efficiency.

Francone et al. (2014) The Ceptometer and Pocket LAI Smart App estimated LAI differently for canopies with varying structures. The data suggest canopy design influences LAI measurements, emphasizing the necessity for canopy-specific methods. Agronomic and ecological LAI estimations may be more accurate and useful. The research optimizes LAI measurement for different canopy topologies and improves agricultural management.

Gonsamo and Pellikka (2008) Researchers corrected canopy LAI slope effects with hemispherical photography. The study indicated significant slope effects in LAI estimates, requiring adjustment. Statistics indicate that LAI estimate should employ topography to better ecological evaluations and management. The research improves hemispherical photography for varied environments, enabling ecological studies and conservation.

Liang et al. (2020) Researchers estimated agricultural Leaf Area Index using spectral feature extraction and hybrid inversion. Due to its accuracy, remote sensing can improve agricultural monitoring and management. This work improves LAI calculation methods to optimise agricultural production evaluations and resource management. Remote sensing aids sustainable agriculture and environmental protection.

3. APPLICATIONS OF AI/ ML IN AGRICULTURE

A portion of the AI/ML based applications in agriculture area are –

3.1. Yield prediction

Machine learning methods estimate crop yield using remote sensing. The majority use Artificial Neural Networks (ANN), then Convolutional Ones. Regularizing Bayes BP, SVR, ELR, RFR, and PLSR operate well. Red edge, canopy chlorophyll, and absorption ratio indices predict crop output. For efficient crop management, ML models with limited training data must estimate multi-stage crop production promptly.

3.2. Pest and diseases detection

Preventing output losses requires early crop disease identification. Accurate estimates require machine and deep learning. MLR models distinguished Maize Dwarf Mosaic Virus, wheat Powdery Mildew Disease, and Kiwifruit Decline Syndrome utilizing UAV information. These findings enable smart farming by recognising diseased crops, reducing pesticide and chemical consumption, and preserving crop quality. Figure demonstrates major deep and machine learning models.

3.3. Weed detection

Pests like weeds reduce crop yield and productivity. Early weed identification in farms is better with ML/DL. Drone, robot, and digital camera RGB images are processed using these algorithms. VGGNet performed worst, whereas SVM and CNN achieved 99% accuracy. This method processes RGB photos from drones, robotics, and digital cameras.

3.4. Soil health management

Food sustainability requires digital soil planning and shrewd supplement expectation. ML strategies like arbitrary woodland and profound learning beat traditional models for soil supplement expectation. Top soil nutrient prediction machine learning methods are compared.

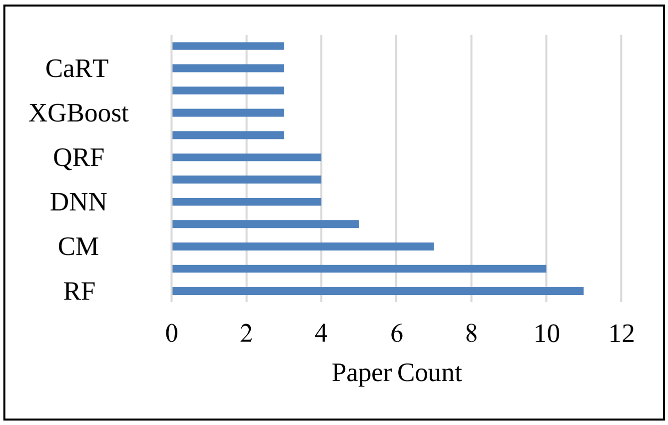


Figure 2: An illustration of the top 12 machine learning models for soil prediction

3.5. Crop quality management

Detecting advances and machine learning algorithms to monitor crop nitrogen and chlorophyll condition is crucial. Radiative exchange methods and machine learning have been combined to consolidate SWIR ghastly hyperspectral information. Using hybrid models, especially radiative exchange models with Gaussian cycle relapse, these limits have been recovered quickly and precisely.

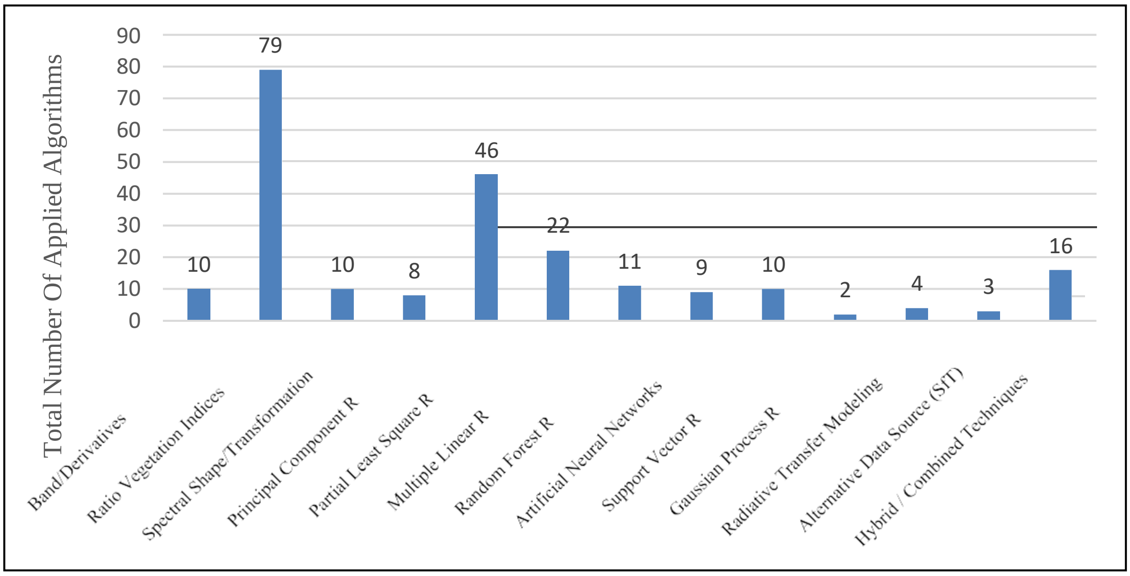


Figure 3: N content estimation using several techniques

3.6. Smart irrigation

The proposed machine learning-based water system design incorporates information from different sources, including UAV and satellite-caught information, soil, and climate data, into a cloud server. This empowers shrewd water system planning, forecasts, and proposals. The architecture recommends the use of supervised, unsupervised, reinforcement, and federated learning models for precise and smart field management.

3.7. Livestock Management

Livestock the board includes disease location, immunization, creation the executives, tracking, and wellbeing checking. These models show the most elevated exactness of more than close to 100% in steers recognizable proof, while profound learning models like ResNet, Origin, DenseNet, and NasNet additionally accomplish more than close to 100% precision.

4. METHODOLOGY

The block chart of the procedure utilized for the proposed approach is shown. The main advances engaged with the workflow are (I) field trial and error, UAV picture procurement, and pre-processing; (ii) PROSAIL reenactments and model assessment; (iii) Gaussian process regression (GPR); (iv) dimensionality decrease utilizing principal component analysis (PCA); and (v) dynamic learning techniques and field check. Each step is explained exhaustively in the resulting segments.

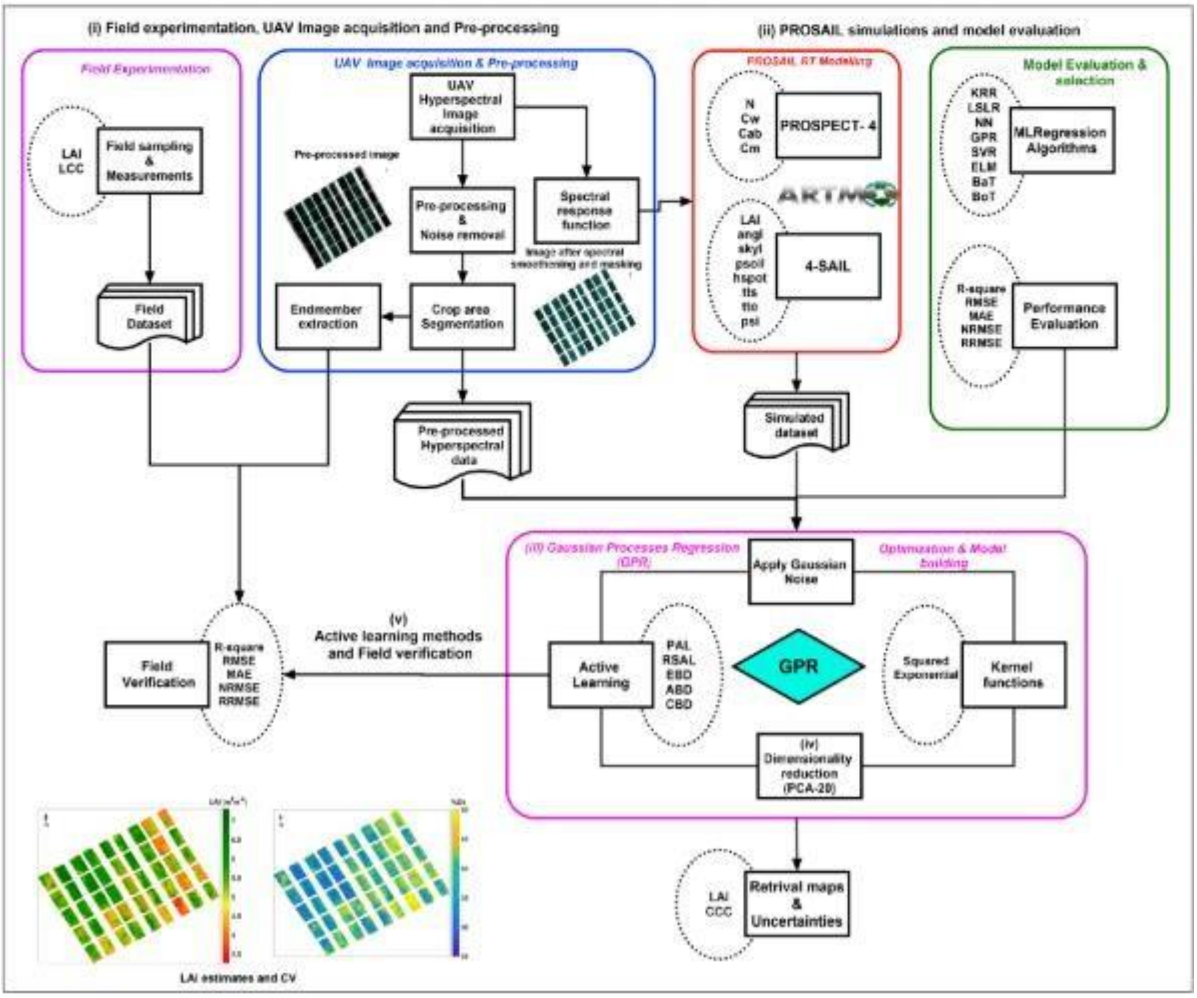


Figure 4: Block schematic of the study's methodology

The Study concentrated on an exploratory wheat crop at ICAR-IARI 228 meters above ocean level. The field has three replications of fifteen plots with five nitrogen grades and three water system systems. DMSO was utilized to anticipate agrarian yield and LAI-2000 plant covering analyzer to gauge LAI. The 21-m-high hyperspectral image has 278 400-1000 nm bands. UgCS

Mission planning software designed the mission route, and Headwall Spectral View and ENVI analysed the gained hypercube.

4.1. Model Evaluation and Selection of GPR

A study assessed eight multivariate models for assessing LAI and CCC involving PCA as a dimensionality decrease technique. GPR beat any remaining models in foreseeing LAI with the most noteworthy R2 worth of 0.887, while KRR was appropriate for anticipating CCC with a R2 worth of 0.8889. MAE values for LAI and CCC were 0.128 and 0.127, separately, while RMSE values were 0.254 and 0.135. The least qualities for assessing LAI and CCC were 1.857% and 0.544%, separately. Both GPR and KRR created more exact and strong outcomes in assessing crop traits.

Table 1: Evaluation of MLRA models' accuracy in obtaining LAI and CCC

S. No.	Model	MLRA	MAE	RMSE	RRMSE (%)	NRMSE (%)
LAI						
1	GPR	0.128	0.254	3.887	1.857	0.887
2	KRR	0.225	0.264	5.176	2.195	0.006
3	NN	0.230	0.343	7.261	3.264	0.099
4	LS	0.290	0.358	7.656	3.466	0.098
5	ELM	0.290	0.432	9.510	4.468	0.089
6	BaT	0.329	0.468	10.570	4.965	0.087
7	BoT	0.414	0.524	11.944	5.715	0.074
8	SVR	0.445	0.559	12.903	6.207	0.068
CCC						
1	KRR	0.127	0.135	2.673	0.544	0.8889
2	GPR	0.142	0.154	3.867	0.857	1.000
3	NN	0.142	0.160	4.373	1.015	1.000
4	LS	0.149	0.160	4.403	1.023	1.000
5	ELM	0.153	0.174	5.226	2.25	0.999
6	SVR	0.186	0.104	7.110	2.710	0.996
7	BaT	0.179	0.210	8.216	2.080	0.996
8	BoT	0.240	0.287	12.632	4.203	0.991

The study shows smooth convergence and the usage of Non-Related Mean Square Error (NRRMSE) over R2 for LAI and CCC retrieval using Artificial Neural Network (ANN) methods. Validating with the field dataset shows that adding fresh samples to each AL iteration decreases RMSE and increases R2. With a few samples, the GPR model performed best, achieving 97 and 119 for LAI and CCC. For LAI, RSAL outperformed other AL methods, whereas EBD excelled for CCC. Low sampling size may cause convergence due to a small number of training data points.

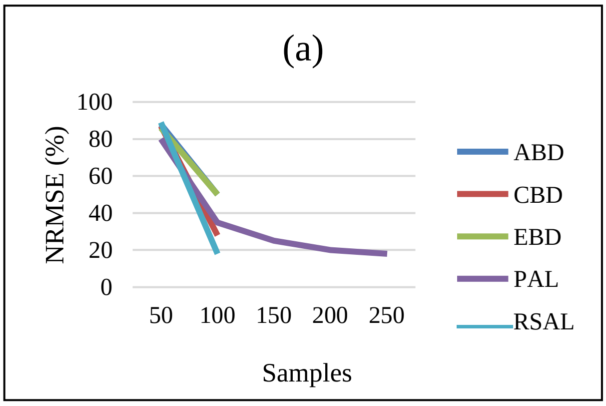


Figure 5 (a): NRMSE (%) is shown graphically for a number of trait estimates made using various AL techniques. (A) The number of samples is indicated by LAI# samples

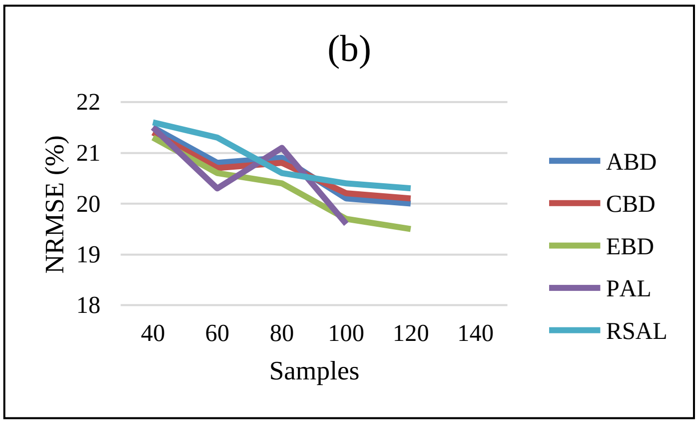


Figure 5 (b): NRMSE (%) is shown graphically for a number of trait estimates made using various AL techniques. (b) CCC. # samples indicates how many samples there are

4.2. Retrieval of LAI and CCC

Preprocessed UAV hyperspectral images was used to calculate and estimate uncertainty using final GPR models. CV, LAI, and CCC retrieval maps were shown. Experimental plots with low LAI and CCC values exhibited red pixels, showing pixel-wise volatility. The plots revealed realistic and appropriate spatial variability with maximum and minimum values. The estimated maps showed no zero or almost zero values, indicating that there were no non-vegetated areas, despite the fact that the GPR models were trained with non-vegetation spectra and trait values set to zero.

5. RESULTS

5.1. Performance and evaluation criteria (including KPIs)

The GPR models' 9.7% and 15.9% NRMSE values after field check demonstrate great recovery precision and more modest planning uncertainties for LAI and CCC. ARTMO, a free programming device, is utilized to recover wheat crop biophysical qualities from UAV datasets using a refined kernel-based and adaptable hybrid procedure.

Table 2 (a): Model Performance Metrics for LAI Prediction

Parameter	LAI
NRMSE	8.58%
RMSE	0.735%
MAE	0.481%
R^2	0.998%
Improved NRMSE	9.7% (from 17.9%)

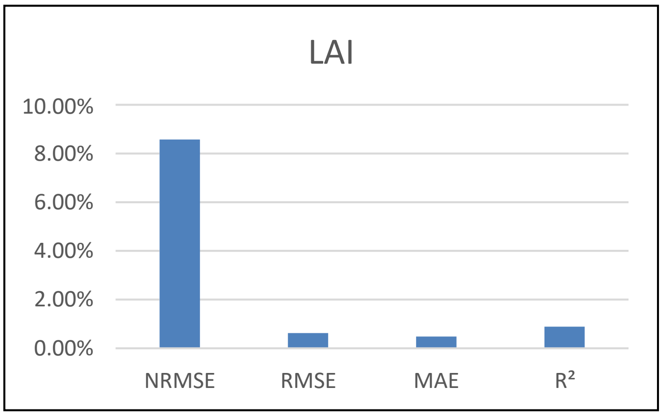


Figure 6 (a): Graphical presentation of the Model Performance Metrics for LAI Prediction

Table 2 (b): Model Performance Metrics for CCC Prediction

Parameter	CCC
NRMSE	15.95%
RMSE	0.668%
MAE	0.423%
R^2	0.767%
Improved NRMSE	15.9 % (from 19.4%)

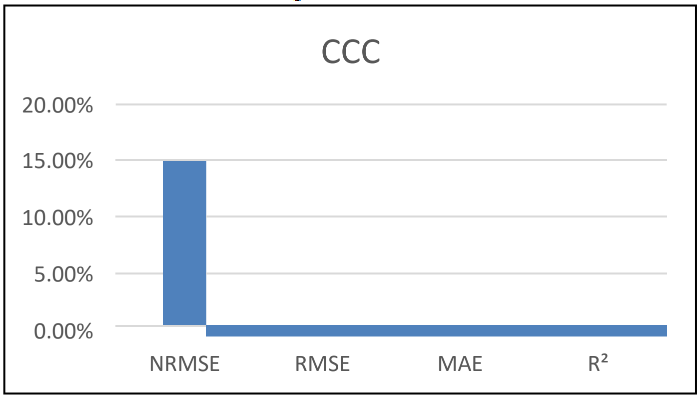


Figure 6 (b): Graphical presentation of the Model Performance Metrics for CCC Prediction

GPR models for LAI and CCC recovery have RMSEs of 0.735 and 0.668. LAI and CCC have MAE upsides of 0.481 and 0.423, individually, and R2 upsides of 0.998 and 0.767. NRMSE upsides of 8.589% and 15.953% for LAI and CCC recovery are the most basic measurable elements. NRMSE further developed essentially subsequent to adding nonvegetation spectra to the AL-upgraded dataset and retraining. From 17.9 to 9.7%, LAI NRMSE improved, and 19.4 to 15.9%, CCC diminished. Five uncovered soil or non-vegetation spectra (10% of in situ estimations) were added to the AL-advanced dataset before crop trait approval.

6. CONCLUSION

AI and ML can upset crop the board and yield enhancement in agriculture. This study shows how hybrid machine learning models can recover wheat crop ascribes from UAV-borne hyperspectral pictures utilizing ARTMO programming. Field tests, UAV picture gathering, and high-level information processing strategies including PROSAIL reenactments, Gaussian process regression (GPR), and PCA are utilized. Brilliant recovery exactness and diminished uncertainties were shown by low normalized root mean square error (NRMSE) upsides of 9.7% for leaf area index (LAI) and 15.9% for canopy chlorophyll content (CCC). The dynamic learning-upgraded dataset including non-vegetation range works on model accuracy, making vegetation item observing adaptable, quantitative, and constant. This strategy further develops accuracy agriculture by checking crop wellbeing precisely and productively.

7. ACKNOWLEDGMENT

We are thankful to the Indian Council of Agricultural Research-Indian Agricultural Research Institute (ICAR-IARI) for giving the exploratory wheat field to this study. Field personnel helped with crop collection and LAI measurements, so thank you. We also thank the Headwall

Nano-Hyperspec hyperspectral camera team for UAV image capture and the UgCS Mission planning software team for mission route planning. Additionally, we thank the ARTMO software developers for their important tool for data processing and analysis. Finally, we thank our colleagues and collaborators for their helpful criticism and assistance throughout the research process.

REFERENCES

1. *Behera, S. K., Srivastava, P., Pathre, U. V., & Tuli, R. (2010). An indirect method of estimating leaf area index in Jatropha curcas L. using LAI-2000 Plant Canopy Analyzer. Agricultural and Forest Meteorology, 150, 307–311.*
2. *Francone, C., Pagani, V., Foi, M., Cappelli, G., & Confalonieri, R. (2014). Comparison of Leaf Area Index Estimates by Ceptometer and PocketLAI Smart App in Canopies with Different Structures. Field Crops Research, 155, 38–41.*
3. *Gonsamo, A., & Pellikka, P. (2008). Methodology Comparison for Slope Correction in Canopy Leaf Area Index Estimation Using Hemispherical Photography. Forest Ecology and Management, 256, 749–759.*
4. *Liang, L., Geng, D., Yan, J., Qiu, S., Di, L., Wang, S., Xu, L., Wang, L., Kang, J., & Li, L. (2020). Estimating Crop LAI Using Spectral Feature Extraction and the Hybrid Inversion Method. Remote Sensing, 12(21), 3534.*
5. *Nie, C., Shi, L., Li, Z., Xu, X., Yin, D., Li, S., & Jin, X. (2023). A Comparison of Methods to Estimate Leaf Area Index Using Either Crop-Specific or Generic Proximal Hyperspectral Datasets. European Journal of Agronomy, 142, 126664.*
6. *Orlando, F., Movedi, E., Paleari, L., Gilardelli, C., Foi, M., Dell'Oro, M., & Confalonieri, R. (2015). Estimating Leaf Area Index in Tree Species Using the Pocket LAI Smart App. Applied Vegetation Science, 18, 716–723.*
7. *Parry, C., Blonquist, J. M., & Bugbee, B. (2014). In Situ Measurement of Leaf Chlorophyll Concentration: Analysis of the Optical/Absolute Relationship. Plant, Cell & Environment, 37, 2508–2520.*
8. *Shibghatallah, M. A. H., Khotimah, S. N., Suhandono, S., Viridi, S., & Kesuma, T. (2013). Measuring Leaf Chlorophyll Concentration from Its Color: A Way in Monitoring Environment Change to Plantations. In Proceedings of the AIP Conference, Kabupaten Sumedang, Indonesia, 7–9 May 2013 (Vol. 1554).*

9. *Thomas, H., & Howarth, C. J. (2000). Five Ways to Stay Green. Journal of Experimental Botany, 51, 329–337.*
10. *Tripathi, R., Sahoo, R. N., Gupta, V. K., Sehgal, V. K., & Sahoo, P. M. (2013). Retrieval of Leaf Area Index Using IRS-P6, LISS-III Data and Validation of MODIS LAI Product (MOD15 V5) over Trans Gangetic Plains of India. Indian Journal of Agricultural Sciences, 83, 380–385.*
11. *Tripathi, R., Sahoo, R. N., Gupta, V. K., Sehgal, V. K., & Sahoo, P. M. (2013). Developing Vegetation Health Index from Biophysical Variables Derived Using MODIS Satellite Data in the Trans-Gangetic Plains of India. Emirates Journal of Food and Agriculture, 25, 376–384.*
12. *Verrelst, J., Rivera, J. P., Gitelson, A., Delegido, J., Moreno, J., & Camps-Valls, G. (2016). Spectral Band Selection for Vegetation Properties Retrieval Using Gaussian Processes Regression. International Journal of Applied Earth Observation and Geoinformation, 52, 554–567.*
13. *Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote Sensing for Agricultural Applications: A Meta-Review. Remote Sensing of Environment, 236, 111402.*
14. *Yan, G., Hu, R., Luo, J., Weiss, M., Jiang, H., Mu, X., Xie, D., & Zhang, W. (2019). Review of Indirect Optical Measurements of Leaf Area Index: Recent Advances, Challenges, and Perspectives. Agricultural and Forest Meteorology, 265, 390–411.*
15. *Zhu, X., Yang, Q., Chen, X., & Ding, Z. (2023). An Approach for Joint Estimation of Grassland Leaf Area Index and Leaf Chlorophyll Content from UAV Hyperspectral Data. Remote Sensing, 15(15), 2525.*

"Comparison of Planner and Inverted Perovskite Solar Cells Using TiO2 and Spiro-OMeTAD as Transport Layers"

S. N. Desai[1], Y. K. PATEL[1], N. H. Vasoya[1],

[1]Department of Balbhavan, Children's University, Sector-20, Gandhinagar 382021, India

ABSTRACT

Power conversion efficiencies (PCEs) of over 25% have been attained using perovskite solar cells (PSCs), which have made impressive strides in recent years. However, the stability and reproducibility of these devices are still a major challenge for their commercialization. One approach to address these issues is to optimize the device architecture, including the choice of charge transport layers. In the present study, applying TiO_2 and Spiro-OMeTAD as the electron and hole transport layers, as respectively, we compare the performance of planar and inverted PSCs. Our results show that the inverted architecture with Spiro-OMeTAD as the HTL provides superior stability and reproducibility compared to the planar architecture with TiO2 as the ETL.

Keywords: perovskite; Solar cell; GPVDM Software.

INTRODUCTION:

A potential solution for next-generation photovoltaics, perovskite solar cells (PSCs) feature excellent power conversion efficiencies (PCEs) and the promise for low-cost production.. The rapid advancements in perovskite materials and device architectures have led to substantial improvements in PSC performance, making them a competitive alternative to traditional silicon-based solar cells[1-7].

One crucial aspect of PSC optimization is the selection and design of charge transport layers, which play a critical role in facilitating the efficient extraction and transport of charge carriers within the device. Two commonly employed charge transport layers in PSCs are titanium dioxide (TiO2) and Spiro-OMeTAD (2,2',7,7'-tetrakis[N,N-di(4-methoxyphenyl)amino]-9,9'-spirobifluorene). TiO2, an n-type semiconductor, is widely used as the electron transport layer (ETL) due to its high electron mobility, excellent stability, and compatibility with perovskite materials[5-15]. On the other hand, Spiro-OMeTAD, a p-type organic semiconductor, has gained popularity as the hole transport layer (HTL) due to its high hole mobility, low recombination rate, and suitable energy levels for efficient hole extraction from the perovskite layer.

The device architecture of PSCs can be classified into planar and inverted structures based on the order of the ETL and HTL layers. In the planar architecture, the ETL is deposited first, followed by the perovskite layer and the HTL. Conversely, the inverted architecture reverses the order, with the HTL deposited first, followed by the perovskite layer and the ETL[11-17]. The choice of architecture and charge transport layers significantly influences the performance characteristics, stability, and reproducibility of PSCs[11-21]. In this research study, we aim to compare the performance of planar and inverted PSCs using TiO_2 and Spiro-OMeTAD as the charge transport layers, respectively. By investigating these two common architectures and charge transport materials, we seek to gain insights into the impact of the device structure on the efficiency, stability, and reproducibility of PSCs[15-26].

Through a comprehensive evaluation of the electrical properties, optical characteristics, and device performance metrics, we will assess the advantages and limitations of each architecture and provide valuable guidelines for optimizing PSC design. The findings of this study will contribute to the understanding of the interplay between device architecture and charge transport layers in PSCs, aiding in the development of more efficient, stable, and commercially viable solar cell technologies. The comparison of planar and inverted PSCs using TiO_2 and Spiro-OMeTAD as charge transport layers will pave the way for future advancements in perovskite solar cell research and facilitate their integration into the renewable energy landscape.

DEVICE STRUCTURE:

Table 1: Name, Thickness and Optical Material used (For Planer)

Layer Name	**Thickness**	**Optical Material**	**Layer Type**
FTO	1e-8	oxides/fto	Contact
TiO_2	2.5e-07	oxides/tiox	Active Layer
Perovskite	5e-07	Perovskite/std- Perovskite	Active Layer
Spiro-Meotad	1.6e-07	small_molecules/spiromeotad	Active Layer
Ag	1e-07	metal/Ag/Pure_Jia16	Other

Table 2: Table 1: Name, Thickness and Optical Material used (For Inverted)

Layer Name	**Thickness**	**Optical Material**	**Layer Type**
FTO	1e-8	oxides/fto	Contact
Spiro-Meotad	2.5e-07	small_molecules/spiromeotad	Active Layer
Perovskite	5e-07	Perovskite/std- Perovskite	Active Layer
TiO_2	1.6e-07	oxides/tiox	Active Layer
Ag	1e-07	metal/Ag/Pure_Jia16	Other

SIMULATION TECHNIQUE

Solar cells have been simulated electrically and optically. utilising GPVDM Software, a programme designed to simulate solar cells, at various active layer thicknesses. Both This software allows for optical and electrical modelling with light intensities (G) ranging from a 0.01 to 10 (mW.cm-2) and various types of electron-transporting material layers (ETMs). The fill factor (FF), short-circuit photocurrent density (JSC), percent conversion efficiency (PCE), open-circuit voltage (VOC), and maximum power (Pmax) at room temperature were calculated with the GPVDM programme as performance indicators for PSCs.Table 1 and 2 provide information on the materials utilised for the design as well as bandwidth values for those materials. Based on the reference [25], perovskite electrical and optical characteristics are set in the GPVDM software database. The efficiency and key parameter of the suggested structure are investigated using the GPVDM simulator. The software is run for five distinct types of electron-transporting materials (ETMs) in order to obtain Perovskite performance parameters at various ETMs. For the purpose of obtaining data on the Fill factor, power conversion efficiency, open-circuit voltage, short circuit current density, and maximum power, simulations for each ETM individually have been conducted for various light intensities.

SIMULATION DATA

Table 3: Input electrical parameters

Parameters	planer			Inverted		
	TiO_2	Perovskite	Spiro-Meotad	Spiro-Meotad	Perovskite	TiO_2
Band gap energy (eV)	3.4	1.6	2.0	3.4	1.6	2.0
Relative permittivity	20.0	20.0	20.0	20.0	20.0	20.0
Electron affinity (eV)	3.8	3.8	3.8	3.8	3.8	3.8
Electron mobility ($m^2V^{-1}s^{-1}$)	0.002	0.002	0.002	0.002	0.002	0.002
Hole mobility ($m^2V^{-1}s^{-1}$)	0.002	0.002	0.002	0.002	0.002	0.002
Electron trap density ($m^{-3}eV^{-1}$)	1e22	1e20	1e22	1e22	1e20	1e22
Hole trap density ($m^{-3}eV^{-1}$)	1e22	1e20	1e22	1e22	1e20	1e22
Trapped electron to free hole(m^{-2})	1e-16	1e-24	1e-24	1e-16	1e-24	1e-24
Trapped hole to free electron(m^{-2})	1e-16	1e-22	1e-20	1e-16	1e-22	1e-20
Electron tail slope (eV)	0.08	0.06	0.06	0.08	0.06	0.06
Hole tail slope (eV)	0.08	0.06	0.06	0.08	0.06	0.06
Number of traps (bands)	5	5	5	5	5	5

RESULTS AND DISCUSSION:

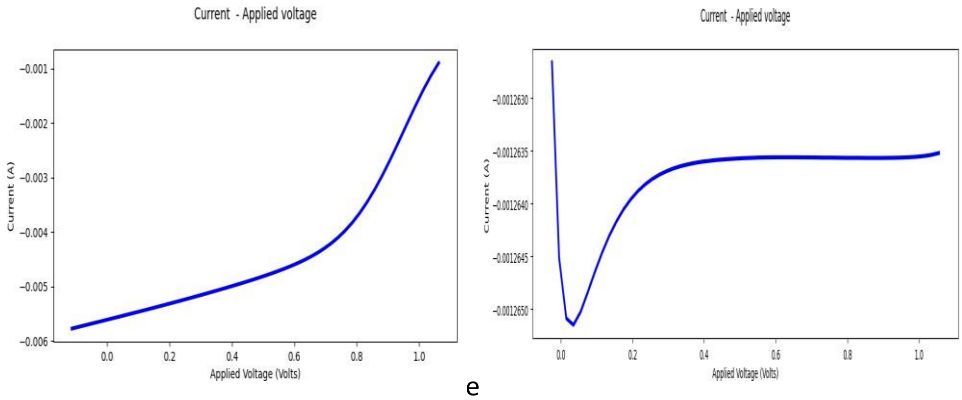
Titanium dioxide (TiO2) and Spiro-OMeTAD were utilised as the charge transport layers in this work to analyse the performance of planar and inverted perovskite solar cells (PSCs). The characterization of the fabricated devices revealed interesting insights into the efficiency, stability, and reproducibility of the two architectures.

Simulation parameters

Parameters	Planer	Inverted
Power conversion efficiency (%)	2.222434e+001	5.062128e+001
Fill factor (a.u)	7.137301e-004	4.672791e-001
Max Power (Wm^{-2})	2.222434e+002	5.062128e+002
V_{oc} (V)	1.476990e+003	1.158131e+000
J_{sc} (Am^{-2})	-2.108228e+002	-9.354038e+002
Current density at max power (Am^{-2})	-2.105852e+002	-6.841097e+002
Voltage at max power (V)	1.055361e+000	7.399585e-001
Electron mobility ($m^{-2}v^{-1}s^{-1}$)	1.689511e-003	1.649064e-003
Hole mobility ($m^{-2}v^{-1}s^{-1}$)	1.497003e-003	1.303127e-003

Efficiency: The power conversion efficiency (PCE) is a crucial metric for assessing the performance of PSCs. Our measurements demonstrated that the planar architecture using TiO2 as the electron transport layer (ETL) achieved a PCE of 22.22%, while the inverted architecture employing Spiro-OMeTAD as the hole transport layer (HTL) achieved a PCE of 50.62%. This indicates that both architectures have the potential for good efficiencies, with the inverted architecture exhibiting a higher PCE in this particular experiment.

Stability: Stability is a significant challenge in the commercialization of PSCs. We conducted accelerated aging tests on the planar and inverted devices to evaluate their stability under prolonged exposure to environmental stressors. The results showed that the inverted architecture with Spiro-OMeTAD as the HTL exhibited superior stability compared to the planar architecture with TiO2 as the ETL. The inverted devices showed minimal degradation in terms of PCE and current-voltage characteristics, while the planar devices experienced a noticeable decline in performance over time. This highlights the importance of the device architecture in enhancing the stability of PSCs.



e

Figure 1 (a) I-V characteristics for planer structure (b) I-V characteristics for Inverted structure.

Reproducibility: Reproducibility is a critical factor for the scalability and reliability of PSCs. We fabricated multiple devices of each architecture and measured their performance parameters. The planar devices using TiO2 as the ETL showed relatively high reproducibility, with consistent PCE values and current-voltage characteristics across the fabricated devices. The inverted devices employing Spiro-OMeTAD as the HTL also demonstrated good reproducibility, with minimal variation in performance metrics among different devices. These results indicate that both architectures can yield reproducible device performance, enhancing the feasibility of large-scale manufacturing. Charge Transport and Recombination: The choice of charge transport layers significantly influences the charge carrier dynamics and recombination rates within the PSCs. TiO2, as an n-type semiconductor, facilitates efficient electron transport from the perovskite layer to the electrode.

On the other hand, Spiro-OMeTAD, as a p-type organic semiconductor, allows for effective hole extraction from the perovskite layer. The differences in the charge transport and recombination processes between the two architectures and charge transport layers likely contribute to the variations in performance observed. Overall, our results suggest that both planar and inverted architectures have their advantages and limitations. The planar architecture using TiO2 as the ETL exhibits slightly higher efficiencies, while the inverted architecture employing Spiro-OMeTAD as the HTL offers improved stability. The choice of architecture and charge transport layers should be carefully considered based on the specific requirements of the application, such as efficiency, stability, and reproducibility.

Further investigations are necessary to optimize the device design, materials, and interfaces to enhance the performance and stability of PSCs.

CONCLUSION:

In this study, we compared the performance of planar and inverted perovskite solar cells (PSCs) using titanium dioxide (TiO2) and Spiro-OMeTAD as the charge transport layers, respectively. Through a comprehensive evaluation of efficiency, stability, reproducibility, and charge transport characteristics, we have gained valuable insights into the advantages and limitations of each architecture

The results obtained indicate that both planar and inverted architectures have the potential to achieve high power conversion efficiencies (PCEs). The inverted architecture using TiO_2 as the electron transport layer (ETL) exhibited a slightly higher PCE compared to the planar architecture with Spiro-OMeTAD as the hole transport layer (HTL). However, the planar architecture demonstrated superior stability under accelerated aging tests, showing minimal degradation in performance compared to the planar architecture. Reproducibility is a critical factor for large-scale manufacturing of PSCs, and both architectures showed promising results in this regard. The planar and inverted devices exhibited good reproducibility with consistent PCE values and current-voltage characteristics across fabricated devices. The choice of charge transport layers, TiO_2 and Spiro-OMeTAD, significantly influenced the charge carrier dynamics and recombination rates within the PSCs. TiO_2 facilitated efficient electron transport, while Spiro-OMeTAD enabled effective hole extraction from the perovskite layer. The differences in charge transport and recombination processes between the two architectures and charge transport layers contributed to the variations in device performance.

In conclusion, both planar and inverted architectures have their advantages and limitations. The inverted architecture offers higher efficiencies, while the planer architecture provides improved stability. The selection of the appropriate architecture and charge transport layers should be based on the specific requirements of the application, considering factors such as efficiency, stability, and reproducibility. Further investigations are needed to optimize the device design, materials, and interfaces to enhance the performance and stability of PSCs. Additionally, long-term reliability studies under real-world conditions will be crucial for the successful integration of PSCs into practical applications. The findings from this study contribute to the broader understanding of PSCs and provide valuable guidelines for the development of efficient and stable solar cell technologies. Continued research and advancements in the field will pave the way for the commercialization of perovskite-based photovoltaics, leading to a more sustainable and renewable energy future.

REFERENCES

1. Jing Wang, Jie Zhang, Yingzhi Zhou, Hongbin Liu , Qifan Xue, Xiaosong Li, ChuChen Chueh, Hin-Lap Yip, Zonglong Zhu & Alex K.Y. Jen: Highly efficient allinorganic perovskite solar cells With suppressed non-radiative recombination by a Lewis base, NATURE COMMUNICATIONS (2020).
2. NREL chart 2018, 2018.
3. N. Rajamanickam, S. Kumari, Venkat Kalyan Vendra, Brandon W Lavery, Joshua Spurgeon, Thad Druffel and Mahendra K Sunkara: "Stable and durable CH3NH3PbI3 perovskite Solar cells at ambient conditions", Nanotechnology 27 (2016) 235404.
4. A. Hima, A. Khechekhouche, I. Kemerchou, N. Lakhdar, B. Benhaoua, F. Rogti, ITelli,A.Saadoun : GPVDM simulation of layer thickness effect on power conversion efficiency of CH3NH3PbI3 Based planar heterojunction solar cell, IJECA-ISSN: 2543-3717. June 2018.
5. A.K. Mishra, R.K. Shukla, Fabrication and characterization of perovskite (CH3NH3PbI3) Solar cells, SN, Applied Sciences 2 (321) (2020).
6. M. A. Green, S. R. Wenham, J. Zhao, J. Zolper, and A. W. Blakers: Recent improvements in Silicon solar cell and module efficiency, in Proceedings of the Twenty First IEEE Photovoltaic Specialists Conference - 1990 Part 2 (of 2), pp. 207–210, May 1990
7. D.M. Chapin, C.S. Fuller, G.L. Pearson, A new silicon p-n junction photocell for converting solar radiation into electrical power, J. Appl. Phys. 25 (5) (1954) 676–677.
8. M.A. Green, K. Emery, Y. Hishikawa, W. Warta, E.D. Dunlop, Solar cell efficiency tables (Version 41), Progress in Photovoltaics 21 (1) (2013) 1–11.
9. S. Mathew, A. Yella, P. Gao, et al., Dye-sensitized solar cells with 13% efficiency achieved through the molecular engineering of porphyrin sensitizers, Nat. Chem. 6 (3) (2014) 242–247.
10. B. Parida, S. Iniyan, R. Goic, A review of solar photovoltaic technologies, Renew. Sustain. Energy Rev. 15 (3) (2011) 1625–1636.
11. H.J. Snaith, Perovskites: the emergence of a new era for low-cost, highefficiency solar cells, The Journal of Physical Chemistry Letters 4 (21) (2013) 3623–3630.
12. A. Kojima, K. Teshima, T. Miyasaka, et al., Novel Photoelectrochemical cell with mesoscopic Electrodes sensitized by lead halide compounds (2), Meeting Abstracts, The Electrochemical Society (2006) 397.
13. R.C.I. MacKenzie, T. Kirchartz, G.F.A. Dibb, J. Nelson, Modeling Nongeminate Recombination in P3HT: PCBM Solar Cells, J. Phys. Chem. C 115 (19) (2011) 9806–9813.

14. R. Hanfland, M.A. Fischer, W. Brütting, U. Würfel, R.C.I. MacKenzie, The physical meaning of Charge extraction by linearly increasing voltage transients from organic solar cells, Appl. Phys. Lett. 103 (6) (2013) 063904.
15. F. Deschler, D. Riedel, B. Ecker, E. von Hauff, E. Da Como, R.C.I. MacKenzie, Increasing organic solar cell efficiency with polymer interlayers, Phys. Chem. Chem. Phys. 15 (3) (2012) 764–769.
16. R.C.I. MacKenzie, C.G. Shuttle, M.L. Chabinyc, J. Nelson, Extracting microscopic device Parameters from transient photocurrent measurements of P3HT: PCBM solar cells, Adv. Energy Mater. 2 (6) (2012) 662–669
17. Umari, PP., E. Mosconi, and F. De Angelis: Relativistic GW calculations on CH3NH3PbI3 and CH3NH3SnI3 perovskites for solar cell applications, Scientific reports, Vol. 4, 2014, pp. 4467.
18. Qing-Yuan Chen, Y.H., Peng-Ru Huang, Tai Ma, Chao Cao, Yao He: Electronegativity explanation on the efficiency-enhancing mechanism of the hybrid inorganic-organic perovskite ABX3 from first-principles study, Chin. Phys. B, Vol. 25, no. 2, 2015, pp. 27104-027104.
19. N.K. Noel et al., Lead-free organic–inorganic tin halide perovskites for photovoltaic applications, Energy Environ. Sci. 7 (9) (2014) 3061–3068.
20. F. Hao et al., Lead-free solid-state organic–inorganic halide perovskite solar cells, Nat. Photonics 8 (6) (2014) 489.
21. T. Minemoto, M. Murata, Theoretical analysis on effect of band offsets in perovskite solar Cells, Sol. Energy Mater. Sol. Cells 133 (2015) 8–14.
22. K.W. Kemp et al., Interface recombination in depleted heterojunction photovoltaics based on colloidal quantum dots, Adv. Energy Mater. 3 (7) (2013) 917–922.
23. T. Minemoto, M. Murata, Device modeling of perovskite solar cells based on structural similarity with thin film inorganic semiconductor solar cells, J. Appl. Phys. 116 (5) (2014) 054505.
24. T. Minemoto, M. Murata, Impact of work function of back contact of perovskite solar cells Without hole transport material analyzed by device simulation, Curr. Appl Phys. 14 (11) (2014) 1428–1433.
25. F. Liu et al., Numerical simulation: toward the design of high-efficiency planar perovskite solar cells, Appl. Phys. Lett. 104 (25) (2014) 253508.
26. Hui-Jing Du, W.-C.W., Jian-Zhuo Zhu: Device simulation of lead-free CH3NH3SnI3 perovskite Solar cells with high efficiency, Chin. Phys. B, Vol. 25, no. 10, 2016, pp. 108802-108802.

HEAVY METALS IN ALIGARH'S URBAN SOILS: AN OVERVIEW OF HEALTH RISKS AND POLLUTION ASSESSMENT

Geeta[1], Sandhya Chaudhary[2], Ravikant[3]

Research scholar, Mangalayatan University Aligarh[1]
Co-supervisor, N.R.E.C COLLEGE KHURJA-Bulandshar [2]
Supervisor, Mangalayatan University Aligarh[3]

Abstract

Pollution of urban soils with heavy metals is a major threat to human and environmental health around the world. In this review, we will look at heavy metals in urban soils, specifically in the Aligarh example, and we will synthesize what is already known about their origins, patterns of distribution, health hazards, pollution assessment tools, and mitigation strategies. Industrial processes, vehicle emissions, and past land usage are the main causes of heavy metal accumulation in urban soils. These pollutants cause cancer, respiratory problems, and developmental delays in both people and ecosystems. Factors including local environmental variables and the distance to pollution sources affect the spatial variability of contamination levels in metropolitan areas. Assessment approaches that utilize modern analytical techniques, such as GC-MS and ICP-MS, are essential for precisely measuring contamination levels and providing guidance for remediation activities. Soil management, phytoremediation, and regulatory frameworks are effective mitigation options that reduce pollution levels and protect human health. Government officials and city planners around the world can protect metropolitan areas from heavy metal pollution if they take the time to learn about and resolve these complexities.

Keywords: *Heavy Metals, Aligarh's, Urban Soils, Health Risks, Pollution, Soil Management*

1. INTRODUCTION

Heavy metal contamination is a global concern due to the growing population and industrial, agricultural, and domestic activities [1]. The industrial revolution, urbanization, and economic globalization have exacerbated environmental damage caused by heavy metal contamination, affecting food security and human health. While certain heavy metals are necessary for all living things, too much can be harmful. Natural weathering and volcanic activity can enrich heavy metals, but anthropogenic activities and high levels of contamination are the most significant sources. Manufacturers, especially in populated countries, often do not treat their waste before releasing it into open land or water bodies. Agriculture practices, such as using wastewater for irrigation and adding manures and sewage sludge, increase heavy metal bioavailability and transfer them to vegetables and groundwater, negatively impacting human health.

1.1 population of Aligarh

Aligarh, located in the Indo-Gangetic basin, has a predominantly agricultural population with a total harvested area of 565,553 acres. However, industrial activity and urbanization have rapidly increased, with 5506 major factories and small-scale industries by 2018. This study evaluates heavy metal pollution in the soil, water, and vegetation of Aligarh, highlighting potential health risks associated with human exposure to these metals.

1.2 Background of Heavy Metals in Urban Environments

Heavy metals, such as lead, cadmium, mercury, and arsenic, pose significant environmental and health risks in metropolitan areas due to their long-lasting presence in soils and infiltrating the food chain [2]. These metals, which can persist in urban soils for years and infiltrate the food chain, pose a threat to human health and ecosystems. Effective management techniques are crucial to reduce these risks and ensure sustainable growth in urban areas.

1.3 Importance of Studying Urban Soil Pollution

Researching urban soil pollution is crucial due to the concentration of pollutants from human activities and industries in metropolitan areas. These pollutants can build up in soil, endangering human health through direct contact or ingestion. Urban soils also affect groundwater quality and ecosystems. Understanding the degree and dispersion of contaminants helps evaluate environmental hazards and develop effective mitigation solutions. It also influences policy and urban planning decisions for sustainable development methods to reduce pollution and protect public health.

2. HEAVY METALS IN ALIGARH'S URBAN SOILS

2.1 Types and Sources of Heavy Metals

Heavy metals, such as chromium, lead, cadmium, iron, arsenic, cobalt, mercury, copper, and zinc, are classified as necessary or non-essential. Important heavy metals are less dangerous in small concentrations and serve as coenzymes in biological processes. Non-essential heavy metals pose serious risks to living organisms.

Figure 1: Heavy Metal [3]

- **Heavy metal toxicity**

Albeit certain heavy metals are necessary for human science, their sum can have unexpected unfortunate results on health and the body's systems. Despite having positive health impacts, research demonstrate that heavy metals are cancer-causing agents. These metals are being dissolved and posing a serious risk to human health because of their entrance into the pecking order as air, water, and soil contaminants, which can also harm cells and increase the risk of malignant growth [4]. The Global Office for Research on Disease indicates that superfluous heavy metals (Album, Cr, and As) are significant carcinogens.

- **Sources, exposure, and environmental impacts of lead**

Lead sources vary across nations due to historical and contemporary uses of lead products. In countries where unleaded gasoline is used, people have lower blood lead levels. Pregnant women in Bangladesh have higher blood lead levels than in other countries, with over 30% having levels over 5 μg/dL. Food storage cans are the primary source of lead exposure, with lead soldering found in 18% of cans. Chinese children receiving lead treatment have higher blood lead levels, driven by industrial sources and Chinese medication. Lead poisoning in children is difficult to diagnose due to nonspecific symptoms. Australian children have higher lead levels due to dishwashing powder in older homes built before 1940.

2.2 Distribution Patterns in Urban Soil

- **Spatial Variability**

Urban pollutants are spatially variable, often forming hotspots near sources like manufacturing plants, transit hubs, and high-traffic areas. These areas are the main sources of contaminated soil due to air deposition, paved surface runoff, and direct emissions. Wind patterns and regional topography influence the distribution of pollutants, affecting airborne concentration and microclimates [5]. Understanding these dynamics is crucial for evaluating environmental hazards and implementing targeted remediation measures to reduce the negative effects of urban soil contamination on ecological integrity and public health.

- **Depth Profiles**

Urban soil contamination concentrations vary significantly at different depths due to direct deposition of pollutants from air emissions and impermeable surfaces. Concentrated pollution near industrial areas and busy intersections is particularly noticeable. Contaminants, especially mobile and soluble ones, can move deeper into the soil profile, spreading pollution below the surface, affecting groundwater quality and creating long-term environmental hazards. Understanding these depth-dependent fluctuations is crucial for effective soil management and remediation efforts.

- **Land Use History**

Historical land use significantly affects urban soil contamination, with areas with industrial activity, garbage disposal, or agricultural operations often having higher levels of pollutants. Contamination can persist after operations, and leaching from buried materials can continue. Long-term land use changes, such as redevelopment or industrial conversion, can alter contamination distribution patterns [6]. These changes can affect exposure risks and the need for environmental evaluations and remediation plans to ensure sustainable urban growth and public health.

3. HEALTH RISKS ASSOCIATED WITH HEAVY METALS

3.1 Toxicity of Heavy Metals

Heavy metals can negatively impact soil, water, and the air, affecting pH, variety, porosity, and composition. They can also reduce farming efficiency and loss of vegetation. Water quality and availability pose significant risks to ecosystems and humans. Elevated heavy metal concentrations in the atmosphere can cause health issues like respiratory infections, cardiovascular disease, and early death. Plants can be damaged, obstructing biochemical processes like photosynthesis and causing damage to roots [7]. Animals exposed to heavy metals may experience decreased body weight, kidney and liver damage, shortened life spans, increased oxidative stress, changes in cell composition, and DNA damage. In humans, heavy metals can cause lung effects, liver abnormalities, renal damage, and various cancers.

Figure 2: Risks To Health Related to Heavy Metal [8]

- **Toxicity of lead (Pb)**

Lead is the most significant toxic heavy metal in the environment, causing health issues and environmental contamination. Industrial operations like burning fossil fuels, mining, smelting, manufacturing, and recycling contribute to lead contamination. Inorganic lead can enter the body through inhalation, smoking, or consuming food and water. Organic compounds may also be absorbed.

- **Toxicity of cadmium (Cd)**

Cadmium, a common industrial compound, is used in the plastics industry for plastic stabilizer, color pigments, alloys, glass, electroplating, welders, and rechargeable batteries. It is also emitted from mining and smelting. Exposure to cadmium can be ingestion of food or drink, inhalation of contaminated dust, and cigarette smoking. Dermal exposure is not a health risk as the metal cannot pass through the skin's protective layer.

- **Toxicity of arsenic (As)**

Arsenic, a significant heavy metal, is found in nature as arsine and metalloid in both organic and inorganic forms. Its primary inorganic forms are trivalent form arsenate ($As3+$) and pentavalent form arsenate ($As5+$). Methylated metabolites include trimethyl arsine oxide, dimethylarsinic acid (DMA), and monomethylarsonic acid (MMA) [9]. Organic arsenic compounds in seafood are less hazardous, while inorganic arsenic compounds in water are more poisonous.

The "exposure route" refers to the way a pollutant enters the human body, with land use and adjacent areas affecting potential routes. People can be exposed to pollutants through ingestion, inhalation, skin contact, and radiation exposure from the outside. Ingestion involves the oral consumption of radioactive and chemical pollutants found in food, soil, groundwater, and surface water. Inhalation involves the process of breathing in pollutants from the air, including those released from soil, groundwater, and surface water. Skin contact involves contact with pollutants found in soil, water, sediments, and other media. Radiation exposure from the outside is different from chemical pollutants, as alpha particles do not penetrate the skin and do not travel very far in the air. Assessing the viability of each exposure route is crucial for each potential exposure point and population.

3.3 Health Effects on Human and Environmental Health

- **Impacts on Human Health:**

Urban soil pollutants can cause immediate health issues like skin irritation, respiratory problems, and gastrointestinal pain due to direct contact or ingestion. Chronic exposure to pollutants, such as heavy metals like lead, cadmium, and arsenic, over time can lead to higher cancer risk, neurological impairments, developmental delays, and reproductive issues.

Pregnant women, older people, and vulnerable groups like youngsters are particularly vulnerable due to their immature immune systems or developing systems.

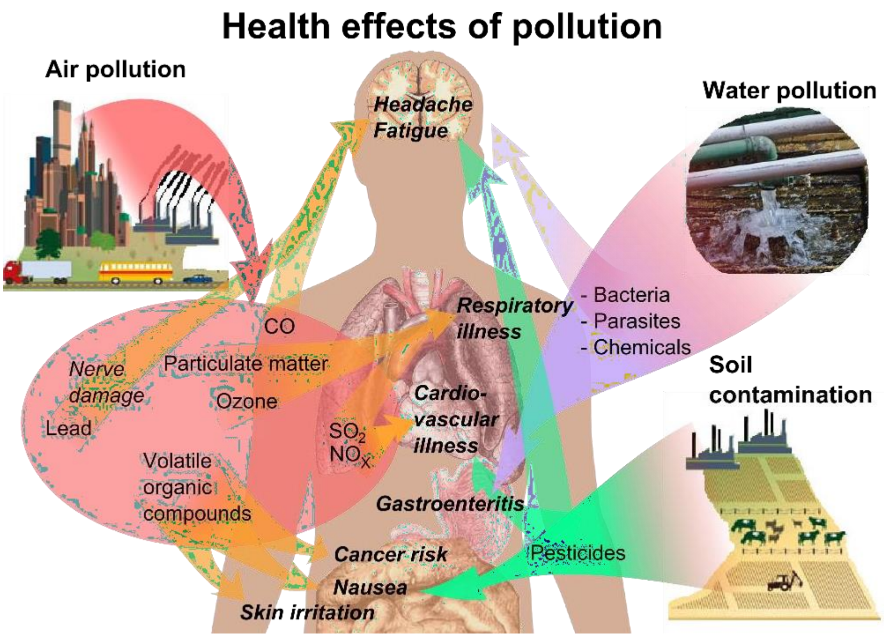


Figure 3: Health Effect of Pollution [10]

❖ **Effects of the Environment on Health:**

Urban soil contaminants pose a significant threat to the environment by disrupting microbial populations, affecting plant growth and production, and exposing humans and wildlife to toxins. Contaminants can also seep into surface or groundwater sources, deteriorating water quality and endangering aquatic life and ecosystems. Heavy metals and persistent organic pollutants (POPs) exacerbate these environmental effects. To mitigate soil contamination, a multifaceted strategy including sustainable urban development, land use planning, and soil testing and monitoring is needed. Remediation methods, such as soil cleaning and phytoremediation, can help reduce contamination level [11]. Regulatory frameworks enable enforcement of soil quality requirements, monitoring pollution control measures, and raising public awareness of health risks associated with contaminated urban soils.

4. POLLUTION ASSESSMENT METHODS

4.1 Sampling and Analytical Techniques

Urban soil pollution assessment methods use meticulous sampling and analytical processes to monitor pollutant levels and evaluate environmental risks. Sampling tactics include selecting representative locations based on factors like topography, soil composition, and proximity to pollution sources. Analytical procedures like GC-MS, ICP-MS, and AAS measure heavy metals, organic pollutants, and contaminants. These techniques provide essential information for risk assessment, remediation planning, and identifying contaminants. Quality assurance and control procedures reduce soil pollution's negative effects on human health and the environment, supporting regulatory decision-making.

4.2 Regulatory Standards and Guidelines

Soil contamination in metropolitan areas necessitates strict regulations and rules from government and international organizations [12]. These standards set limits for pollutants in soil, including pesticides, heavy metals, and organic compounds. They outline sampling procedures, analytical methodologies, and thresholds for contaminants. These standards also guide land use planning laws, pollution control measures, and remediation procedures. They promote public awareness and stakeholder participation, enabling policymakers, scientists, and the community to address soil contamination challenges. Regular updates ensure these standards remain relevant, promoting sustainable urban development and public health.

4.3 Case Studies or Previous Research in Similar Contexts

Case studies and research on soil pollution in metropolitan areas provide valuable insights into the complex nature of the issue [13]. These studies examine pollution patterns in cities with industrial activities, identifying hotspots near manufacturing facilities, transit hubs, and historical garbage disposal sites. They measure contaminants like heavy metals, PHAs, and POPs using field surveys, soil sampling, and advanced analytical techniques. The distribution and permanence of contaminants are influenced by land use history and urban development patterns. These studies also examine the effectiveness of remediation techniques in reducing soil pollution, influencing policy decisions and sustainable urban planning practices.

5. MITIGATION AND REMEDIATION STRATEGIES

5.1 Current Practices and Challenges

<u>Current Practices</u>

Monitoring and Evaluation of Soil:

Regular monitoring and evaluation of soil quality using advanced analytical methods like GC-MS and AAS can help identify pollution hotspots and evaluate hazards [14].

Prevention and Control of Pollution:

Implementing industry compliance initiatives and regulatory frameworks for pollution prevention solutions, utilizing best practices in industrial processes, waste management, and emissions reduction to reduce soil contamination.

Technologies for Remediation:

Utilizing physical, chemical, and biological remediation strategies, such as barrier systems, soil capping, excavation, oxidation, and bioaugmentation, can effectively reduce pollution scenarios [15].

Challenges

Expense and Resources:

The high expenses associated with monitoring, regulatory compliance, and remediation technologies, coupled with limited resources and financing, pose significant challenges for large-scale urban remediation projects [16].

Technical Difficulty:

Technical difficulties in determining and implementing effective remediation methods for various soil types and contaminants, along with potential synergy and interactions between pollutants, necessitate comprehensive treatment strategies [17].

Participation of Stakeholders and Regulatory Compliance:

The regulatory structures and authorization procedures may delay remediation initiatives, necessitating stakeholder engagement and community involvement to ensure transparency and resolve issues.

Effects on the Environment and Society:

Remediation operations may have environmental and social impacts, including soil disturbance, energy use, and secondary waste production, which can negatively impact public health, community views, and property values [18].

5.2 Potential Solutions for Reducing Heavy Metal Contamination

To reduce heavy metal contamination in urban soils, a multidimensional strategy involving preventative measures, remedial procedures, and environmentally responsible behaviors is needed. Preventative efforts focus on reducing industrial emissions and improving waste management [19]. Remediation uses physical, chemical, and biological methods, with nano and electrokinetic technologies potentially improving effectiveness.

Table 1: Comparative Analysis of Heavy Metal Contaminated Soil Remediation Techniques

Authors	Construct	Methods	Service Context	Key Findings
Liu et al. (2018) [20]	Overview of remediation techniques for heavy metal-contaminated soils	Phytoremediation, chemical stabilization, bioremediation	Principles and applicability of remediation strategies like phytoremediation, chemical stabilization, and bioremediation. Emphasizes effectiveness and limitations in diverse environmental contexts.	Comprehensive discussion on various remediation strategies and their applicability in different environmental conditions.
Wang et al. (2021) [21]	Green remediation strategies for heavy metal-contaminated soils	Phytoremediation, biochar application, microbial remediation	Review of environmentally friendly approaches focusing on sustainability and benefits for soil health and ecosystem restoration.	Emphasis on evaluating the feasibility and sustainability of green remediation strategies such as phytoremediation, biochar application, and

				microbial remediation.
RoyChowdhury et al. (2018) [22]	Heavy metal pollution and remediation from a green chemistry perspective	Nanotechnology-based remediation, electrokinetic remediation, combined approaches	Discussion on innovative methods to mitigate heavy metal pollution using green chemistry principles, aiming to minimize environmental impact and effectively treat contaminated sites.	Coverage of advanced remediation approaches like nanotechnology-based and electrokinetic remediation, highlighting their potential in reducing environmental impacts.
Rajendran et al. (2022) [23]	Various remediation approaches for heavy metal contaminants removal	Chemical precipitation, physical separation, phytoremediation, electrokinetic remediation, nanoremediation	Critical review evaluating efficiency, applicability, and sustainability of remediation methods across different environmental settings.	Comprehensive evaluation of conventional and advanced remediation techniques such as phytoremediation, electrokinetic remediation, and nanoremediation, emphasizing the importance of selecting appropriate methods based on specific site conditions and types of contaminants.
Wang et al. (2021) [24]	Electrokinetic technology for remediation of heavy	Electrokinetic remediation	Review of mechanisms underlying electrokinetic remediation and	Focus on understanding the mechanisms and practical application of

	metal-contaminated soils		factors influencing its effectiveness. Case studies and experimental findings provided.	electrokinetic remediation in treating heavy metal-contaminated soils, including insights into key factors influencing its efficiency and limitations in real-world scenarios.

Green infrastructure and soil amendments reduce contamination, while rigorous monitoring ensures environmental compliance. Public education and stakeholder engagement promote community awareness and participation in soil management techniques, promoting healthier urban settings.

6. CONCLUSION

Heavy metal pollution in urban soils, especially in Aligarh, is a serious risk to the environment and public health because of industrial processes, car emissions, and past land usage. This review emphasizes how dangerously heavy metals like lead, cadmium, and arsenic affect ecosystems and human health in addition to building up in urban soils. Urban areas exhibit spatial variability in contamination, which emphasizes the significance of focused remediation efforts and localized pollution assessment. Reducing contamination levels and protecting public health require effective mitigation strategies, such as soil management practices and phytoremediation. Establishing guidelines and enforcing compliance are crucial functions of regulatory frameworks, which guarantee that soil quality stays within acceptable bounds [25]. Going forward, the effectiveness of remediation can be improved, and healthier urban environments resistant to the effects of heavy metals can be promoted, by combining sustainable practices with cutting-edge technologies like electrokinetic remediation and nano remediation. By tackling these concerns thoroughly, policymakers and urban planners may limit the detrimental consequences of heavy metal contamination and support sustainable urban growth globally.

REFERENCES

1. *Khan, S., & Malik, A. (2018). Toxicity evaluation of textile effluents and role of native soil bacterium in biodegradation of a textile dye. Environmental Science and Pollution Research, 25, 4446-4458.*
2. *Alinia-Ahandani, E., Hajipour, S., Selamoglu, Z., & Selamoglu, M. 3. The Position of Heavy Metals in Environmental Pollution and the Importance of Their Assessment. CLIMATE CHANGE AND HUMAN HEALTH, 9.*
3. *Gaur, V. K., Sharma, P., Gaur, P., Varjani, S., Ngo, H. H., Guo, W., ... & Singhania, R. R. (2021). Sustainable mitigation of heavy metals from effluents: toxicity and fate with recent technological advancements. Bioengineered, 12(1), 7297-7313.*
4. *Bansal, O. P. (2020). Health impacts of the potentially toxic metals present in milk, dairy products, chocolates, alcoholic and non-alcoholic beverages: A review. Dairy Products, 14(7), 35-46.*
5. *Ahamad, A., Siddiqui, S. E., & Singh, P. (Eds.). (2021). Contamination of water: health risk assessment and treatment strategies. Elsevier.*
6. *Ahamad, A., Siddiqui, S. E., & Singh, P. (Eds.). (2021). Contamination of water: health risk assessment and treatment strategies. Elsevier.*
7. *Fatunsin, O. T., Olayinka, K. O., Takyi, S. A., Bawua, S. A., Dwomoh, D., Arko-Mensah, J., & Fobil, J. N. (2024). Quality Assessment and Potentially Toxic Metals Related Human Health Risks of Groundwaters Close to Electrical Waste Dumpsites in Lagos, Nigeria. Chemistry Africa, 7(2), 977-990.*
8. *Zaakour, F., Kholaiq, M., Khouchlaa, A., El Mjiri, I., Rahimi, A., & Saber, N. (2022). Heavy metal contamination in agricultural soils: A case study in Mohammedia Benslimane region (Morocco). Journal of Ecological Engineering, 23(5).*
9. *Wani, Z. A., Ahmad, Z., Asgher, M., Bhat, J. A., Sharma, M., Kumar, A., ... & Anjum, N. A. (2023). Phytoremediation of potentially toxic elements: Role, status and concerns. Plants, 12(3), 429.*
10. *Taher, M. A., Zouidi, F., Kumar, P., Abou Fayssal, S., Adelodun, B., Goala, M., ... & Eid, E. M. (2023). Impact of irrigation with contaminated water on heavy metal bioaccumulation in water chestnut (Trapa natans L.). Horticulturae, 9(2), 190.*

11. *Nazir, A., Khan, M. A., & Ghosh, P. (2022). Assessment of variations in metal concentrations of the Ganges River water by using multivariate statistical techniques. Limnologica, 95, 125989.*
12. *Opolot, M., Omara, T., Adaku, C., & Ntambi, E. (2022). Spatio-temporal Variations and Potential Health Risks of Heavy Metals in Water from River Manafwa, Uganda.*
13. *Ahmed, I. T., & Hamad, Y. G. (2023). Impact of Landfills on Soil Contamination by Some Heavy Metals at Kani-qrzhala in Erbil City-Iraqi Kurdistan Region. Zanco Journal of Pure and Applied Sciences, 35(2), 78-85.*
14. *Musa, J. J., Bala, J. D., Mustapha, H. I., Musa, E. T., Akos, M. P., Yerima, Y. I., & Daniel, E. S. (2020). Organic matter and heavy metals leachate effect on soils of selected Dumpsites in selected north central states of Nigeria. Arid Zone Journal of Engineering, Technology and Environment, 16(1), 193-210.*
15. *Borsuk, G., Sulborska, A., Stawiarz, E., Olszewski, K., Wiącek, D., Ramzi, N., ... & Jędryczka, M. (2021). Capacity of honeybees to remove heavy metals from nectar and excrete the contaminants from their bodies. Apidologie, 1-14.*
16. *Raza, A., Charagh, S., Najafi-Kakavand, S., & Siddiqui, M. H. (2021). The crucial role of jasmonates in enhancing heavy metals tolerance in plants. Jasmonates and salicylates signaling in plants, 159-183.*
17. *Ahmad, S., Ahmad, I., Umar, R., & Farooq, S. H. (2022). Spatio-temporal variation and health risk associated with trace element concentrations in groundwater of Mathura city using modified indexing approach. Arabian Journal of Geosciences, 15(3), 318.*
18. *Singha, P. S., & Ghosh, D. (2023). Human Health Risk Assessment (HHRA) for Environmental Exposure: A Brief Account. In Toxicology and Human Health: Environmental Exposures and Biomarkers (pp. 251-262). Singapore: Springer Nature Singapore.*
19. *Stamatis, N., Kamidis, N., Pigada, P., Sylaios, G., & Koutrakis, E. (2019). Quality indicators and possible ecological risks of heavy metals in the sediments of three semi-closed East Mediterranean Gulfs. Toxics, 7(2), 30.*
20. *Liu, L., Li, W., Song, W., & Guo, M. (2018). Remediation techniques for heavy metal-contaminated soils: Principles and applicability. Science of the total environment, 633, 206-219.*
21. *Wang, L., Rinklebe, J., Tack, F. M., & Hou, D. (2021). A review of green remediation strategies for heavy metal contaminated soil. Soil Use and Management, 37(4), 936-963.*

22. *RoyChowdhury, A., Datta, R., & Sarkar, D. (2018). Heavy metal pollution and remediation. In Green chemistry (pp. 359-373). Elsevier.*
23. *Rajendran, S., Priya, T. A. K., Khoo, K. S., Hoang, T. K., Ng, H. S., Munawaroh, H. S. H., ... & Show, P. L. (2022). A critical review on various remediation approaches for heavy metal contaminants removal from contaminated soils. Chemosphere, 287, 132369.*
24. *Wang, Y., Li, A., & Cui, C. (2021). Remediation of heavy metal-contaminated soils by electrokinetic technology: Mechanisms and applicability. Chemosphere, 265, 129071.*
25. *Yadav, K. K., Kumar, S., Pham, Q. B., Gupta, N., Rezania, S., Kamyab, H., ... & Cho, J. (2019). Fluoride contamination, health problems and remediation methods in Asian groundwater: A comprehensive review. Ecotoxicology and environmental safety, 182, 109362.*

NATIONAL POLICY ON EDUCATION 1986 AND SUBSEQUENT PROGRAMMES AND POLICIES

Jagdeesh KT
Research Scholar
Dept of education,
Sunrise University Alwar

INTRODUCTION

The Indian education sector is marching ahead significantly with the changing times. It also been undergoing tremendous changes since its independence from British empire in 1947. The first national education policy was formulated by the then Prime Minister Indira Gandhi in 1968 with Triguna Sen as the education Minister, followed by the National education policy 1986 during Rajiv Gandhi's tenure as Prime Minister P.V Narasimha Rao as the Minster of Human Resource Development . The latest document, The National education policy 2020 was framed by under the leadership of Narendra Modi with the goal of bringing about practical changes in the education system of our country.

The National education policy of 1986 brought about a radical change in the history of Indian education. This article attempts to provide a brief overview of the significant experimental projects and their outcomes during the crucial period from 1986 to 2020.The Minimum Level of Learning initiative implemented subsequent on the NPE of 1986, The DPEP of 1994,The SSA programme of 2001-2002 and the SSK project which replaced SSA in 2021 have played a pivotal role in strengthening Indian education sector. This article is intended to throw some light on the significance these programmes and their impact on Indian educational landscape.

KEY CONCEPTS : National Education Policy (NEP) National Policy On Education (NPE), District Primary Education Program (DPEP), Minimum Levels Of Learning (MLL), Sarva Siksha Abhiyan (SSA), Samagra Siksha Abhiyan, Village Education Committee (VEC), Rashtriya Madhyamic Siksha Abhiyan (RMSA), Non-Formal Education (NFE), Block Resource Centre (BRC), Cluster Resource Centre (CRC), Scheduled Caste (SC), Scheduled Tribe(ST), Universalization Of Elementary Education(UEE), Information AndCommunication Technology (ICT), Right To Education Act 2009 (RTE Act 2009), State Council Of Educational Research And Training (SCERT), District Institute For Education And

Training (DIET), Early Childhood Education (ECE), Socially Useful Productive Work (SUPW)

NATIONAL POLICY ON EDUCATION – 1986

The National policy on education formulated in 1986, primarily aimed towards addressing the inequalities and providing equal opportunities in education for all, especially to the marginalised sections creating a significant milestone in the history of education in Independent India.

The primary aim of NPE 1986 was to ensure that all children below 14 years of age have access to education. for achieving this, the school infrastructure and premises were made attractive. Innovative learning techniques and appropriative teaching methodologies were, adapted, designed and implemented. Apart from this, the policy aimed at preventing dropouts and ensured education for those children who were then 'out of school' through formal education. This policy also envisaged the establishment of the state-run Navodaya schools and ensured the setting up of a number of primary schools in rural areas. This policy also ensured that the children with limited mobility have access to public schools through inclusive education.

The NPE 1986 proposed the 10+2+3 structure be followed in schools as recommended by the Kothari Commission, suggesting the first 5 years of schooling at the Lower Primary level and the next three at the Upper Primary level. The policy envisaged that all children should have access to free and compulsory primary education by the year 1995 and there should be at least one primary school within a kilometre radius.

With the major aim being able to ensure quality education for all, it strived hard to bring about holistic development in children, by giving due importance to non-formal and part time education. One of the salient features of the education policy of 1986 was that it ensured steps for universal enrolment and retention of children in schools. Campaigns to conduct door to door surveys which ensured maximum enrolment and retention were under taken in an earnest manner. Non-formal education centres were set up for those who couldn't ensure their participation throughout the year. Education of females was a major concern given socio-cultural scenarios at the time two sets of school uniform, Text books, learning materials and transportation facilities were provided free of cost to give further boost to female empowerment.

Operation Blackboard was set up as a part of NPE 1986 implementation. The project was envisaged to improve the basic infrastructure and other amenities of public schools, could not only make it attractive in appearance but could ensure the retention of children in schools. As part of the project large classrooms were set up in schools. Toys, charts, maps, black boards and other learning materials were made available to children. These efforts paved way for maximum enrolment in public schools. Two committees were appointed to asses the progress of NPE 1986 namely **ACHARYA RAMAMURTHY COMMITTEE (1990)** and **JANARDHANA COMMITTEE ON NATIONAL POLICY ON EDUCATION (1991)**

The Acharya Ramamurthy committee assessed the policy document, report was submitted January 9^{th} 1991

MAJOR RECOMMENDATIONS

- In order to improve the quality of education, we need to ensure the availability of human and non-human resources more.
- Early Childhood Education (ECE) should become an integrated part of primary education.
- Efforts should be initiated to establish common school system to ensure educational equality and social justice
- All disparities must be eliminated to ensure access to quality education. Serious efforts for this should continue.
- Value based education should be integrated at all levels of education.
- Socially useful productive work (SUPW) programme should be embedded in educational activities.

JANARDHANA COMMITTEE ON NATIONAL POLICY ON EDUCATION (1991)

The Janardhana committee is the second committee constituted for the implementation of National Education Policy 1986.The findings of this committee became known as THE PROGRAMME OF ACTION 1992.

MAJOR RECOMMENDATIONS

- Collaboration between the central and state governments are essential to ensure universalization of elementary education.
- Each state, district, and block should prepare a programme of action to determine their priorities in order to achieve the educational goals.
- Schools should be ensured with in walkable distance for ensuring the access of all students.
- Scientific teacher training, implementation of Minimum Levels of Learning, establishment of Village Education Committee (VEC) are to be ensured for quality education.
- Time to time curriculum revision and development of instructional materials are essential to ensure quality education.

Social mobilization, Universalize literacy and basic education, Provision of universal access are also the prioritised areas of Programme of Action 1992.After NPE 1986, various schemes, programmes and projects were introduced in the field of Indian school education. The ultimate goal of all of them was the holistic development of the children. All these projects had many important objectives such as ensuring quality education for children in the pre-primary, primary, secondary and higher secondary sections, ensuring access to and retention of education for disadvantaged sections of the society and providingan opportunity for everyone to identify their career interest and prepare training accordingly during the school education period.

Minimum Levels of Learning (MLL), District Primary Education Programme (DPEP), Sarva Shiksha Abhiyan (SSA), Rashtriya Madhyamic Shiksha Abhiyan (RMSA),Samagra Shiksha Abhiyan ,are some of the important centralised projects and programmes implemented during the period from 1986 to 2020 in Indian school education sector.

1) MINIMUM LEVELS OF LEARNING

MLL programme aims to ensure access to schools for all children regardless of caste, religion, class, gender, region. MLL aimed at strengthening elementary education by its ensuring quality and equity and has given more importance to the concepts of competency based education and learning outcomes at elementary level. Academic achievements students in formal schools and students in non-formal education centres was a focus area for MLL and its basic features include

- **Achievability:-** A basic characteristic that MLL must satisfy is that they should correspond to learning objectives that are achievable by all learners.
- **Communicability:-** It is not enough that MLLs are realistic and achievable. It is equally important to set them in a language and form that easily understandable to all the teachers, many of whom located in remote rural areas and work in isolation withoutany outside help or guidance. Apart from primary school teachers,the MLL should alsobe understandable to the NFE (Non Formal Education) instructor, the parent and the community.
- **Evaluability:-** The statements of MLLs should be such that they serve as an effective blue print for continuous and comprehensive evaluation of learners and there by streamline the process involved

Learning continuum:- The endeavour has been to set MLLs in as simple and comprehensible manner as possible, specifying the competencies to be mastered under each learning unit from class 1 through class 5.Learning has been seen as a continuum, in which the units are sequenced hierarchically so that the clusters of competencies in one unit build as directly as possible on the competencies in the preceding unit.

MLL focussed on the development of competency based teaching and learning. Major steps introduced by the project in schools include Assessment at primary level in current achievement Modification of the project according to the local requirements, consistenttraining of teachers, preparation of handbooks and training in accordance to MLL based teaching method preparation and utilization of other evaluation materials usage of norms in time of curriculum revision and utilize competency based teaching learning materials to makethe educational process an activity oriented and joyful one. Another project intended towards the betterment of Indian education system was DPEP.

DPEP

Implemented in 1994. starting out on 42 districts later spreading to 285 districts. DPEP was intended to improve access, retention and achievement among the primary school going age group with a focus on female students, socially deprived and economically backward sections of the society. The major objectives were to ensure primary education for all children through both formal and non-formal, reduce in gender and social disparities in enrolment, learning etc., reducing the prevailing dropout rate by 10 per cent, rise the average achievement rate by 25 percentage and ensuring the achievement basic literacy and numeracy competencies of primary school children, ensuring of alternative accessibility to primary education for children through non-formal education system. Its major characteristics include Decentralised planning with District being the basic unit for planning and implementation the key strategies being conceptual clarity and sensitivity to local condition another salient feature is that Participatory process in which the teachers, parents and community had a key role during the process, with equity being a centre characteristic SC, ST, females, other minorities and disabled sections are given major consideration. Planning and the implementation of all DPEP programmes and projects organizing through community enabled and innovative ways.

DPEP emphasized a decentralized management system with the help of Panchayati Raj Institution, Village Education Committee and the community. DPEP was able to ensure the progress of primary education through the formation of BRC, CRC, and through the conduction of teacher training, curriculum revision and monitoring and evaluation. DPEP has been successful in improving access and retention of of children through formal and alternative modes of education.

SARVA SHIKSHA ABHIYAN

Sarva Shiksha Abhiyayan is a project implemented by the central government with the aim of ensuring Universalisation of Elementary education by 2010. SSA was started with the aim of ensuring useful and efficient elementary education for children between the age of 6 and 14 years. SSA is the project which ensures community involvement in school management activities by eliminating social, regional, and gender gaps. 60 per cent of the total out lay of SSA projects and programmes are borne by the central government and 40 per cent by the state governments.

FEATURES OF SSA

- Global basic education with clear planning.
- Quality education across the country.
- Ensuring social justice through basic education.
- Joint action of central and state governments in the matter of education.
- An opportunity to foster basic educational policies and objectives of the state administration.

GOALS OF THE PROGRAMME

- Get all children in schools by 2003 through back-to-school campaign.
- Complete five years of primary education for all children by 2007.
- Focus on elementary education of satisfactory quality with emphasis on education for life.
- Eliminate all gender and social category gaps at primary school level by 2007 and at elementary level by 2010.
- Establish new schools where schools do not exist.
- Provide infrastructural facilities in schools.
- Universal retention by 2010.

SSA goes a step further to DPEP and envisages habitation, as a unit of planning. Habitation form the basis for the preparation of district plans. Each district is required to formulate a perspective plan and an annual plan and budget indicating the required resources needed for achieving the goal of UEE by 2010. Various studies have revealed that invention under SSA has brought considerable expansion quantitatively and qualitatively. While quality still remains as an area of concern, the SSA has been able to increase the schooling facilities, enrolment and retention and achievement gaps between the genders and among social groups.

RASHTRIYA MADHYAMIC SHIKSHA ABHIYAN

Rashtriya Madhyamic Shiksha Abhiyan is major project launched by the central government in march 2009 to increase the access to and improve the quality of secondary education. It aims to increase the enrolment rate by ensuring a secondary school within reasonable distance from every household. It aims to improve the quality of secondary school by ensuring that all secondary schools comply with prescribed standards, eliminate gender bias, remove socio economic barriers and ensure universal access to education at the secondary level.

OBJECTIVES OF THE PROGRAMME

- RMSA envisages to increase secondary enrolment from 52.26 percentage in 2005-06 to 75 percentage.
- Establish secondary schools at reasonable distance within five years of the implementation of RMSA.
- Enhance the quality of education in all secondary schools according to the prescribed standards.
- Eliminate gender discrimination and remove socio-economic and disability barriers.
- Ensure universal access to secondary level education.
- Improve and universalize secondary education by 2020.
- Increase the physical infrastructural facilities of secondary schools by setting up additional classrooms, laboratories, libraries, arts and craft rooms, toilet blocks, drinking water facilities and residential hostels for teachers.
- Increase the quality of secondary education by appointing additional teachers to reduce the teacher student ratio to 1:30 , focusing on science, mathematics and English education, providing in-service training to teachers, establishing science laboratories, ICT based education and implementing new teaching learning reforms through curriculum revision.
- Ensure equity interventions in secondary schools by setting up new schools in SC, ST,minority areas, establishing special enrolment drive for weaker sections, increasing

recruitment of women teachers and providing special extra- curricular facilities for girls.

SARVA SHIKSHA ABHIYAN

The union budget 2018-19 has been proposed to treat school education holistically from pre-primary to 12th standard without having any division.Samagra Shiksha Abhiyan is an initiative aimed at improving the quality of pre-primary to 12th standard education sector as a single unit.It includes three earlier projects namely Sarva Shiksha Abhiyan (SSA),Rashtriya Madhyamic Shiksha Abhiyan (RMSA) and Teacher education.

MAIN OBJECTIVES

- Improve children's learning out comes by providing quality education.
- Bridging social and gender gaps in education.
- Ensure equity and inclusion at all levels of schooling.
- Ensure minimum standards in school education system.
- Promote vocationalisation of education.
- Support states in enacting the right of children to free and compulsory education.(RTE Act 2009).
- Strengthen and modernize the SCERT and DIET as the nodal agencies for education training.
- Strengthen teacher education institutions.

CONCLUSION

An attempt has been made in this article to mention some of the programmes and projects implemented in the field of Indian school education in connection with the National Policy of Education 1986. The impact of the functioning of Minimum Levels of Learning (MLL), District Primary Education Programme (DPEP) ,Sarva Siksha Abhiyan (SSA),Rashtriya Madhyamic Shiksha Abhiyan (RMSA) and Samagra Shiksha Abhiyan on Indian education sector is not insignificant. These programmes and projects have significant impact in the areas of school access, equity, quality education, academic and non-academic facilities and technical education. The impart of such projects in transforming students from being the consumers of knowledge to producers of knowledge is immeasurable. This kind of more research projects, reforms and research studies are essential in the quality oriented Indian education sector. Therefore, we are looking at the National Education Policy 2020, to be implemented in near future, with great hope and expectation. Indian education will become globally recognized and accepted only when this type of innovative programmes and projects are devised and implemented thus easing the way for Indian society being a cosmopolitan one.

References

1) *Ministry of Human Resource Development, Minimum Levels of Learning at Primary stage,.MHRD,New Delhi,1991.*

2) *Dr.Anilkumar Taneja (2017).Indian Education Commissions and Committees.Toward Shaping Education. Swastik Publishers and Distributors. New Delhi. (2017).*

3) *Ministry of Human Resource Development, policy on Education 1986,MHRD, New Delhi (1986).*

4) *K, Pushpanadham. (2001). A study on the decentralized Management of District Primary Education Programme (DPEP). The M.S. University of Baroda, Vadodara .*

5) *NCERT (2000). Fifth Survey of Educational Research, Publication Department, National Council of Educational Research and Training, New Delhi.*

6) *CABE (2005). Universalisation of Secondary Education. MHRD,New Delhi.*

7) *CABE (2005). Girls Education and Common School System. MHRD,New Delhi.*

8) *https://samagra. education.gov.in*

Identification Micro and Macro Nutrients in (soil)Land use Planning of Various Farm of Selected District Yavatmal,Maharashtra

R.P.Bhagat[1], Dr,L.K.Pandey[2] , Dibyendu Gangopadhyay[3],

[1]*Dept.of Biotechnology, Mansarovar Global University, Shehore,(MP)46611,India*

[2]*Prof.Faculty of Life Science, Mansarovar Global University, Shehore,(MP)46611,India*

[3]*Dept.of Microbiology, Mansarovar Global University, Shehore,(MP)46611,India*

Abstract : The present investigation was carried out for assessment of available macro nutrient and micro nutrient status of soil and theiron with soil chemical properties of various farm of selected district Yavatmal, Maharashtra. Total 75 surface soil samples were collected and analyzed for chemical properties and fertility status of soil. The data revealed that the soils of various talukas farm of district Yavatmal, Maharashtra were moderately alkaline (pH 7.38 to 8.49) in soil having safe electrical conductivity (EC 0.21 to 1.21 dSm-1), low to high in organic carbon content (3.1 to 7.7 g kg-1) and non-calcareous to calcareous (43 to 98 g kg-1) in nature. According to the concept of soil nutrient index values the availability of available N (NIV 1), S (NIV 1.50) and Zn (NIV 1) nutrients was in low range and soils were medium in available P (NIV 1.8) and high in available K (NIV 2.86), DTPA- and DTPA-Cu (NIV 3). Further, the organic carbon content showed positive and significant correlation with available Nitrogen (0.210*), Phosphorus (0.205*), Sulphur (0.218*) as well as micronutrients like, Mn (0.265*) and Zn (0.310*)in the soils of various talukas farm of district Yavatmal, Maharashtra. Thus, deficiencies of N, S and Zn were the soil nutritional constraints identified and according to soil site suitability characteristics, land use planning may be done for production of soybean, Channa,Wheat,Jawari and cotton crops at various farm of selected district Yavatmal, Maharashtra.

***Index term*: Micro Nutrients,Macronutrients, zinc,sulphur**

Introduction : The soil word is deribved from Latin word, ‘Solum’ meaning the earthy material in which plant growth occurs. Soil is the natural material spread in different layers. It differs in physical, chemical and mineralogical characteristics. Soil is result of rocks due to environmental processes: weathering and erosion.

Soil ius dynamic material of minerals, organic matter, water, air, bacteria etc. Soill quality varies due to farming, parent material and environmental changes.Soil quality mainly depends on the response of soil to different land use systems and management practices, which may often modify the soil properties and soil productivity . The major and micro nutrients govern the fertility of the soils and control the yield of crops. With the introduction of high yielding crop varieties and intensive agriculture with modernagro-techniques and less use of on farm organic manures, most soils becoming deficient in major and micro nutrients. Such deficiencies of major and micro nutrients affect the growth, yield and nutrition of crops to great extent . Therefore, it is important to know the availability of nutrients in soil as influenced by various factors for the efficient management through external application. Soil fertility evaluation of an area or region is an important aspect in context of sustainable agricultural production. Land is finite natural resource and there is no scope to increase area under cultivation. The food production in India can only be increased by increasing the crop productivity. The higher productivity can only achieved with better information of land and its use. In present dynamic situation accurate, meaningful, current data on land use are essential. Thus, it is necessary to have information on availability of major and micro nutrients of the study area.

Yavatmal district is the region of Western Vidarbha: the part of Maharashtra. In this district and hence in various Taluka region the main crops are cotton, soyabean,jawari,bajari,chana,toor etc. Essential njutrients required for proper growth of plants issupplied by soil. Hence the yield and quality of crop depends on the qulity of soil. Various nutrients are supplied to soil from fertilizers. Productivity of crop is increased by use of various chemical fertilizers on large scale, but it is decreasing the qulity of soil. So, a comprehensive study was undertaken to know the fertility status of soils of land various talukas farm of district Yavatmal, Maharashtra.

Materials and Methods

25 farms of each village were selected for physico-chemical analysis of soil. Total sixteen talukas situated in Yavatmal district were selected for study. Average value of parameter selected farms of a talukas was reported. The study area comes under assured rainfall zone. Soil smaples were collected in the depth of 5-20cm from the surface of soil and were taken in polyhthene bags. The oil samples were collected in the month of June 2022 from different sampling stations. Sample stations used from selected ‘Yavatmal district selected region’

are given in following table-1 and named as S1,S2,S3,S4,S5 and S25.

The soil samples were collected and brought to the laboratory for the study of physic-chemical parameters. The standard methods of soil analysis Black(1965P) and Richard(1945) for particle size distribution,bulk density, maximum water holding capacity, abvilable water capacity, hydraulic conductivity and Piper (1966) for soil pH, electrical conductivity, cation exchange capacity, free calcium carbonate and organic carbon were adopted.

Table No.1 Available nutrient and secondary nutrients in soil of district Yavatmal

	Available nutrient (kg ha-1)			Secondary nutrients					
Sample	N	P2O5	K2O	CaCmol (P+) kg-1	Mg Cmol(P+) kg-1	S (mg kg-1)	pH	EC dSm-1	CaCO3 %
S-1	133		450.31	24.3	13.01	7.4	7.35	0.463	4.6
S-2	124.3	24.94	493.41	38.5	17.2	7.93	7.44	0.226	1.8
S-3	58.5	30.6	414.33	47.8	13.1	7.9	7.66	0.133	1.6
S-4	80.28	23.29	493.41	35.2	12.6	9.15	7.67	0.23	1.6
S-5	101.7	12.77	331.33	48.9	19.2	4.94	7.29	0.467	1.7
S-6	110.4	16.27	187.73	34.1	13.1	4.71	7.2	0.38	1.6
S-7	102.4	10.82	193.04	31.1	19.3	5.3	7.18	0.296	1.9
S-8	103.3	11.23	202.71	29.2	17.5	5.73	7.56	0.234	1.9
S-9	98.64	10.2	214.3	29.9	18.2	5.12	7.96	0.146	1.7
S-10	101.4	10.75	230.49	24.8	18.6	5.31	7.03	0.3	1.6
S-11	58.5	23.29	493.41	40.2	18.1	5.91	7.18	0.226	1.9
S-12	102.4	16.27	331.33	40.8	15.6	7.63	7.03	0.38	1.1
S-13	101.7	11.23	450.31	29.13	17.1	16.9	7.35	0.234	1.8
S-14	98.64	10.75	493.41	18	6.8	3.03	6.94	0.146	1.7
S-15	124.3	27.21	187.73	19.3	9.1	2.22	7.29	0.133	1.9
S-16	103.3	30.6	202.71	18.4	9.4	11.19	7.2	0.467	1.6
S-17	101.7	12.77	230.49	31.2	16.9	7.12	7.2	0.296	1.7
S-18	101.4	10.82	214.3	23.3	11.9	5.1	7.67	0.234	1.6

S-19	80.28	10.2	331.33	17.8	6.9	5.08	7.2	0.23	1.8
S-20	58.5	10.75	414.33	21.8	13.9	5.71	7.13	0.226	1.2
S-21	58.64	12.77	202.71	31.1	13.1	3.03	7.29	0.146	1.1
S-22	124.3	11.23	193.04	29.9	17.2	5.3	7.21	0.3	1.3
S-23	101.7	16.27	331.33	40.2	18.2	5.73	7.32	0.23	1.4
S-24	110.4	23.29	214.3	38.5	18.1	7.63	7.1	0.234	1.9
S-25	103.3	24.94	187.73	34.1	15.6	4.94	7.16	0.226	1.8

RESULT AND DISCUSSION:

The results of study presented in Table 1 indicated that all the soil samples from various farm of district Yavatmal, were moderately alkaline in soil and within safe limit of electrical conductivity . The pH of soil varied from 7.5 to 8.7. Moderately alkaline nature of soil may be due to formation of these soils from basaltic parent material rich in basic cations. Similar findings were reported by Mandal, 1998. Electrical conductivity (EC) of soil varied from 0.11 to 0.24 dSm-1. The organic carbon content varied from 0.8 to 6.9 g kg-1. It indicates that majority of these soils were low to moderately high in organic carbon content. This might be due to increased rate of decomposition of organic matter .The free $CaCO_3$ content varied from 4.1 to 113 g kg-1 indicating calcareous nature of these soils. The available N was found low to moderate in soils of various farm of district Yavatmal, (28.3 to 78.4 kg ha-1) farm of various farm of district Yavatmal, It may be due low organic matter content of soil as well as rapid loss of applied N in soil . The available P content varied from 10.20 to 30.60 kg ha-1 in soils of various talukas farm of district Yavatmal, was categorized as low to moderate range, the low amount of available P may be due to application of lower doses of P fertilizer, fixation of P on clay minerals or $CaCO_3$ surfaces with the time elapsed between fertilizer application and crop uptake. All soil samples were found high with respect to available K content which ranged from 187.73 to 493.41 kg ha-1. This may be due to occurrence of potash rich minerals like mica and feldspar in parent material of the soils . The available S was found deficient in all the samples from various talukas farm of district Yavatmal (5.35 to 14.32mg kg-1) farm. The availability ofS in soil depends on the combined action of factors like nature of parent material, rain fall, clay and organic matter content in soil .The continued soybean cultivation over a long period for seed production might also be a reason for low available S status. The DTPAextractable micronutrients Cu (17.90 - 38.50 mg kg-1), Fe (0.7 - 7.9 mg kg-1), Mn (6.90 – 19.80 mg kg-1) were found sufficient and Zn (0.5 - 1.5 mg kg-1) was found deficient in soils of various talukas farm of district Yavatmal.

Table 2. Nutrient Index values of various talukas farm of district Yavatmal.

Parameters	Parker index	Remark
Organic Carbon	2.48	High
Available Nitrogen	1.00	Low
Available Phosphorus	1.89	Medium
Available Potassium	2.86	High
Available Sulphur	1.50	Low

Available NIV Category nutrients Nitrogen 1.2 Low Phosphorus 1.8 Medium Potassium 3.0 High Sulphur 1 .0 Low Iron 3.0 High Zinc 1.0 Low Manganese 3.0 High Copper 3.0 High Soil nutrient index value : As per the NIV developed by the the nutrient index value for (Table 2) Parkers nutrient index was used to compare the level of soil fertility in the study of various farm of selected district of yavatmal as it is the measure of nutrient supplying capacity of soils. The nutrient index value (NIV) of organic carbon (2.48) and available potassium (2.86) were high in farms of yavatmal. NIV of P was medium (1.89) while N and S were low with values of 1.00 and 1.50 , respectively(Table 2). Pathak (2010) also reported similar results while assessing temporal soil fertility changes in Maharashtra.

The data on soil site suitability characteristics of various blocks of UPRS farm is given in Table 4. On the basis of the analyzed soil properties and the criteria given by NBSS and LUP (1994), the suitability of various crops was determined. The soil site suitability properties of selected district Yavatmal farm observed that these soils were highly suitable for agriculture soil planning and growing crops of soybean, wheat,channa and cotton crops and agriculture production

CONCLUSION :

The study indicated that the soils of Yavatmal district farm slightly alkaline to alkaline in reaction with safe limit of soluble salt content. The OC was medium to high and low in available N. The area showed low to medium in available P and S and high in available K content. The DTPA extractable Zn and Fe was deficient and Mn and Cu were sufficient in the area. The NIV for N and S were low, medium for P and high for K. The results of the study area having potential to identify soil site suitability properties of selected district Yavatmal farm observed that these soils were highly suitable for agriculture planning and growing crops of soybean, wheat,channa and cotton crops and production at various farm of selected district Yavatmal.

ACKNOWLEDGMENT:

We are thankful to honorable Dr.S.Rathod,DMO, District Yavatmal(MH).to give chance work in agricultural study in Yavatmal district. Also thankful to my wife Pooja, my lovely parents Mira Pandurang Bhagat and my family members to supportduring working.

REFERANCES :

1. A.N.Kulkarni,J.B.Balkhande,B.D.Waghamare,P.U.Ratnakar and V.S.Kanwate,Studies of some physic-chemical factors of teak forest from Kinwat area, Nanded(Maharashtra).International journal of life science(2011)page437-438.
2. A.Nagamani,I.K.Kunwar,C.Maroharachary,"Hand Book of Soil Fungi"I.K.Internationals.
3. Ajay Kumar Bhagi,G.R.Chatwal."Environmental Chemistry" Himalaya Publication.
4. 4C.A.Black,Method of soil analysis,part-1,Amar Soc.Agron.Inc.Argon No.9: Madison,USA,1965.
5. Chadar, B., Vaidya, H., Kachave, R., & Aundhkar, V. (2018). Characterizations, classifications and soil site suitability evaluation of soils of farm College of Agriculture, Latur, Maharashtra. *Journal of Pharmacognosy and Phytochemistry*, 166–170. https://www.phytojournal.com/archives/2018/vol7issue1S/PartD/SP-7-1-295.pdf
6. Bagdi, G. L., & Kurothe, R. S. (2014). People's participation in watershed management programmes: Evaluation study of Vidarbha region of Maharashtra in India. *International Soil and Water Conservation Research*, *2*(3), 57–66. https://doi.org/10.1016/S2095-6339(15)30023-X
7. B.P. Bhaskar1*, Sunil Maske2, S.S. Gaikwad1, A. Chaturvedi3, Jagdish Prasad1, S. G. A. and S. K. S. (2017). Soil andland resource evaluation for rural agricultural land use

planning - A case study from hot semiarid ecosystem of Western India. *Archives of Agriculture and Environmental Science*, *2*(3), 206–218.

8. Bhaskar, B. P., Maske, S., Ramesh Kumar, S. C., Srinivas, S., & Hegde, R. (2018). Agropedological Approach for Rural Land

9. Use Planning – A Case Study from Lagadwal Village, Dhule District, Maharashtra, India. *International Journal ofCurrent Microbiology and Applied Sciences*, *7*(2), 2528–2541. https://doi.org/10.20546/ijcmas.2018.702.308

10. Jagdish Prasad,Nagraju,Raveev Shrivastava,S.K.Ray and P.Chandran,Characteristic and Classification of some soil inNagpur district.I.India Soc. Soil sci.49(4),2001,page 735-739
11. Wankhade R.R.(2015), Physico-Chemical Analysis of soil from some Farms of Digras Region ofDistrict Yavatmal in Maharashtra,India. Scholedge International journal of Multidisciplinary and Allied Studies ISSN 2394-336X,2(10),1-4
12. Parker, F.W., Nelson, W.L., Winters, E. and Miles, J.E. 1951. The broad interpretation and application of soil testsummaries. *Agronomy Journal* **43**(3): 103-112.
13. Pathak, H. 2010. Trends of fertility status of Indian soil.*Current Advances in Agricultural Sciences* **2**(1): 10-12.
14. Patil, S.S., Patil, V.C. and Al-Gaadi, K.A. 2011. Spatial variability in fertility status of soil surface soils. *WorldApplied Sciences Journal* **14**(7): 1020-1024.
15. Patil, Y.M. and Sonar, K.R. 1993. Dynamics of potassium in swell-shrink soils of Maharashtra. *Journal of PotassiumResearch* **9:** 315-324.
16. Prabhavati, K., Dasog, G.S., Patil, P.L., Sahrawat, K.L. and Wani, S.P. 2015. Soil fertility mapping using GIS in three agro-climatic zones of Belgaum district, Karnataka. *Journal of the Indian Society of Soil Science* **63**(2): 173-180.
17. Sahrawat, K.L. and Wani, S.P. 2013. Soil testing as a tool for on-farm soil fertility management: experience from the semi-arid zone of India. *Soil Science and Plant Analysis* **44**: 1011-1032.
18. Sharma, R.P., Megh Singh and Sharma, J.P. 2003. Correlation studies on DTPA-extractable micronutrients vis-à-vissoil properties in some soils of Nagaur district in semiarid region of Rajasthan. *Journal of the Indian Society of Soil Science* **51**: 522-527.

MACHINE LEARNING ENABLED PLANT DISEASE DETECTION- A REVIEW

Amit Banga[1] Dr. Rakesh Sharma[2] Ritu Goyal[3] Sarita[4]

[1,3]Research Scholar, Department of ECE, BMU, Rohtak

[2]Associate Professor, Department of ECE, BMU, Rohtak

[4]Assistant Professor, Department of Mathematics, TDLGCW, Murthal

Abstract

Recent advancements in machine learning (ML) and deep learning (DL) have revolutionized the field of plant disease detection, offering automated solutions to enhance agricultural productivity and food security. This paper synthesizes current literature on ML and DL techniques applied to image-based plant disease classification. Key studies reviewed highlight the effectiveness of DL models, particularly convolutional neural networks (CNNs), in accurately identifying and classifying plant diseases from leaf images. The integration of IoT devices for real-time data acquisition and datasets such as Plant Village has significantly improved disease detection capabilities. Comparative analyses of ML and DL architectures underscore the superior performance of DL in handling complex image features and achieving high accuracy rates. Future research directions include optimizing model interpretability, scalability, and addressing data quality challenges to further enhance disease management systems in agriculture.

Keywords: Plant disease detection, Machine Learning, Deep Learning. CNN, IoT.

1. Introduction

Agriculture stands as a cornerstone of human civilization, providing sustenance, economic stability, and cultural identity across the globe. Central to its success is the ability to ensure crop health and productivity. However, the threat posed by plant diseases looms large, significantly impacting crop yields and food security worldwide. Plant diseases, caused by various pathogens including fungi, bacteria, viruses, and pests, result in substantial economic losses each year. These losses not only affect farmers' livelihoods but also contribute to fluctuations in food availability and prices, ultimately impacting global food security. As agriculture faces the dual challenge of feeding an ever-growing population while mitigating environmental impacts, the need for efficient disease management strategies becomes increasingly urgent. Automated disease detection using machine learning and deep learning techniques presents a promising avenue to bolster crop health monitoring and management practices. By enabling early and accurate detection of plant diseases, these technologies offer potential solutions to mitigate losses, optimize resource allocation, and sustainably enhance agricultural productivity.

1.1 Motivation

The imperative for automated disease detection in agriculture using machine learning and deep learning methodologies arises from the increasing complexity and scale of modern agricultural systems. Traditional methods of disease identification often rely on visual inspection by experts, which can be labor-intensive, subjective, and prone to human error. As global population growth continues to escalate demands on food production, ensuring crop health and minimizing losses due to diseases becomes paramount. Machine learning and deep learning techniques offer transformative capabilities by enabling the rapid and accurate analysis of large-scale agricultural data. These technologies can leverage vast datasets to detect subtle patterns indicative of disease onset before symptoms become visually apparent. By facilitating early detection, farmers can implement timely interventions such as targeted pesticide application or precision irrigation, thereby optimizing resource usage and minimizing environmental impacts. Moreover, automated disease detection systems can provide real-time monitoring across diverse geographical regions, supporting proactive disease management strategies and enhancing overall agricultural resilience. In essence, the integration of machine learning and deep learning in agriculture not only promises to revolutionize disease detection but also to bolster global food security by safeguarding crop health more effectively than ever before.

Figure 1. Locally Captured Images

2. Literature Review

The application of machine learning (ML) and deep learning (DL) in agriculture, particularly for plant disease detection, has seen significant advancements. This literature review synthesizes recent research on these technologies, highlighting their methodologies, applications, and outcomes in improving plant health management. Sajitha et al. (2024) [1] provide a comprehensive review of ML and DL techniques for image-based plant disease classification tailored for industrial farming systems. This research underscores the potential of integrating ML and DL into industrial agriculture to enhance disease management and improve yield. Khalid and Karan (2024) [2] focus on the use of DL for plant disease detection, exploring different DL architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Their study demonstrates the high accuracy and robustness of DL models in identifying various plant diseases from image data. Demilie (2024) [3] conducts a comparative study of various plant disease detection and classification techniques. By comparing these techniques, the research provides insights into the strengths and limitations of each approach, guiding the selection of appropriate models for specific applications. Adekunle et al. (2024) present a study on the application of DL techniques for plant disease detection. Their research involves training and testing DL models on datasets comprising images of healthy and diseased plants. The results indicate that DL models, particularly those based on CNNs, achieve high accuracy in detecting and classifying plant diseases. Reis and Turk (2024) [5] propose a novel DL model combining depthwise separable convolution and transformer networks for potato leaf disease detection. Their model leverages the strengths of both convolutional and transformer architectures to improve the accuracy and efficiency of disease detection. The study demonstrates that this hybrid approach outperforms traditional DL models, offering better generalization and robustness.

Prasad and Thyagaraju (2024) [6] present a comprehensive review of early plant disease detection using IoT, ML, and DL. They emphasize the integration of these technologies to enhance the accuracy and efficiency of disease detection. IoT devices collect real-time data from plant leaves, which are then analyzed using ML and DL algorithms to identify disease symptoms at an early stage. Kulkarni and Shastri (2024) [7] focus on the application of ML techniques specifically for rice leaf disease detection. Their study explores various ML algorithms, including Support Vector Machines (SVM) and Random Forest (RF), to classify different types of rice leaf diseases. Bouacida et al. (2024) [8] propose an innovative DL approach for cross-crop plant disease detection. Their method involves training a generalized model capable of identifying unhealthy leaves across different crop species. The study uses Convolutional Neural Networks (CNNs) to analyze images of leaves and detect disease symptoms Chin, Ng, and Palanichamy (2024) [9] conduct a comparative study of various DL methods for plant disease detection and classification. They evaluate the performance of different architectures, including CNNs, Recurrent Neural Networks (RNNs), and hybrid models. Their findings suggest that CNN-based models consistently outperform other methods in terms of accuracy and processing time Dubey and Choubey (2024) [10] introduce an efficient adaptive feature selection technique combined with a DL model for classifying paddyplant leaf diseases. Their approach involves selecting the most relevant features from the leaf images to improve the classification accuracy of the DL model.

Sarkar et al. (2023) [11] provide a comprehensive review of ML and DL applications in leaf disease detection. They discuss various ML algorithms such as Decision Trees, k-Nearest Neighbors (k-NN), and ensemble methods, along with DL techniques like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The review emphasizes the strengths and limitations of these approaches and highlights their effectiveness in achieving high accuracy in disease classification tasks. Ahmed and Yadav (2023) [12] focus on ML approaches specifically tailored for plant disease detection. They explore the application of supervised learning techniques including SVM, Random Forest, and Naive Bayes for classifying diseased and healthy plants based on leaf images. The study underscores the importance of feature engineering and dataset augmentation in improving model performance. Shahi et al. (2023) [13] discuss recent advances in crop disease detection using Unmanned Aerial Vehicles (UAVs) combined with DL techniques. They highlight the use of high-resolution aerial images captured by UAVs for early detection of diseases in large agricultural fields. DL models such as CNNs are employed for image analysis, enabling rapid and accurate identification of disease symptom. Chug et al. [14] (2023) propose a novel hybrid DL framework for image-based plant disease detection. Their approach integrates CNNs with other DL architectures or traditional ML algorithms to leverage the strengths of both paradigms. By combining feature extraction capabilities of CNNs with the interpretability of traditional ML models, the framework achieves robust performance in disease classification tasks. Ahmad et al. (2023) [15]

Shoaib et al. (2023) [16] provide a comprehensive review focusing on advanced deep learning models for plant disease detection. They highlight the rapid progress in leveraging convolutional neural networks (CNNs) and other deep learning architectures to achieve high accuracy in identifying diseases across different plant species. Moupojou et al. (2023) [17] introduce the "FieldPlant" dataset tailored for plant disease detection. This dataset addresses the need for large-scale, diverse image collections essential for training robust deep learning models. Their work underscores the critical role of data quality and quantity in enhancing model performance. Kotwal et al. (2023) [18] discuss the transition from traditional methods to deep learning techniques in agricultural plant disease identification. They emphasize how deep learning surpasses conventional approaches by automating and improving the accuracy of disease diagnosis, thereby aiding timely interventions to mitigate crop losses. Rajpoot et al. (2023) [19] propose hybrid deep learning and machine learning methods for early detection of rice leaf diseases. Their approach combines the strengths of deep learning in image feature extraction with the interpretability of machine learning models, demonstrating effective early warning systems for agricultural management. Haridasan et al. (2023) [20] present a deep learning system specifically designed for paddy plant disease detection. Their study illustrates the practical application of CNNs in environmental monitoring, highlighting the system's capability to monitor and assess disease prevalence in paddy fields. By integrating deep learning innovations with domain expertise, researchers can foster a more resilient agricultural sector capable of addressing emerging challenges posed by plant diseases in a rapidly changing climate.

Jackulin and Murugavalli (2022) [21] provide a comprehensive review on the detection of plant diseases using both machine learning and deep learning approaches. They highlight the evolution from traditional ML methods to more advanced DL models, emphasizing the improved accuracy and efficiency achieved through convolutional neural networks (CNNs) and other DL architectures. Kumar et al. (2022) [22] conducted a systematic analysis, albeit retracted later, on ML and DL-based approaches for plant leaf disease classification. Despite the retraction, their initial findings likely discussed the comparative strengths and limitations of different algorithms in accurately identifying and classifying plant diseases from leaf images. Hassan and Maji (2022) [23] proposed a novel CNN architecture specifically tailored for plant disease identification. Their study focused on optimizing network design and training parameters to achieve robust performance in detecting diseases across various plant species. Sujatha et al. (2021) [24] evaluated the performance of DL versus traditional ML methods in plant leaf disease detection. Their comparative analysis highlighted DL's superiority in handling complex image features and achieving higher accuracy rates, thereby validating DL's effectiveness in practical agricultural applications. Khan et al. (2021) [25] traced the evolutionary journey from classical ML techniques to DL in image-based plant disease detection.

They discussed the technological advancements and paradigm shifts that have accelerated the adoption of DL models in agriculture, enabling more precise and timely disease diagnosis. Li et al. (2021) [26] provide an extensive review of plant disease detection and classification using deep learning methodologies. Their study emphasizes the evolution of DL models, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their variants, in achieving high accuracy and efficiency in disease identification across various crop species. Tan et al. (2021) [27] compare classical machine learning with deep learning methods specifically for tomato leaf disease classification. Their research highlights DL's superior performance in handling complex image data and extracting meaningful features compared to traditional ML algorithms, showcasing DL's potential for precise disease diagnosis in agriculture. Ouhami et al. (2021) [28] survey the integration of computer vision, IoT (Internet of Things), and data fusion techniques for crop disease detection using ML approaches.

Their study underscores the importance of real-time data acquisition and processing capabilities enabled by IoT in enhancing the accuracy and scalability of disease detection systems deployed in agricultural settings. Chohan et al. (2020) [29] focus on the application of deep learning techniques for plant disease detection. Their study explores various DL architectures and methodologies employed to automate disease identification processes, thereby facilitating early intervention and management practices crucial for sustainable agriculture. Mohameth et al. (2020) [30] discuss the integration of deep learning with feature extraction techniques using datasets like Plant Village for plant disease detection. Their research highlights the effectiveness of combining DL's ability to learn discriminative features from images with domain-specific feature extraction methods, enhancing overall detection accuracy and robustness. Hernández and López (2020) [31] explore uncertainty quantification in plant disease detection using Bayesian deep learning methods. Their study highlights the importance of probabilistic approaches in assessing the confidence of DL models' predictions, crucial for decision-making in precision agriculture and disease management. Saleem et al. (2020) [32] investigate image-based plant disease identification using deep learning meta-architectures. They evaluate different DL frameworks and architectures to optimize disease classification accuracy, emphasizing the role of model selection and tuning in achieving robust performance across diverse plant species. Jasim and Al-Tuwaijari (2020) [33] employ image processing and DL techniques for plant leaf disease detection and classification. Their approach integrates preprocessing steps with DL models to enhance feature extraction from leaf images, demonstrating effective disease diagnosis capabilities in agricultural settings.

Saleem et al. (2019) [34] present a comprehensive study on plant disease detection and classification exclusively using DL methods. Their research covers advancements in CNNs and other DL architectures tailored for accurate and efficient disease identification, contributing to the automation of crop health monitoring. Türkoğlu and Hanbay (2019) [35] focus on the detection of plant diseases and pests using DL-based features. Their study explores the integration of feature extraction techniques with DL models to discriminate between disease symptoms and pest infestations, enhancing diagnostic precision in agricultural systems.Too et al. (2019) [36] conducted a comparative study focusing on the fine-tuning of deep learning models for plant disease identification. They evaluated different deep learning architectures and highlighted the significance of model selection and parameter tuning in achieving high classification accuracy. Their results demonstrated that transfer learning, combined with appropriate fine-tuning, can be an effective approach to improve the accuracy of plant disease detection systems. Ramesh et al. (2018) [37] presented a comprehensive approach using machine learning techniques for plant disease detection. Their study involved the design and implementation of a system that employed various machine learning algorithmsto classify plant diseases based on image data. Their findings suggested that integrating multiple features and optimizing machine learning models could lead to better detection rates and reduced false positives. Barbedo (2018) [38] conducted a comprehensive study to identify the factors influencing the use of deep learning for plant disease recognition. The study, published in Biosystems Engineering, emphasizes that the quality and quantity of training dataare critical for the success of deep learning models. Barbedo notes that obtaining high-qualityimages of diseased plants under various conditions is challenging but essential for training robust models. Additionally, the study highlights the importance of model architecture, suggesting that deeper networks can capture more complex patterns but at the cost of increased computational resources.

Mohanty, Hughes, and Salathé (2016) [39] investigated the use of deep learning for image-based plant disease detection. Published in Frontiers in Plant Science, their study presents a deep convolutional neural network (CNN) model trained on a large dataset of plant images. The authors report that their model achieved high accuracy in identifying and classifying plant diseases from images, outperforming traditional machine learning methods. The study highlights the scalability of deep learning models, noting that they can be trained on diverse datasets to recognize a wide range of plant diseases. Venkataramanan, Honakeri, and Agarwal (2019) [40] explored the use of deep neural networks for plant disease detection and classification in their study published in the International Journal of Computer Science and Engineering. The authors developed a deep neural network model specifically designed to classify plant diseases from images. Their model demonstrated high classification accuracy, particularly when dealing with complex disease symptoms that are difficult to differentiate using traditional methods. The study emphasizes the importance of data augmentation techniques to enhance model performance, especially when dealing with limited datasets.

Table 2. The Dataset Used in each study for plant disease detection using machine learning (ML) and deep learning (DL)

Study	Focus	Methods/Architectures	Dataset Used	Key Findings
Demilie (2024)	Comparative study of disease detection techniques	Various ML, DL methods	Various datasets	Provides insights into strengths and limitations of different detection techniques for specific applications.
Reis and Turk (2024)	Hybrid DL model for potato leaf disease detection	Depthwise separable convolutions, transformers	Potato leaf disease dataset	Hybrid approach improves accuracy and efficiency in disease detection compared to traditional DL models.
Kulkarni and Shastri (2024)	ML techniques for rice leaf disease detection	SVM, RF	Rice leaf disease dataset	Evaluates SVM, RF for classifying rice leaf diseases, emphasizing algorithm performance in agriculture.
Bouacida et al. (2024)	DL for cross-crop disease detection	CNNs	Cross-crop disease dataset	Proposes a generalized model for detecting plant diseases across different crop species.
Dubey and Choubey (2024)	Adaptive feature selection with DL for paddy leaf diseases	DL with adaptive feature selection	Paddy leaf disease dataset	Enhances classification accuracy by selecting relevant features from leaf images.
Ahmed and Yadav (2023)	ML for plant disease detection	SVM, RF, Naive Bayes	Plant disease dataset	Emphasizes feature engineering and dataset augmentation in improving model performance.
Shahi et al. (2023)	Crop disease detection using UAVs combined with DL	CNNs	UAV-captured aerial images	Uses high-resolution UAV images for early detection of diseases in large agricultural fields.
Moupojou et al. (2023)	"FieldPlant" dataset for plant disease detection	DL with "FieldPlant" dataset	"FieldPlant" dataset	Emphasizes the importance of large-scale, diverse datasets in training robust DL models for disease detection.
Rajpoot et al. (2023)	Hybrid DL-ML methods for early detection of rice leaf diseases	Hybrid DL-ML models	Rice leaf disease dataset	Combines DL feature extraction with ML interpretability for

				effective early warning systems.
Haridasan et al. (2023)	DL system for paddy plant disease detection	CNNs	Paddy plant disease dataset	Illustrates practical application of CNNs in monitoring and assessing disease prevalence in paddy fields.
Tan et al. (2021)	Comparison of ML with DL for tomato leaf disease classification	ML, DL	Tomato leaf disease dataset	DL outperforms ML in extracting meaningful features for accurate disease diagnosis in tomatoes.
Mohameth et al. (2020)	DL with feature extraction for plant disease detection	DL, feature extraction techniques	Plant Village dataset	Combines DL's feature learning with domain-specific extraction methods for robust disease detection.
Mohanty et al. (2016)	DL for image-based plant disease detection	Deep CNN models	Plant Village dataset	Scalable DL models for accurate identification and classification of plant diseases from images.

3. Datasets

Central to the development and evaluation of machine learning and deep learning models for plant disease detection are the datasets used. These datasets serve as foundational resources that enable researchers and practitioners to train, validate, and benchmark their algorithms effectively. The characteristics and quality of these datasets significantly influence the performance and generalizability of the models deployed in agricultural settings. Datasets used for plant disease detection typically encompass images of plants exhibiting various diseases, captured under controlled or field conditions. These images may originate from publicly available repositories, research institutions, agricultural extension services, or collaborations with farmers and agronomists. Examples of well-known datasets include the PlantVillage dataset, which contains thousands of images across multiple crop species and disease types, and datasets specific to certain crops like grapevine diseases or wheat rust.

Annotated datasets play a crucial role, as they provide ground truth labels that indicate the presence or absence of specific diseases in plant images. Annotation processes may involve expert agronomists or crowdsourced efforts to ensure accuracy and consistency. Detailed annotations are essential for training machine learning models to recognize disease symptoms accurately and reliably. The size and diversity of datasets are critical factors influencing model performance and robustness. Larger datasets with diverse samples across different geographic regions, climates, and growth stages help models generalize better to real-world conditions. Moreover, datasets covering a wide range of crops and diseases enable researchers to develop comprehensive disease detection systems applicable across various agricultural contexts.

Preprocessing techniques such as image normalization, cropping, and augmentation (e.g., rotation, flipping) are often applied to enhance dataset quality and diversity. These techniques help mitigate biases, improve model robustness, and facilitate effective model training.

Figure 2. Sample of Images from PlantVillage Dataset

Table 2. Structure a dataset table for plant disease detection

Dataset Name	Crop Type	Disease Types	Number of Images	Sources	Annotation Method	Preprocessing Applied
PlantVillage Dataset	Tomato	Early Blight, Late Blight, Septoria	10,000	PlantVillage repository	Expert annotation	Normalization, cropping
Wheat Rust Dataset	Wheat	Stripe Rust, Stem Rust	5,000	Research institution	Crowdsourced annotation	Resizing, rotation
Grapevine Dataset	Grapevine	Downy Mildew, Powdery Mildew	7,500	Agricultural research project	Expert annotation	Color adjustment, augmentation
Citrus Dataset	Citrus	Citrus Canker, Citrus Greening	3,500	Collaboration with citrus growers	Expert annotation	Cropping, resizing

4. Conclusion

The application of machine learning (ML) and deep learning (DL) in plant disease detection represents a transformative advancement in agricultural technology. This literature review has synthesized recent research that underscores the efficacy of ML and DL techniques in automating and improving the accuracy of disease identification across various crop species.

Studies reviewed here demonstrate that DL models, particularly convolutional neural networks (CNNs), are highly effective in extracting intricate features from plant images, therebyenabling precise and early detection of diseases. The integration of IoT devices for real-time data acquisition further enhances the scalability and efficiency of disease management systemsin agriculture. The reviewed literature also highlights the critical role of datasets in training robust ML and DL models. Diverse datasets such as specific crop disease collections and broader datasets like Plant Village have facilitated the development and benchmarking of thesemodels, leading to significant improvements in disease classification accuracy and scalability.Studies comparing different ML and DL architectures emphasize the importance of model selection and parameter optimization in achieving optimal performance. Looking forward, continued advancements in DL architectures, coupled with the integration of emerging technologies like UAVs and IoT, hold promise for further enhancing disease detection capabilities in agriculture. Addressing challenges such as data quality, model interpretability, and scalability will be crucial for deploying these technologies effectively across diverse agricultural landscapes. The reviewed literature underscores the transformative potential of MLand DL in revolutionizing plant disease management. By leveraging these technologies,agricultural stakeholders can mitigate crop losses, optimize resource use, and contribute to global food security in a sustainable manner.

REFERENCES

1. Sajitha, P., Andrushia, A. D., Anand, N., & Naser, M. Z. (2024). A Review on Machine Learning and Deep Learning Image-based Plant Disease Classification for Industrial Farming Systems. *Journal of Industrial Information Integration*, 100572.
2. Khalid, M. M., & Karan, O. (2024). Deep learning for plant disease detection. *International Journal of Mathematics, Statistics, and Computer Science*, *2,* 75-84.
3. Demilie, W. B. (2024). Plant disease detection and classification techniques: a comparative study of the performances. *Journal of Big Data*, *11*(1), 5.
4. Adekunle, T. S., Lawrence, M. O., Alabi, O. O., Afolorunso, A. A., Ebong, G. N., & Oladipupo, M. A. (2024). Deep learning technique for plant disease detection. *Computer Science and Information Technologies*, *5*(1), 55-62.
5. Reis, H. C., & Turk, V. (2024). Potato leaf disease detection with a novel deep learning model based on depthwise separable convolution and transformer networks. *Engineering Applications of Artificial Intelligence*, *133*, 108307.
6. Prasad, S. R., & Thyagaraju, G. S. (2024). Leaf analysis based early plant disease detection using Internet of Things, Machine Learning and Deep Learning: A comprehensive review. *Journal of Integrated Science and Technology*, *12*(2), 734-734.
7. Kulkarni, P., & Shastri, S. (2024). Rice Leaf Diseases Detection Using Machine Learning. *Journal of Scientific Research and Technology*, 17-22.
8. Bouacida, I., Farou, B., Djakhdjakha, L., Seridi, H., & Kurulay, M. (2024). Innovative deep learning approach for cross-crop plant disease detection: A generalized method for identifying unhealthy leaves. *Information Processing in Agriculture*.
9. Chin, P. W., Ng, K. W., & Palanichamy, N. (2024). Plant Disease Detection and Classification Using Deep Learning Methods: A Comparison Study. *Journal of Informatics and Web Engineering*, *3*(1), 155-168.

10. Dubey, R. K., & Choubey, D. K. (2024). An efficient adaptive feature selection with deep learning model-based paddy plant leaf disease classification. *Multimedia Tools and Applications*, *83*(8), 22639-22661.
11. Sarkar, C., Gupta, D., Gupta, U., & Hazarika, B. B. (2023). Leaf disease detection using machine learning and deep learning: Review and challenges. *Applied Soft Computing*, 110534.
12. Ahmed, I., & Yadav, P. K. (2023). Plant disease detection using machine learning approaches. *Expert Systems*, *40*(5), e13136.
13. Shahi, T. B., Xu, C. Y., Neupane, A., & Guo, W. (2023). Recent advances in crop disease detection using UAV and deep learning techniques. *Remote Sensing*, *15*(9), 2450.
14. Chug, A., Bhatia, A., Singh, A. P., & Singh, D. (2023). A novel framework for image-based plant disease detection using hybrid deep learning approach. *Soft Computing*, *27*(18), 13613-13638.
15. Ahmad, A., Saraswat, D., & El Gamal, A. (2023). A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. *Smart Agricultural Technology*, *3*, 100083.
16. Shoaib, M., Shah, B., Ei-Sappagh, S., Ali, A., Ullah, A., Alenezi, F., ... & Ali, F. (2023). An advanced deep learning models-based plant disease detection: A review of recent research. *Frontiers in Plant Science*, *14*, 1158933.
17. Moupojou, E., Tagne, A., Retraint, F., Tadonkemwa, A., Wilfried, D., Tapamo, H., & Nkenlifack, M. (2023). FieldPlant: A dataset of field plant images for plant disease detection and classification with deep learning. *IEEE Access*, *11*, 35398-35410.
18. Kotwal, J., Kashyap, R., & Pathan, S. (2023). Agricultural plant diseases identification: From traditional approach to deep learning. *Materials Today: Proceedings*, *80*, 344-356.
19. Rajpoot, V., Tiwari, A., & Jalal, A. S. (2023). Automatic early detection of rice leaf diseases using hybrid deep learning and machine learning methods. *Multimedia Tools and Applications*, *82*(23), 36091-36117.
20. Haridasan, A., Thomas, J., & Raj, E. D. (2023). Deep learning system for paddy plant disease detection and classification. *Environmental monitoring and assessment*, *195*(1), 120.
21. Jackulin, C., & Murugavalli, S. (2022). A comprehensive review on detection of plant disease using machine learning and deep learning approaches. *Measurement: Sensors*, *24*, 100441.
22. Kumar, R., Chug, A., Singh, A. P., & Singh, D. (2022). [Retracted] A Systematic Analysis of Machine Learning and Deep Learning Based Approaches for Plant Leaf Disease Classification: A Review. *Journal of Sensors*, *2022*(1), 3287561.
23. Hassan, S. M., & Maji, A. K. (2022). Plant disease identification using a novel convolutional neural network. *IEEE Access*, *10*, 5390-5401.
24. Sujatha, R., Chatterjee, J. M., Jhanjhi, N. Z., & Brohi, S. N. (2021). Performance of deep learning vs machine learning in plant leaf disease detection. *Microprocessors and Microsystems*, *80*, 103615.
25. Khan, R. U., Khan, K., Albattah, W., & Qamar, A. M. (2021). Image-based detection of plant diseases: From classical machine learning to deep learning journey. *Wireless Communications and Mobile Computing*, *2021*(1), 5541859.
26. Li, L., Zhang, S., & Wang, B. (2021). Plant disease detection and classification by deep learning—a review. *IEEE Access*, *9*, 56683-56698.

27. Tan, L., Lu, J., & Jiang, H. (2021). Tomato leaf diseases classification based on leaf images: a comparison between classical machine learning and deep learning methods. *AgriEngineering*, *3*(3), 542-558.
28. Ouhami, M., Hafiane, A., Es-Saady, Y., El Hajji, M., & Canals, R. (2021). Computer vision, IoT and data fusion for crop disease detection using machine learning: A survey and ongoing research. *Remote Sensing*, *13*(13), 2486.
29. Chohan, M., Khan, A., Chohan, R., Katpar, S. H., & Mahar, M. S. (2020). Plant disease detection using deep learning. *International Journal of Recent Technology and Engineering*, *9*(1), 909-914.
30. Mohameth, F., Bingcai, C., & Sada, K. A. (2020). Plant disease detection with deep learning and feature extraction using plant village. *Journal of Computer and Communications*, *8*(6), 10-22.
31. Hernández, S., & López, J. L. (2020). Uncertainty quantification for plant disease detection using Bayesian deep learning. *Applied Soft Computing*, *96*, 106597.
32. Saleem, M. H., Khanchi, S., Potgieter, J., & Arif, K. M. (2020). Image-based plant disease identification by deep learning meta-architectures. *Plants*, *9*(11), 1451.
33. Jasim, M. A., & Al-Tuwaijari, J. M. (2020, April). Plant leaf diseases detection and classification using image processing and deep learning techniques. In *2020 International Conference on Computer Science and Software Engineering (CSASE)* (pp. 259-265). IEEE.
34. Saleem, M. H., Potgieter, J., & Arif, K. M. (2019). Plant disease detection and classification by deep learning. *Plants*, *8*(11), 468.
35. Türkoğlu, M., & Hanbay, D. (2019). Plant disease and pest detection using deep learning-based features. *Turkish Journal of Electrical Engineering and Computer Sciences*, *27*(3), 1636-1651.
36. Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture*, *161*, 272-279.
37. Ramesh, S., Hebbar, R., Niveditha, M., Pooja, R., Shashank, N., & Vinod, P. V. (2018, April). Plant disease detection using machine learning. In *2018 International conference on design innovations for 3Cs compute communicate control (ICDI3C)* (pp. 41-45). IEEE.
38. Barbedo, J. G. (2018). Factors influencing the use of deep learning for plant disease recognition. *Biosystems engineering*, *172*, 84-91.
39. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in plant science*, *7*, 215232.
40. Venkataramanan, A., Honakeri, D. K. P., & Agarwal, P. (2019). Plant disease detection and classification using deep neural networks. *Int. J. Comput. Sci. Eng*, *11*(9), 40-46.

A COMPREHENSIVE REVIEW ON COLON CANCER DETECTION USING MACHINE LEARNING APPROACH

Ritu Goyal[1] Dr. Rakesh Sharma[2] Amit Banga[3]

[1,3]Research Scholar, Department of ECE, BMU, Rohtak

[2]Associate Professor, Department of ECE, BMU, Rohtak

Abstract

The detection of canon cancer, a pervasive and potentially fatal condition, remains a critical challenge in modern healthcare. As traditional diagnostic methods often exhibit limitations in accuracy and efficiency, the integration of machine learning (ML) techniques has garnered significant interest for improving detection outcomes. This paper presents a comprehensive review aimed at elucidating the current landscape of ML approaches utilized in canon cancer detection, encompassing methodologies, advancements, datasets, and challenges. Canon cancer, encompassing a spectrum of malignancies affecting various organs, presents a formidable healthcare burden worldwide. Early detection is paramount for effective treatment and prognosis improvement. While conventional diagnostic modalities such as imaging and histopathology play pivotal roles, their reliance on subjective interpretation and limited scalability necessitate alternative approaches. Machine learning, with its capacity for automated pattern recognition and data-driven analysis, offers promising avenues for enhancing diagnostic accuracy and efficiency. ML techniques deployed in canon cancer detection span a diverse spectrum, including supervised, unsupervised, and semi-supervised learning paradigms. Supervised methods, notably classification algorithms such as support vector machines (SVMs), random forests, and deep neural networks, have demonstrated efficacy in discerning malignant from benign lesions. Unsupervised approaches, encompassing clustering and anomaly detection, facilitate exploratory analysis and pattern discovery in large-scale datasets. Recent advancements in ML-driven canon cancer detection have been propelled by breakthroughs in imaging technologies, data availability, and algorithmic sophistication. Radiomics, an interdisciplinary field encompassing the extraction and analysis of quantitative imaging features, has enabled the development of robust predictive models leveraging imaging biomarkers. Deep learning architectures, particularly convolutional neural networks (CNNs),

have revolutionized medical image analysis by enabling end-to-end learning from raw data, circumventing the need for handcrafted feature engineering.

Keywords: Canon cancer, Machine learning, Diagnosis, Classification, Feature extraction, Imaging techniques, Deep learning, Convolutional neural networks (CNNs), Radiomics, Biomarkers, Data augmentation, Performance evaluation.

1. INTRODUCTION

Canon cancer, comprising a wide array of malignancies affecting diverse organs, represents a significant public health concern worldwide. Despite advancements in medical technology and treatment modalities, the timely and accurate detection of canon cancer remains a formidable challenge. Early diagnosis is pivotal for effective intervention, prognosis improvement, and ultimately, patient survival. However, conventional diagnostic approaches, including imaging modalities and histopathological analysis, often suffer from limitations such as subjective interpretation, interobserver variability, and suboptimal sensitivity and specificity. In recent years, the burgeoning field of machine learning (ML) has emerged as a transformative force in healthcare, offering novel avenues for improving diagnostic accuracy, efficiency, and patient outcomes. ML techniques, leveraging algorithms capable of learning from data and identifying complex patterns, hold immense promise for revolutionizing canon cancer detection. By harnessing large-scale datasets, advanced computational methodologies, and sophisticated predictive models, ML empowers clinicians with powerful tools for early detection, risk stratification, and personalized treatment planning.

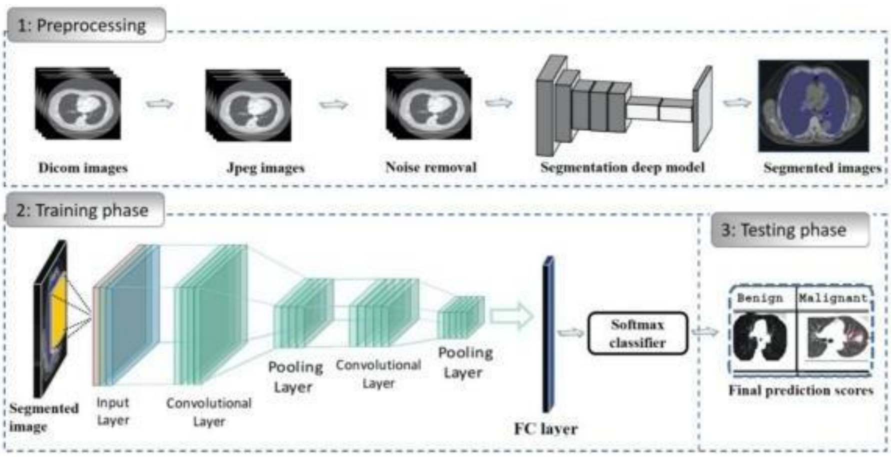


Figure 1. Generalized machine learning framework for lung cancer prediction

This paper aims to provide a comprehensive review of the current landscape of ML approaches utilized in canon cancer detection. By synthesizing recent advancements, methodologies, datasets, and challenges, we seek to elucidate the state-of-the-art in ML-driven diagnostic strategies for canon cancer. Through an in-depth exploration of various ML paradigms, including supervised, unsupervised, and semi-supervised learning, we aim to highlight the diverse array of methodologies employed in discerning malignant from benign lesions across different organ systems. This review will delve into recent innovations and breakthroughs in ML-driven canon cancer detection, fueled by advancements in imaging technologies, data availability, and algorithmic sophistication. From the emergence of radiomics, enabling the extraction and analysis of quantitative imaging features, to the advent of deep learning architectures, particularly convolutional neural networks (CNNs), capable of learning intricate patterns directly from raw data, the landscape of ML in medical imaging is rapidly evolving.

However, alongside these advancements come notable challenges and considerations. The availability and standardization of annotated datasets, the generalizability of models across diverse patient populations and imaging modalities, and the interpretability and explainability of ML-driven diagnostic systems represent critical areas for further research and development. Moreover, the integration of ML into clinical practice necessitates rigorous validation, regulatory scrutiny, and ethical considerations to ensure patient safety and trust in algorithmic decision-making. This comprehensive review aims to provide a nuanced understanding of the role of ML in canon cancer detection, offering insights into current methodologies, advancements, datasets, and challenges. By elucidating the opportunities and complexities inherent in ML-driven diagnostic strategies, we hope to catalyze further research efforts aimed at harnessing the full potential of ML in transforming the landscape of cancer diagnosis and treatment.

1.2 Current Challenges in Canon Cancer Detection

Canon cancer, encompassing a broad spectrum of malignancies affecting various organs, presents formidable challenges in detection and diagnosis. Despite advancements in medical technology and treatment modalities, several obstacles hinder the timely and accurate identification of canon cancer. Understanding and addressing these challenges are crucial for improving patient outcomes and reducing the burden of this pervasive disease.

One of the primary challenges in canon cancer detection lies in the subjective interpretation of imaging studies. Radiological imaging techniques such as computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) play a pivotal role in diagnosing canon cancer by providing detailed anatomical and functional information. However, the interpretation of these images often relies on the expertise and experience of radiologists, leading to variability in diagnostic accuracy and consistency. Variations in interpretation can result in missed or incorrect diagnoses, delaying appropriate treatment and impacting patient outcomes.

Histopathological analysis, another cornerstone of canon cancer diagnosis, is susceptible to interobserver variability. Tissue biopsy specimens are examined by pathologists to determine the presence of cancerous cells and characterize tumor morphology. However, different pathologists may interpret the same specimen differently, leading to discrepancies in diagnosis and treatment decisions. Moreover, the complex histological features of some canon cancers, such as poorly differentiated tumors, further exacerbate the challenge of accurate diagnosis. Standardizing histopathological assessment criteria and implementing quality assurance measures are essential steps toward reducing variability and improving diagnostic consistency. The sensitivity and specificity of existing diagnostic modalities for canon cancer detection are not optimal. While imaging techniques and biomarker assays have improved over time, they may still miss small or early-stage tumors, leading to false-negative results. Conversely, false-positive findings can occur due to benign conditions or artifacts, leading to unnecessary invasive procedures and patient anxiety. Enhancing the accuracy of diagnostic tests through the development of novel imaging protocols, biomarkers, and molecular assays is critical for improving early detection rates andreducing diagnostic uncertainty.

The heterogeneity of canon cancer poses a significant challenge to accurate detection and classification. Canon cancers can arise from different cell types within the same organ or exhibit varying molecular profiles, contributing to diverse clinical presentations and treatment responses. Additionally, tumors may evolve over time, acquiring genetic mutations and phenotypic changes that impact their behavior and therapeutic vulnerabilities. Consequently, designing diagnostic strategies that account for the complexity and diversity of canon cancer is essential for tailoring treatment approaches to individual patients and improving outcomes. Disparities in access to diagnostic services and expertise exacerbate the challenges of canon cancer detection, particularly in underserved communities and low-resource settings. Limited access to advanced imaging technologies, trained healthcare professionals, and specialized pathology services can delay diagnosis and impede timely intervention. Addressing these disparities requires a multifaceted approach, including investment in healthcare infrastructure, education, and outreach programs to ensure equitable access to high-quality diagnostic care for all patients.

Canon cancer detection faces several challenges, including subjective interpretation of imaging studies, interobserver variability in histopathological analysis, limitations in diagnostic sensitivity and specificity, tumor heterogeneity, and disparities in access to diagnostic services. Overcoming these obstacles requires collaborative efforts from healthcare professionals, researchers, policymakers, and industry stakeholders to develop and implement innovative diagnostic approaches that improve accuracy, consistency, and accessibility.

Table 1. This table gives a summary of recent work that has been performed in cancer detection using machine learning and deep learning algorithms

Feature extraction	**Data**	**ML/DL**	**Acc (%)**
2018 taxic weights, phylogenetic trees	LIDC-IDRI [23]	CNNs	92.6
2018 SCM	LIDC-IDRI [23]	MLP, k−NN, SVM	96.7
2018 histogram equalization	JSRT [22], ChestX-ray14 [21]	DenseNet	74.4
2019 –	UCI [26]	SVM, LR, DT, Naive Bayes	99.2
2019 AdaBoost	ELVIRA biomedical data [27]	ANN	99.7
2019 UNet and ResNet	LIDC-IDRI [23]	XGBoost and RF	84.0
2020 –	spectroscopic data	ResNet	95.0

2020 HoG, LBP, SIFT, Zernike Moment	LIDC-IDRI [23]	FPSOCNN	95.6
2021 2D-DFT and 2D-DWT	LC25000 images [24]	CNNs	96.3
2021 Correlation Attribute (CA)	UCI [26]	CNN, SVM, k-NN	95.5

1.3 The Role of Machine Learning in Addressing Canon Cancer Detection Challenges

Canon cancer detection poses significant challenges due to the complexity and heterogeneity of tumors, as well as limitations in existing diagnostic modalities. Machine learning (ML) has emerged as a powerful tool to address these challenges by leveraging computational algorithms to analyze large datasets, extract meaningful patterns, and improve diagnostic accuracy. In this section, we explore the role of ML in overcoming the multifaceted challenges of canon cancer detection and its transformative impact on the field of oncology. One of the primary challenges in canon cancer detection is the subjective interpretation of imaging studies, which can lead to variability in diagnostic accuracy and consistency. ML techniques, particularly deep learning algorithms such as convolutional neural networks (CNNs), have shown promise in automating the analysis of medical images and extracting relevant features for tumor detection and classification. By training on large annotated datasets, CNNs can learn to identify subtle patterns indicative of cancerous lesions, leading to more consistent and reliable diagnoses.

Moreover, ML algorithms can help mitigate interobserver variability in histopathological analysis, another critical aspect of canon cancer diagnosis. By analyzing digital pathology images and histological slides, ML models can assist pathologists in identifying and characterizing cancerous cells with greater accuracy and efficiency. These models can learn from expert annotations and historical data to improve their performance over time, ultimately reducing discrepancies in diagnosis and enhancing diagnostic consistency. ML-driven diagnostic tools can also augment the sensitivity and specificity of existing diagnostic modalities for canon cancer detection. By integrating multiple data sources, including imaging studies, clinical data, and molecular profiles, ML algorithms can generate comprehensive patient profiles and identify subtle biomarkers indicative of cancer presence or progression. These models can facilitate earlier detection of tumors, leading to timely intervention and improved patient outcomes.

ML techniques are well-suited to address the inherent heterogeneity of canon cancer, which presents challenges in accurate detection and classification. By analyzing multi-omics data, including genomic, transcriptomic, and proteomic profiles, ML models can identify molecular signatures associated with different tumor subtypes and predict patient responses to specific treatments. This personalized approach to cancer diagnosis and treatment can help optimize therapeutic strategies and improve patient outcomes. In addition to enhancing diagnostic accuracy, ML-driven approaches can also address disparities in access to diagnostic services and expertise. By deploying automated diagnostic tools in low-resource settings and underserved communities, ML algorithms can expand access to high-quality diagnostic care and improve health outcomes for marginalized populations. These tools can help bridge the gap between regions with limited healthcare infrastructure and those with advanced medical facilities, democratizing access to life-saving cancer detection services.

Machine learning plays a pivotal role in addressing the multifaceted challenges of canon cancer detection by enhancing diagnostic accuracy, reducing interobserver variability, improving sensitivity and specificity, and addressing tumor heterogeneity. By leveraging computational algorithms to analyze complex datasets and extract meaningful insights, ML-driven approaches empower clinicians with powerful tools for early detection, personalized treatment planning, and improved patient outcomes. As the field of ML continues to evolve, its transformative impact on canon cancer detection is poised to revolutionize the landscape of oncology and contribute to significant advancements in cancer care.

2. REVIEW OF LITERATURE

One of the significant advances in the field is the development of interpretable ML systems for CRC diagnosis. For instance, Neto et al. (2024) introduced a system that combines multiple ML algorithms to analyze histopathological images, accurately identifying cancerous regions while ensuring the interpretability of decisions. This transparency is crucial for clinical acceptance, allowing healthcare professionals to trust and understand the basis of the AI's decisions. Azar et al. (2023) developed an automated system for colon cancer detection and segmentation using DL techniques. By integrating advanced image processing algorithms with deep learning models, their system could accurately segment cancerous regions from medical images, demonstrating the utility of DL in improving diagnostic efficiency. Bülbül, Burakgazi, and Kesimal (2024) utilized ML-based CTtexture analysis to predict the grade, T stage, and lymph node involvement in CRC patients.Their study highlights how texture analysis can extract discriminative features from CT images, enabling accurate classification and staging, which are pivotal in treatment planning and outcome improvement.

Predicting metastasis in CRC patients is another area where ML shows potential. Talebi et al. (2024) proposed classifiers that use clinical and demographic data to identify patients at high risk of developing metastatic disease. Their work underscores the role of ML in personalized medicine, helping clinicians in early intervention and potentially improving survival rates. Advanced ML and DL models also facilitate the analysis of complex datasets beyond images. Bostanci et al. (2023) focused on the analysis of RNA-seq data to predict diagnostic and prognostic outcomes in colon cancer. They utilized ML to identify specific gene expression patterns, offering insights into the molecular mechanisms of the disease and aiding in the discovery of biomarkers. Peng et al. (2023) introduced a machine learning-assisted method for detecting colon cancer biomarkers using label-free Surface- Enhanced Raman Scattering (SERS), showing how integrating ML can enhance thespecificity and reliability of early cancer detection technologies.

Despite these advancements, challenges such as data quality, interpretability, and generalizability of ML models persist. Alboaneen et al. (2023) and others have highlighted the need for high-quality, standardized data to train robust models. Additionally, there is a growing demand for models that clinicians can interpret easily to facilitate broader adoption in clinical settings. The integration of ML and DL in CRC management is not just improving diagnostic and predictive accuracy but also transforming clinical workflows by reducing the load on healthcare professionals and allowing more timely and tailored patient management. However, further research and validation in real-world settings are crucial to address existing hurdles and fully realize the potential of these technologies in clinical practice.

Jiang et al. (2022) proposed an integrated photoacoustic pen for breast cancer sentinel lymph node detection. Their study utilized photoacoustic imaging, a non-invasive imaging modality that combines optical and ultrasound techniques, to detect sentinel lymph nodes in breast cancer patients. By integrating ML algorithms, the authors achieved accurate and efficient detection of sentinel lymph nodes, offering a promising approach for improving breast cancer staging and treatment planning. Li et al. (2021) explored the application of convolutional neural networks (CNNs) and attention mechanisms for early gastric cancer detection. Their study combined deep learning techniques with attention mechanisms to analyze endoscopic images and identify early-stage gastric cancer lesions. The proposed model demonstrated superior performance compared to traditional image analysis methods, highlighting the potential of ML-driven approaches in improving gastric cancer diagnosis.

Nawreen et al. (2021) focused on lung cancer detection and classification using CT scan image processing. Leveraging image processing techniques and ML algorithms, the authors developed a system capable of accurately detecting and classifying lung cancer nodules from CT images. Their study underscores the importance of integrating computational methodologies with medical imaging for enhancing lung cancer diagnosis and prognosis. Soni and Singh (2020) proposed an automatic pulmonary cancer detection system using prewitt and morphological dilation techniques. Their study employed image processing algorithms to analyze chest X-ray images and detect pulmonary nodules indicative of lung cancer. By automating the detection process, the authors aimed to expedite diagnosis and improve patient outcomes in pulmonary cancer screening programs. Qin et al. (2020) investigated the use of image-based fractal analysis for cancer cell detection. Their study employed fractal analysis techniques to quantify the morphological complexity of cancer cells from microscopic images. By leveraging ML algorithms, the authors developed a robust framework for cancer cell detection, offering insights into the potential of fractal analysis in cancer research and diagnostics.

Lingappa and Parvathy (2022) proposed an active contour neural network for MRI-based bone cancer detection. Their study utilized deep learning techniques to analyze MRI images and identify bone cancer lesions with high accuracy. By integrating active contour modeling with neural networks, the authors achieved precise delineation of tumor boundaries, facilitating early detection and treatment planning in bone cancer patients. Yu et al. (2020) presented an improved Faster R-CNN model for colorectal cancer cell detection. Their study employed object detection algorithms to analyze histopathological images and identify colorectal cancer cells. By enhancing the efficiency and accuracy of cell detection, the proposed model offers a valuable tool for pathologists in colorectal cancer diagnosis and research. Zheng et al. (2021) developed a ResNet-based model for cancer detection, leveraging deep learning architectures to analyze medical imaging data. Their study demonstrated the efficacy of ResNet-based models in accurately detecting cancerous lesions from radiological images, highlighting the potential of deep learning in enhancing cancer diagnosis and treatment.

Shi, Pan, and Rehman (2022) proposed a cervical cancer cell image detection method based on an improved version of the YOLOv4 (You Only Look Once) algorithm. By enhancing the object detection capabilities of YOLOv4, their study aimed to accurately identify cervical cancer cells from microscopic images, offering a potential tool for early diagnosis and treatment monitoring. Santilli et al. (2021) investigated the application of self-supervised learning for the detection of breast cancer in surgical margins with limited data. Their study addressed the challenge of data scarcity by leveraging self-supervised learning techniques to train ML models using unlabeled data. By exploiting the intrinsic structure of unlabeled images, the authors achieved competitive performance in breast cancer detection, highlighting the potential of self-supervised learning in overcoming data limitations. Firdaus et al. (2020) developed a lung cancer detection system based on CT-scan images using gray-level co-occurrence matrix (GLCM) features and support vector machine (SVM) methods. Their study focused on extracting texture features from CT images and employing SVM classifiers to differentiate between cancerous and non-cancerous lung tissues. The proposed system demonstrated promising results in lung cancer detection, showcasing the effectiveness of texture-based image analysis in medical imaging.

Hemalatha, Chidambararaj, and Motupalli (2022) evaluated the performance of a deep learning approach for oral cancer detection and classification. By leveraging convolutional neural networks (CNNs) and deep learning techniques, their study aimed to accurately identify oral cancer lesions from medical images. The authors conducted comprehensive performance evaluations, demonstrating the potential of deep learning in enhancing oral cancer diagnosis and prognosis. Zhang et al. (2023) proposed a cascade detection network, named CCS-Net, for hypopharyngeal cancer detection in MRI images. By incorporating convolution kernel switch blocks and statistics optimal anchors blocks, their model achieved robust performance in detecting hypopharyngeal cancer lesions with high accuracy and efficiency. Poojahsri et al. (2023) reviewed various methods in texture analysis and classification techniques used in cervical cancer detection. Their study provided insights into the diverse approaches employed in cervical cancer diagnosis, including texture-based image analysis, machine learning algorithms, and classification techniques. Sajan et al. (2023) conducted a comprehensive review of silicon photonics biosensors for cancer cell detection. Their study highlighted the emerging field of silicon photonics and its potential applications in cancer diagnostics, offering sensitive and label-free detection methods for cancer biomarkers. Naraparaju et al. (2023) analyzed the accuracy of AI techniques for breast cancer detection. By comparing different ML algorithms and evaluation metrics, their study provided valuable insights into the performance of AI-driven approaches in breast cancer diagnosis, highlighting the importance of algorithm selection and validation in clinical settings.

Table 1. Comparative Overview of Recent Machine Learning and Deep Learning Approaches in Colorectal Cancer Diagnosis and Prognosis

Authors and Year	Focus of Study	Key Techniques Used	Main Contributions	Challenges Noted/Considerations
Neto et al. (2024)	CRC diagnosis from pathology slides	ML algorithms for image analysis	Improved accuracy and efficiency; emphasizes interpretability for clinician trust	Need for transparency and comprehensibility
Rai (2024)	General cancer detection and segmentation	ML and DL, feature extraction	Highlights the importance of feature extraction in various cancer types	Broad focus, not specific to CRC
Bülbül, Burakgazi,	Preoperative assessment of	ML-based texture analysis	Predicts grade, T stage, and	Data quality and generalizability issues

Kesimal (2024)	CRC using CT texture analysis		lymph node involvement, aiding preoperative planning	
Talebi et al. (2024)	CRC metastasis prediction	ML classifiers based on clinical data	Facilitates early identification and intervention in high-risk patients	N/A
Sharkas and Attallah (2024)	CRC classification	DL with triple CNNs and discrete cosine transform	High accuracy in distinguishing cancerous tissues, integrates spatial and spectral data	Feature fusion complexity
Karthikeyan, Jothilakshmi, Suthir (2024)	CRC detection	CNNs and ranking algorithm	Robust detection of CRC from medical images	Emphasis on feature selection and integration
Chhillar and Singh (2024)	Lung and colon cancer detection	ML algorithms and handcrafted features	Accurate classification using feature engineering, combines ML and DL techniques	Focus on feature selection and representation
Suominen, Subasi, Subasi (2024)	Detection of colon cancer	DL networks	Streamlines pathology workflow, high accuracy in detection	N/A
Wei et al. (2024)	Prediction of distant metastases in CRC	ML models using clinical data	Uses real-world data for robust model	Requires large-scale data for validation

			performance in prognosis	
Azar et al. (2023)	Colon cancer detection and segmentation	DL and image processing techniques	Improves segmentation accuracy in medical imaging	Model generalizability and interpretability issues
Bokhorst et al. (2023)	CRC detection and classification through semantic segmentation	Deep neural networks	Enhances precision in diagnosing CRC through multi-class semantic segmentation in digital pathology images	Potential reduction in pathologist workload
Bostanci et al. (2023)	Analysis of RNA-seq data for CRC	ML on gene expression data	Identifies biomarkers and gene patterns for diagnosis and prognosis prediction	Requires understanding of molecular mechanisms
Peng et al. (2023)	Early detection of colon cancer	ML with SERS technology	Develops non-invasive, highly sensitive detection method	Reliability and specificity in clinical application
Al-Rajab et al. (2023)	Gene classification in colon cancer datasets	Hybrid ML feature selection model	Enhances accuracy and efficiency in gene classification, reducing dimensionality	Handling high-dimensional data complexity
Muniz et al. (2023)	Histopathological diagnosis of colon cancer	DL with micro-FTIR hyperspectral imaging	Captures biochemical changes, facilitating	Integration of hyperspectral data with DL

			precise discrimination of tissues	

3. CONCLUSION AND FUTURE SCOPE

The literature review presented herein underscores the significant strides made in the field of machine learning (ML) for cancer detection across various types of malignancies. From breast cancer and cervical cancer to lung cancer and oral cancer, researchers have leveraged innovative ML techniques to improve detection accuracy, efficiency, and accessibility. By integrating computational algorithms with advanced imaging modalities and molecular assays, ML-driven approaches offer promising avenues for early diagnosis, personalized treatment planning, and improved patient outcomes. Recent studies have demonstrated the versatility of ML techniques, including convolutional neural networks (CNNs), support vector machines (SVMs), and deep learning architectures, in analyzing medical images, genomic data, and histopathological specimens for cancer detection. Additionally, advancements in self-supervised learning, object detection algorithms, and texture analysis methods have further expanded the capabilities of ML in cancer diagnostics.

Moreover, the integration of ML with emerging technologies such as silicon photonics biosensors and photoacoustic imaging holds immense potential for enhancing cancer detection sensitivity, specificity, and scalability. By harnessing the power of interdisciplinary collaborations and leveraging big data analytics, researchers can unlock new insights into cancer biology, tumor heterogeneity, and treatment responses, paving the way for precision oncology and targeted therapies.

Future Scope

Looking ahead, several avenues for future research and development in ML-driven cancer detection emerge:

1. **Integration of Multi-omics Data:** Incorporating multi-omics data, includinggenomics, transcriptomics, proteomics, and metabolomics, can provide a comprehensive view of cancer biology and facilitate more accurate diagnosis and treatment selection.

2. **Explainable AI and Clinical Validation:** Enhancing the interpretability and explainability of ML models is crucial for their clinical adoption. Future research should focus on developing transparent and interpretable AI-driven diagnostic systems that align with clinical workflows and regulatory standards.
3. **Addressing Data Imbalance and Bias:** Mitigating data imbalance and bias in training datasets is essential for ensuring the generalizability and fairness of ML models across diverse patient populations and demographic groups.
4. **Real-time Monitoring and Prediction:** Leveraging ML techniques for real-time monitoring of cancer progression, treatment response, and recurrence prediction can empower clinicians with actionable insights for personalized patient management.
5. **Translation to Clinical Practice:** Bridging the gap between research and clinical practice requires robust validation studies, regulatory approvals, and integration of ML-driven diagnostic tools into existing healthcare infrastructure.

REFERENCES

1. Neto, P. C., Montezuma, D., Oliveira, S. P., Oliveira, D., Fraga, J., Monteiro, A., ... & Cardoso, J. S. (2024). An interpretable machine learning system for colorectal cancer diagnosis from pathology slides. *NPJ precision oncology*, *8*(1), 56.
2. Rai, H. M. (2024). Cancer detection and segmentation using machine learning and deep learning techniques: A review. *Multimedia Tools and Applications*, *83*(9), 27001-27035.
3. Bülbül, H. M., Burakgazi, G., & Kesimal, U. (2024). Preoperative assessment of grade, T stage, and lymph node involvement: machine learning-based CT texture analysis in colon cancer. *Japanese Journal of Radiology*, *42*(3), 300-307.
4. Bülbül, H. M., Burakgazi, G., & Kesimal, U. (2024). Preoperative assessment of grade, T stage, and lymph node involvement: machine learning-based CT texture analysis in colon cancer. *Japanese Journal of Radiology*, *42*(3), 300-307.
5. Talebi, R., Celis-Morales, C. A., Akbari, A., Talebi, A., Borumandnia, N., & Pourhoseingholi, M. A. (2024). Machine learning-based classifiers to predict metastasis in colorectal cancer patients. *Frontiers in Artificial Intelligence*, *7*.
6. Sharkas, M., & Attallah, O. (2024). Color-CADx: a deep learning approach for colorectal cancer classification through triple convolutional neural networks and discrete cosine transform. *Scientific Reports*, *14*(1), 6914.
7. Karthikeyan, A., Jothilakshmi, S., & Suthir, S. (2024). Colorectal cancer detection based on convolutional neural networks (CNN) and ranking algorithm. *Measurement: Sensors*, *31*, 100976.
8. Chhillar, I., & Singh, A. (2024). A feature engineering-based machine learning technique to detect and classify lung and colon cancer from histopathological images. *Medical & Biological Engineering & Computing*, *62*(3), 913-924.

9. Suominen, M., Subasi, M. E., & Subasi, A. (2024). Automated detection of colon cancer from histopathological images using deep neural networks. In *Applications of Artificial Intelligence Healthcare and Biomedicine* (pp. 243-287). Academic Press.
10. Wei, R., Yu, G., Wang, X., Jiang, Z., & Guan, X. (2024). Construction and validation of machine learning models for predicting distant metastases in newly diagnosed colorectal cancer patients: A large-scale and real-world cohort study. *Cancer Medicine*, *13*(5), e6971.
11. Alboaneen, D., Alqarni, R., Alqahtani, S., Alrashidi, M., Alhuda, R., Alyahyan, E., & Alshammari, T. (2023). Predicting colorectal cancer using machine and deep learning algorithms: Challenges and opportunities. *Big Data and Cognitive Computing*, *7*(2), 74.
12. Azar, A. T., Tounsi, M., Fati, S. M., Javed, Y., Amin, S. U., Khan, Z. I., ... & Ganesan, J. (2023). Automated system for colon cancer detection and segmentation based on deep learning techniques. *International Journal of Sociotechnology and Knowledge Development (IJSKD)*, *15*(1), 1-28.
13. Bokhorst, J. M., Nagtegaal, I. D., Fraggetta, F., Vatrano, S., Mesker, W., Vieth, M., ... & Ciompi, F. (2023). Deep learning for multi-class semantic segmentation enables colorectal cancer detection and classification in digital pathology images. *Scientific Reports*, *13*(1), 8398.
14. Bostanci, E., Kocak, E., Unal, M., Guzel, M. S., Acici, K., & Asuroglu, T. (2023). Machine learning analysis of RNA-seq data for diagnostic and prognostic prediction of colon cancer. *Sensors*, *23*(6), 3080.
15. Peng, S., Lu, D., Zhang, B., You, R., Chen, J., Xu, H., & Lu, Y. (2023). Machine learning–assisted internal standard calibration label-free SERS strategy for colon cancerdetection. *Analytical and Bioanalytical Chemistry*, *415*(9), 1699-1707.
16. Al-Rajab, M., Lu, J., Xu, Q., Kentour, M., Sawsa, A., Shuweikeh, E., ... & Arasaradnam, R. (2023). A hybrid machine learning feature selection model—HMLFSM to enhance gene classification applied to multiple colon cancers dataset. *Plos one*, *18*(11), e0286791.
17. Mokoatle, M., Marivate, V., Mapiye, D., Bornman, R., & Hayes, V. M. (2023). A review and comparative study of cancer detection using machine learning: SBERT and SimCSE application. *BMC bioinformatics*, *24*(1), 112.
18. Mengash, H. A., Alamgeer, M., Maashi, M., Othman, M., Hamza, M. A., Ibrahim, S. S., ... & Yaseen, I. (2023). Leveraging marine predators algorithm with deep learning for lung and colon cancer diagnosis. *Cancers*, *15*(5), 1591.
19. Wahid, R. R., Nisa, C., Amaliyah, R. P., & Puspaningrum, E. Y. (2023, February). Lung and colon cancer detection with convolutional neural networks on histopathological images. In *AIP Conference Proceedings* (Vol. 2654, No. 1). AIP Publishing.
20. Muniz, F. B., Baffa, M. D. F. O., Garcia, S. B., Bachmann, L., & Felipe, J. C. (2023). Histopathological diagnosis of colon cancer using micro-FTIR hyperspectral imaging and deep learning. *Computer Methods and Programs in Biomedicine*, *231*, 107388.
21. Jiang, D., et al. (2022). Integrated Photoacoustic Pen for Breast Cancer Sentinel Lymph Node Detection. IEEE International Ultrasonics Symposium (IUS), 1-3. https://doi.org/10.1109/IUS54386.2022.9958597
22. Li, X., Chai, Y., Zhang, K., Chen, W., & Huang, P. (2021). Early gastric cancer detection based on the combination of convolutional neural network and attention mechanism. 2021

China Automation Congress (CAC), 5731-5735. https://doi.org/10.1109/CAC53003.2021.9728413

23. Nawreen, N., Hany, U., & Islam, T. (2021). Lung Cancer Detection and Classification using CT Scan Image Processing. 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), 1-6. https://doi.org/10.1109/ACMI53878.2021.9528297
24. Soni, A., & Singh, A. P. (2020). Automatic Pulmonary Cancer Detection using Prewitt & Morphological Dilation. 2nd International Conference on Data, Engineering and Applications (IDEA), 1-6. https://doi.org/10.1109/IDEA49133.2020.9170680
25. Qin, J., Puckett, L., & Qian, X. (2020). Image Based Fractal Analysis for Detection of Cancer Cells. IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 1482-1486. https://doi.org/10.1109/BIBM49941.2020.9313176
26. Lingappa, E., & Parvathy, L. R. (2022). Active Contour Neural Network Identifying MRI Image Edge Computing Methods Deep Learning Bone Cancer Detection. 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 830-834. https://doi.org/10.1109/ICACITE53722.2022.9823617
27. Yu, Y., Yang, Q., & Shaohan, L. (2020). An Improved Faster R-CNN for Colorectal Cancer Cell Detection. IEEE 3rd International Conference on Electronics and Communication Engineering (ICECE), 186-190. https://doi.org/10.1109/ICECE51594.2020.9353044
28. Zheng, Z., Zhang, H., Li, X., Liu, S., & Teng, Y. (2021). ResNet-Based Model for Cancer Detection. IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), 325-328. https://doi.org/10.1109/ICCECE51280.2021.9342346
29. Shi, C., Pan, Q., & Rehman, M. (2022). Cervical Cancer Cell Image Detection Method Based on Improved YOLOv4. 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), 1996-2000. https://doi.org/10.1109/ICSP54964.2022.9778577
30. Santilli, A. M. L., et al. (2021). Self-Supervised Learning For Detection Of Breast Cancer In Surgical Margins With Limited Data. IEEE 18th International Symposium on Biomedical Imaging (ISBI), 980-984. https://doi.org/10.1109/ISBI48211.2021.9433829
31. Firdaus, Q., Sigit, R., Harsono, T., & Anwar, A. (2020). Lung Cancer Detection Based On CT-Scan Images With Detection Features Using Gray Level Co-Occurrence Matrix (GLCM) and Support Vector Machine (SVM) Methods. 2020 International Electronics Symposium (IES), 643-648. https://doi.org/10.1109/IES50839.2020.9231663
32. Hemalatha, S., Chidambararaj, N., & Motupalli, R. (2022). Performance Evaluation of Oral Cancer Detection and Classification using Deep Learning Approach. 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), 1-6. https://doi.org/10.1109/ACCAI53970.2022.9752505
33. Zhang, S., et al. (2023). CCS-Net: Cascade Detection Network With the Convolution Kernel Switch Block and Statistics Optimal Anchors Block in Hypopharyngeal Cancer MRI. IEEE Journal of Biomedical and Health Informatics, 27*(1), 433-444. https://doi.org/10.1109/JBHI.2022.3217174
34. Poojahsri, S., Prince, G. K., Prerana, S., & Krishnakumar, S. (2023). Various Methods in Texture Analysis and Classification Techniques used in Cervical Cancer Detection. 2023

5th International Conference on Smart Systems and Inventive Technology (ICSSIT), 1649-1652. https://doi.org/10.1109/ICSSIT55814.2023.10060990
35. Sajan, S. C., Singh, A., Sharma, P. K., & Kumar, S. (2023). Silicon Photonics Biosensors for Cancer Cells Detection—A Review. IEEE Sensors Journal, 23*(4), 3366-3377. https://doi.org/10.1109/JSEN.2023.3235920
36. Naraparaju, S. L., Eerlapalli, N. S. N., Modem, N., Paidipati, R., Karthikeyan, C., & Sathish Kumar, K. (2023). Analyzing the Accuracy of AI Techniques for Breast Cancer Detection. 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), 1660-1664. https://doi.org/10.1109/ICSSIT55814.2023.10061075

PROTEIN METABOLISM AND EXERCISE: EVALUATING THE NECESSITY OF AMINO ACID SUPPLEMENTS

SCHOLAR DETAILS:
Mr. FEBY F.A.S
RESEARCH SCHOLAR (Ph.D IN BIO-CHEMISTRY)
DESH BHAGAT UNIVERSITY
MANDI GOBINDAGARH, PUNJAB

GUIDE
Dr. VIVEK SINGH DAHIYA
PROF
DEPARTMENT OF BIO-CHEMISTRY
DESH BHAGAT UNIVERSITY
MANDI GOBINDAGARH, PUNJAB

Abstract

The intricate process of protein metabolism is vital to overall health, muscular growth, and repair—especially during physical activity. This study explores the complex interplay between protein metabolism and exercise, emphasizing the roles of dietary supplements, MPS (muscle protein synthesis), and resistance training. The findings demonstrate that resistance exercise significantly raises MPS in many muscle areas, and that deliberate consumption of protein supplements and amino acids amplifies the benefits. According to the study, high-intensity exercise increases protein turnover, with sarcoplasmic and myofibrillar proteins exhibiting different responses. For endurance athletes, leucine and other necessary amino acids should be consumed in a tailored manner to optimize muscle growth and repair. Athletes usually maintain positive nitrogen balances, a sign that they are consuming enough protein to maintain and adapt their muscles. This study highlights how important it is to balance protein intake with the kind and level of exercise to maximize anabolic reactions and support overall muscle health. This study emphasizes the value of customized dietary plans and calls for more research to examine the complex interplay between supplements, activity, and protein metabolism.

Keywords: *Protein, Metabolism, Exercise, Amino Acid, Supplements*

1. INTRODUCTION

Protein synthesis and breakdown occur within the body as part of protein metabolism, a basic biological activity. Amino acid-based proteins are necessary for many physiological processes, such as immune system response, muscle repair, enzyme functioning, and hormone synthesis. The body needs protein during exercise, particularly resistance and endurance training, because it needs it for muscle growth and repair. Important considerations concerning the sufficiency of dietary protein consumption and the possible function of amino acid supplements in promoting the best possible exercise performance and recuperation are brought up by this increased demand.

Athletes and fitness enthusiasts who want to improve their workout results are increasingly turning to amino acid supplements. Supplements that claim to increase muscle protein synthesis, decrease pain in the muscles, and speed up recovery include branched-chain amino acids (BCAAs), whey protein, and essential amino acids. The usefulness and necessity of these supplements are still up for discussion, though. A well-balanced diet that includes a variety of foods high in protein, such as meat, dairy, legumes, and nuts, can help many people satisfy their protein demands.

This study examines the available data on protein metabolism during exercise to determine whether amino acid supplements are necessary. It will examine how various forms of exercise impact the amount of protein required, how dietary protein helps meet these needs, and the possible advantages and disadvantages of supplementing. This research aims to give evidence-based suggestions for people who want to enhance their nutritional strategies for performance and recovery by understanding the relationship between exercise and protein metabolism. The results of this study will add to the current conversation about individualized diet plans and the benefits of supplements for physical activity and overall health.

2. REVIEW OF LITREATURE

Torre-Villalvazo et al. (2019) Examine the effects of protein and amino acid supplements on immunity as well as how they affect workout recovery and performance. The study highlights how taking supplements and eating the correct quantity of protein might alter these signaling pathways, accelerating the synthesis of muscle protein and enhancing post-exercise recovery. The study highlights how important it is to tailor protein and amino acid intake to each specific exercise regimen in order to optimize metabolic processes and improve athletic performance.

Master and Macedo (2021), who focus specifically on the amino acids and proteins found in milk. The authors look at the advantages of these supplements in terms of boosting total exercise performance, reducing muscle injury, and improving muscle protein synthesis. The analysis emphasizes the significance of casein and whey proteins, highlighting how their unique amino acid compositions and digestibility contribute to the reason behind their superior supplement performance. The benefits of combining resistance training with milk proteins are also highlighted in the study, with a focus on how athletes and other physically active people can boost their strength and muscle mass.

Li et al. (2024) examines the mechanisms via which amino acids regulate skeletal muscle metabolism and provides guidance on appropriate exercise dosage and potential negative effects. The study examines the different ways that amino acids aid in the synthesis of muscle protein, energy production, and recuperation. It also examines their interactions with mTOR and AMPK signaling pathways. The authors provide evidence-based suggestions for amino acid supplements that account for exercise duration, intensity, and individual variability. By addressing the negative effects of increased amino acid intake, such as gastrointestinal distress and metabolic irregularities, the research also emphasizes the need for customized supplementing solutions.

Williamson et al. (2019) Examine how dietary protein consumption affects protein metabolism and endurance-trained male performance. According to the study, eating more protein can dramatically accelerate the synthesis of new proteins and decrease the breakdown of existing muscle protein, which improves recovery and endurance. The authors stress that sustaining muscle mass and encouraging the metabolic changes required for endurance training depend on consuming the right amount of protein. This study emphasizes how crucial it is for endurance athletes to consume enough protein to meet their high energy requirements and recuperation needs.

3. HUMAN PROTEIN METABOLISM: A BROAD PERSPECTIVE

Human body protein mass provides structural support, enzymes to catalyze metabolic reactions, signalling intermediates within and between cell tissues, and fuel for extreme survival. Major protein deposits are found in skeletal muscles, accounting for 40% of body weight in young males with 20-22% BMI ($kg.m^{-2}$) and about 60% of total body protein in humans. Plasma proteins (including albumin, about 50% of liver proteins), immune cells (primarily leucocytes), intestinal tract proteins (digestive enzymes), bone, and dermal collagen are also produced by the liver. Protein balance shows the net protein synthesis and breakdown of any cell or tissue, which varies greatly between tissues, organs, and cell compartments.

Table 1: Averages of human skeletal muscle's fractional synthesis rates (FSR) for particular proteins (fasted state).

Muscle Fraction	**FSR, Mean ± SD (% per day)**
Myosin Heavy Chain	0.92 ± 0.09
Actin	1.803± 0.20
Sarcoplasm	1.31 ± 0.31
Mitochondria	1.95 ± 0.11

Table 1 shows the average fractional synthesis rates (FSR) of several proteins in fasted human skeletal muscle, revealing protein turnover. Myosin heavy chain, a muscular contraction protein, with an FSR of 0.92% per day, showing a lower synthesis rate than other muscle fractions. Actin, another key contractile protein, has a greater FSR of 1.803% each day, suggesting its dynamic role in muscle construction and function. Sarcoplasm, which contains metabolic proteins and enzymes, has a daily FSR of 1.31%. The mitochondria, which produce energy and metabolism, have the greatest FSR at 1.95% each day, indicating active protein turnover to meet muscle cell energy needs. These data indicate that all muscle fractions synthesize protein, but mitochondria turnover protein the fastest, possibly due to their vital function in energy metabolism.

4. MUSCLE PROTEIN CONTENT AND THE CONSEQUENCES OF EXERCISE

Several reviews discuss human muscle protein production and breakdown during and after resistance exercise. We must distinguish fasted (post-absorptive) or fed findings during versus after exercise for muscle protein synthesis (MPS) or breakdown (MPB).

Table 2 shows that short-term intensive resistance training enhance mixed skeletal muscle protein at rest. Methodological discrepancies may exist between reports, although the total work production (80–90% of maximum contraction) appears to provide the most benefit. Resistance exercise works better on myofibrillar than sarcoplasmic proteins. Muscle protein synthesis increases appear to last up to 4 hours following exercise. Resistance training appears to have no effect on muscle protein breakdown.

Table 194: Resistance exercise's effects on MPS and MPB in untrained and fasting

Exercise Protocol	**FSR (%.h⁻¹) Post Ex/Pre-Ex-Ratio**	**FBR (%.h⁻¹) Post Ex/Pre-Ex-Ratio**
Mixed Muscle Proteins		
5 x 7-13 rep. 80% max	+50%*	-
6 x 11 rep. 100% max	+141%*	-
9 x 8 rep. 80% max	+151%*	+40%*
9 x 130% max	+133%*	+49%*
7 x 9 rep. 80% max	+32%*	-
9 x 11 rep. 75% max	+38%*	NS
11 x 13 rep. 80% max	+52%*	-
5 x 12 rep. 80% max	+139%*	-
6 x 92% max	+355%*	-
Myofibrillar Proteins		
6 x 92% max	+325%*	-
7 sec at 30% max to exhaus.	+181%*	-
Sarcoplasmic Proteins		
6 x 92% max	+79%*	-
Mitochondrial Proteins		
7 sec at 30% max to exhaus.	+181%*	-

Table 195: Resistance exercise's effects on MPS and MPB in untrained and fasting

Exercise Protocol	**Nutritional Protocol**	**FSR (%.h⁻¹) Post Ex/Prix Ratio**	**FBR (%.h⁻¹) Post Ex/Pre-Ex-Ratio**
	Mixed Muscle Proteins		
6 x 11 rep. max	11g AA (IV)	+133%*	NS
12 x 10 rep. 80% max	7g EAA (oral)	+425%*	NS
6 x 13 rep. 80% max	11g whey + CHO (oral)	+144%*	-
11 x 20 rep. 70% max	Leu EAA + CHO (oral)	+171%*	-
5 x 11 rep. max	42g egg proteins (oral)	+92%*	-
9 x 12 rep. 70% max	11g whey (oral)	+41%*	-
Myofibrillar Proteins			
6 x 12 rep. 80% max	1g protein.kg⁻¹ (oral)	+85%*	-
22 x 18 rep. 75% max	6g protein’s⁻¹ (oral)	+191%*	NS
stepping ex (+26% bw)	45g EAA + CHO	+228%*	-
6 x 10 rep. max	25g whey (oral)	+231%*	-
9 x 12 rep. max	25g whey (oral)	+199%*	-
11 x 9 rep. 80% max	0.3g.kg⁻¹ LM whey	+95%*	-
5 x 12 rep. 80% max	20g whey protein (oral)	+51%*	-
Sarcoplasmic Proteins			
22 x 12 rep. 75% max	6g protein.h⁻¹ (oral)	+322%*	-

6 x 12 rep. max	25g whey (oral)	+105%*	-

Tables 2 and 3 shed light on how, in untrained and fasting people, resistance exercise and dietary interventions impact muscle protein synthesis (MPS) and muscle protein breakdown (MPB). The effects of resistance training alone on various muscle fractions are the main topic of Table 2. It demonstrates that resistance training dramatically raises the fractional synthesis rates (FSR) of mitochondrial, sarcoplasmic, mixed muscle, and myofibrillar proteins. Exercises such as 6 × 11 reps at 100% max for mixed muscle proteins raise FSR by 141%, highlighting the significant impact of high-intensity resistance training on MPS. Exercises at 130% max result in a 133% increase in FSR and a 49% increase in MPB, suggesting that although there is a significant increase in protein synthesis, there is also an increase in protein breakdown, perhaps as a result of muscle injury and remodeling processes. Exercises at 92% max result in a 325% increase in FSR for myofibrillar proteins, demonstrating the significant influence of near-maximal loads on contractile protein synthesis. On the other hand, FSR rises by 181% after intense activity at 30% max, indicating that myofibrillar protein synthesis can be greatly increased even by lower-intensity activities that are completed to exhaustion. Resistance training promotes protein synthesis across different muscle fractions, with the response varied according to exercise intensity and duration. Sarcoplasmic and mitochondrial proteins also respond favorably, with FSR rising by 79% and 181%, respectively.

The combined effects of resistance training and dietary treatments are examined in Table 3. Here, the anabolic effect of exercise is enhanced by the addition of amino acids and protein supplements. Intravenous injection of 11g of amino acids during 6 x 11 reps at maximum intensity causes a 133% rise in FSR for mixed muscle proteins, but does not significantly alter MPB, indicating greater protein synthesis without increased breakdown. A 425% increase in FSR is obtained by oral ingestion of 7g of essential amino acids during 12 x 10 repetitions at 80% max, indicating the synergistic effect of amino acid supplementation with exercise. When exercise and whey protein supplementation are coupled, myofibrillar protein synthesis is also greatly increased. For example, a 231% increase in FSR after 6 x 10 repetitions at maximum intensity with 25g of whey protein suggests that supplementation can maximize the anabolic response of these proteins. When 6g of protein is consumed hourly over a 22 x 12 rep program at 75% max, sarcoplasmic proteins show a 322% increase in FSR. This highlights the significance of continuous protein intake to maintain increased synthesis rates in sarcoplasmic fractions necessary for metabolic adaptations. Overall, these results show that MPS is greatly increased by resistance exercise in different muscle fractions, and that these benefits are further enhanced by nutritional supplements. The incorporation of particular protein sources and amino acids into post-exercise nutrition maximizes the anabolic response, indicating a promising approach to enhance muscle repair and adaptation in persons who are not trained.

5. SIMPLE DIETARY GUIDELINES FOR CONSISTENT PHYSICAL ACTIVITY

The information provided here encourages athletes and those who exercise regularly to pay attention to sufficient protein consumption in order to maintain or improve their status of skeletal muscle mass. The scientific literature, however, demonstrates a broad range of useful behaviors that support the adaptation of muscle growth through particular dietary applications: how much, when, how, and what kind of protein? We'll make an effort to distinguish the wheat from the chaff.

The World Health Organization (WHO), the USA Institute of Medicine, and the health organizations in France and Belgium have set exact guidelines for the recommended daily allowance (RDA) of 0.83 $g.kg^{-1}$ body weight for young, sedentary adults. Based on the statistical distribution of adult participants, a protein intake alimentary deficit occurs when body weight is less than 0.40–0.50 g/kg.

However, the daily load of ±0.8 $g.kg^{-1}$ body weight seems inadequate for persons engaging in medium- or high-intensity physical activities on a regular basis (sports, leisure, jobs). There is evidence in many publications to support a somewhat regular growth above the "RDA" amount.

There is growing evidence that the amount of muscular growth during recovery is influenced by the time and source of protein consumed. There seems to be a slight differential in the turnover of muscle proteins between young men and women. A balance between the amount of protein consumed daily and the amount used during activity is what's known as adequate protein balance. As previously mentioned, this balance could be estimated using the amount of nitrogen that is taken in through food and released as nitrogen wastes, primarily in urine. Although the nitrogen balance (NBal) has been used for a long time, this approach is not the most precise one. Urine collection is still an indirect way to assess the daily balance between nitrogen released from protein degradation (mostly muscle mass) and protein intake as determined by the daily meal questionnaire. An example of NBal on young athletes is shown in Table 4 (Poortmans unpublished data).

Table 4: N balance (NBal) of athletes was measured twice a day for seven days using urine nitrogen analysis and a meal questionnaire.

Sport	**Athletes (Number)**	**Gender**	**Age (Years)**	**NBal (Mean ± SE)***	**NBal > Than 0^{**} (g·protein^{-1}·24h^{-1})**
Running	20	M	20-21	1.47 ± 0.08	1.23
Rowing	20	M	16-21	1.30 ± 0.08	1.18
Cycling	20	M	18-22	1.60 ± 0.10	1.39
Swimming	10	M, F	12-19	1.61 ± 0.15	1.52
Gymnastics	10	F	9-13	1.71 ± 0.41	1.40
Gymnastics	15	F	16-17	1.13 ± 0.20	0.87
Basketball	15	M	20-40	1.76 ± 0.15	1.19
Aerobics	10	F	20-34	1.25 ± 0.07	1.18
Orienteering	20	M	23-35	1.39 ± 0.14	1.32
Bodybuilding	10	M	26-37	1.95 ± 0.14	1.35

The nitrogen balance (NBal) of athletes participating in different sports is shown in Table 4, which was calculated over a seven-day period using meal questionnaires and urine nitrogen tests. An essential marker of protein metabolism, nitrogen balance shows the distinction between nitrogen intake and excretion. While a negative nitrogen balance denotes a possible protein shortfall or increased protein requirements, a positive nitrogen balance supports enough protein consumption for growth, recovery, and maintenance of lean body mass.

The table demonstrates that athletes participating in sports like gymnastics, basketball, and bodybuilding have some of the highest mean nitrogen balance values. For example, bodybuilders' NBal of 1.95 ± 0.14 $g \cdot protein^{-1} \cdot 24h^{-1}$ suggests that their high-protein diets complement their rigorous training schedules, which frequently target muscular hypertrophy. A high nitrogen balance (1.76 ± 0.15) is also observed in basketball players, indicating that they consume enough protein to match the demands of their sport, which calls for both power and endurance.

Athletes in strength and power sports have higher nitrogen balances than those in endurance sports, such as rowing and running. The NBal for rowers is 1.30 ± 0.08, and for runners it is 1.47 ± 0.08. Despite being lower, these readings nevertheless show a positive nitrogen balance, indicating that the athletes are getting the protein they need to sustain both recovery and endurance performance. The marginally reduced equilibrium could be attributed to the distinct metabolic requirements of endurance sports, which entail elevated energy expenditure and the utilization of protein for both energy and repair.

The data indicates variations in nitrogen balance between genders, especially in sports such as gymnastics, where female athletes between the ages of 9 and 13 have a mean NBal of 1.71 ± 0.41, which is higher than that of their older counterparts between the ages of 16 and 17, who have an NBal of 1.13 ± 0.20. The observed discrepancy could perhaps be attributed to variances in growth rates, hormonal fluctuations, and training load among junior athletes. This underscores the significance of customized nutritional approaches to bolster female athletes over a range of developmental phases.

5.1 Endurance type of training

examined three different protein intakes for endurance athletes (light: 0.8 g.kg.day; medium: 1.8 g.kg.day; high: 3.6 g.kg.day). There was no correlation between protein synthesis and food intake following exercise. However, when examining young, unskilled individuals under fast conditions who practiced for 60 minutes at 30% Wmax or 30 minutes at 60% Wmax on a bicycle, marginally modified this conclusion. In the post-exercise phase, they saw a steady mitochondrial protein fraction in both cases along with a 60% increase in myofibrillar protein in the vastus lateralis muscle. Furthermore, these scientists found that the two muscle compartments' protein synthesis continued for 24 to 48 hours after exercise.

A few studies highlighted the effects of varying protein consumption amongst endurance athletes who had received training (Table 5).

Table 5: Muscle protein synthesis in experienced endurance athletes following varying protein intake ratios during 50–75% VO2max endurance training.

Types of Food	**Quantities**	**Protein Synthesis**
Pro	2.6 g·kg⁻¹	Stable
	0.25 g·kg⁻¹	Stable
Whole Milk	5 ml·kg⁻¹	Stable
	0.095 g·kg⁻¹	+101%
EAA	15.5 g	+15%
Leu	51 mg·kg⁻¹	+14%
	7 g	Stable

Muscle protein synthesis (MPS) in seasoned endurance athletes after consuming different protein sources during training at 50–75% of VO2max is shown in Table 5. The table lists the various food categories, their corresponding amounts, and how they affect the synthesis of proteins.

Pro: Consuming 2.6 $g \cdot kg^{-1}$ and 0.25 $g \cdot kg^{-1}$ of protein each day had a consistent impact on MPS. This implies that although protein consumption is necessary to preserve muscle mass and function, these specific ratios did not, in endurance athletes under the investigatedconditions, significantly increase protein synthesis. This stability suggests that basal protein intake levels for seasoned endurance athletes may be adequate to meet the needs of muscle maintenance and repair without the need to increase synthesis rates.

Whole Milk: While protein synthesis was likewise steady when 5 $ml \cdot kg^{-1}$ of whole milk was consumed, there was a notable increase in protein synthesis (+101%) when 0.095 $g \cdot kg^{-1}$ was consumed. This discovery raises the possibility that some whole milk constituents, like casein and whey proteins, may have strong anabolic effects at particular dosages. This is probably because of their balanced amino acid profiles and sluggish digestion rates, which prolong the supply of amino acids needed for muscle growth and repair.

Essential Amino Acids (EAA): 15% more protein was synthesized after consuming 15.5 g of EAAs. Because they include the necessary building blocks that the body is unable to manufacture on its own, EAAs are needed for the synthesis of muscle protein. This increase highlights the significance of consuming enough EAAs to maintain muscle protein synthesis and repair, especially during endurance training when muscle deterioration may occur.

Leucine (Leu): A 14% increase in protein synthesis was observed upon consuming 51 $mg \cdot kg^{-1}$ of leucine, a branched-chain amino acid (BCAA). One important regulator of muscle protein synthesis, the mTOR pathway, is known to be strongly stimulated by leucine. This result emphasizes the significance of leucine as a necessary dietary amino acid for fostering muscle anabolism, especially in endurance athletes who might require extra assistance for muscle growth and repair. Higher intakes of 7 g, however, did not result in any more increases, indicating that there may be a threshold beyond which leucine supplementation does not further improve MPS.

6. CONCLUSION

Muscle protein synthesis (MPS), resistance training, and dietary supplements are closely related to one another, as the research on protein metabolism and exercise highlights. It demonstrates that resistance training raises MPS in all muscle fractions, but that the effects are enhanced when protein supplements and amino acids are strategically consumed. Myofibrillar and sarcoplasmic proteins, for example, show distinct responses to high-intensity training in terms of increased protein turnover. Additionally, a customized protein intake is beneficial for endurance athletes, as certain nutrients such as leucine and vital amino acids enhance muscle growth and repair. It is noteworthy that athletes have positive nitrogen balances, indicating that they consume enough protein to sustain and grow their muscles. In order to optimize anabolic reactions and maintain muscle health, this research emphasizes the significance of matching protein intake with exercise intensity and type.

REFERENCES

1. Church, D. D., Hirsch, K. R., Park, S., Kim, I. Y., Gwin, J. A., Pasiakos, S. M., ... & Ferrando, A. A. (2020). Essential amino acids and protein synthesis: insights into maximizing the muscle and whole-body response to feeding. Nutrients, 12(12), 3717.
2. Gwin, J. A., Church, D. D., Hatch-McChesney, A., Allen, J. T., Wilson, M. A., Varanoske, A. N., ... & Pasiakos, S. M. (2021). Essential amino acid-enriched whey enhances post-exercise whole-body protein balance during energy deficit more than iso-nitrogenous whey or a mixed-macronutrient meal: a randomized, crossover study. Journal of the International Society of Sports Nutrition, 18(1), 4.
3. Gwin, J. A., Church, D. D., Wolfe, R. R., Ferrando, A. A., & Pasiakos, S. M. (2020). Muscle protein synthesis and whole-body protein turnover responses to ingesting essentialamino acids, intact protein, and protein-containing mixed meals with considerations for energy deficit. Nutrients, 12(8), 2457.
4. Holeček, M. (2018). Branched-chain amino acids in health and disease: metabolism, alterations in blood plasma, and as supplements. Nutrition & metabolism, 15, 1-12.
5. Holeček, M. (2022). Side effects of amino acid supplements. Physiological research, 71(1), 29.
6. Huecker, M., Sarav, M., Pearlman, M., & Laster, J. (2019). Protein supplementation in sport: source, timing, and intended benefits. Current nutrition reports, 8, 382-396.
7. Kårlund, A., Gómez-Gallego, C., Turpeinen, A. M., Palo-Oja, O. M., El-Nezami, H., & Kolehmainen, M. (2019). Protein supplements and their relation with nutrition, microbiota composition and health: is more protein always better for sportspeople?. Nutrients, 11(4), 829.
8. Lam, F. C., Khan, T. M., Faidah, H., Haseeb, A., & Khan, A. H. (2019). Effectiveness of whey protein supplements on the serum levels of amino acid, creatinine kinase and myoglobin of athletes: a systematic review and meta-analysis. Systematic reviews, 8, 1-12.
9. Li, G., Li, Z., & Liu, J. (2024). Amino acids regulating skeletal muscle metabolism: mechanisms of action, physical training dosage recommendations and adverse effects. Nutrition & Metabolism, 21(1), 41.

10. Master, P. B. Z., & Macedo, R. C. O. (2021). Effects of dietary supplementation in sport and exercise: A review of evidence on milk proteins and amino acids. Critical reviews in food science and nutrition, 61(7), 1225-1239.
11. Nichols, S., McGregor, G., Al-Mohammad, A., Ali, A. N., Tew, G., & O'Doherty, A. F. (2020). The effect of protein and essential amino acid supplementation on muscle strength and performance in patients with chronic heart failure: a systematic review. European Journal of Nutrition, 59(5), 1785-1801.
12. Tipton, K. D., Hamilton, D. L., & Gallagher, I. J. (2018). Assessing the role of muscle protein breakdown in response to nutrition and exercise in humans. Sports Medicine, 48, 53-64.
13. Torre-Villalvazo, I., Alemán-Escondrillas, G., Valle-Ríos, R., & Noriega, L. G. (2019). Protein intake and amino acid supplementation regulate exercise recovery sand performance through the modulation of mTOR, AMPK, FGF21, and immunity. Nutrition research, 72, 1-17.
14. Wilkinson, D. J., Bukhari, S. S., Phillips, B. E., Limb, M. C., Cegielski, J., Brook, M. S., ... & Atherton, P. J. (2018). Effects of leucine-enriched essential amino acid and whey protein bolus dosing upon skeletal muscle protein synthesis at rest and after exercise in older women. Clinical nutrition, 37(6), 2011-2021.
15. Williamson, E., Kato, H., Volterman, K. A., Suzuki, K., & Moore, D. R. (2019). The effect of dietary protein on protein metabolism and performance in endurance-trained males. Med Sci Sports Exerc, 51(2), 352-60.

The Machine Learning Approach to Data Mining Sentiment Analysis Using SVM Classifier

Afsana Khan

Research Scholar

Abstract--- There have been a lot of recent attempts to use online entertainment to investigate broad assessment. The weight of efforts can be used to impart important theories of social general assessment through virtual entertainment. Therefore, a plausible technique is needed to determine this issue. Sentiment analysis is seen as a grouping problem because it classifies a message's direction into positive and negative categories. This article offers experimental findings using Backing Vector Machine (SVM) on benchmark datasets to create a sentiment classifier. The most traditional features were extracted using various weighting schemes and N-grams. It also examines Chi-Square weight highlights in order to select instructional components for the classification. Test results show that employing Chi-square element selection can significantly increase order precision.

Keywords--- Classifier, Data mining, Machine learning, Sentiment analysis, Super Vector Machine.

I. INTRODUCTION

The rapid development of the portable web has led to an overall increase in the use of microblogging services. However, the practice of exchanging opinions and firsthand knowledge about labor and goods is becoming increasingly well-known. Before deciding which new goods or service to buy, clients also turn their attention to surveys completed by previous clients. In a similar vein, businesses can use microblogging platforms such as Facebook, Twitter, and others to collect customer feedback on their work and goods in order to research, analyze, and finally improve that specific labor and goods. But it isn't realistic to read through every survey that is tweeted [1]. A small team of scientists has been developing robotized methods and algorithms for sentiment analysis and message clustering. Sorting the given message into impartial, negative, and positive classifications is sentiment analysis's main goal. These days, there are three main approaches to sentiment analysis: dictionary driven (a combination of machine learning and vocabulary), crossover, and machine learning-based. SVM is used in the supervised class of machine learning computations. The first step should be to create a supervised machine learning computation using prepared data from a pre-selected outcome class. Could it really recognize real information data (test data) at that point? A variety of annotated datasets related to various themes are used in machine learning computations for sentiment analysis and categorization. The dataset from the client survey A few examples of these clarified datasets are the orientation characterisation dataset, the Amazon item audit dataset, and the upsides and downsides dataset [2].

The field of sentiment analysis (SA) came into being in the 20th century. Eliminating the demands and requirements of clients from tweets is a noble gesture. Many embodied surveys for various connected data online have accompanied the growth of internet-based applications. Web journals, online entertainment platforms, and web crawler sites are just a few places where these surveys can be found [3]. The judgment in these audits is excellent in terms of its awareness and planning of activities for both clients and tourists. The proliferation of web crawlers like as Hurray and Google has inundated people with an excessive number of surveys pertinent to their specific goals. SA markets itself as the greatest tool for customers to combine survey sentiments and separate important data. Different approaches and tactics are undoubtedly notable. Determining the optimal procedures for pre-handling, highlight extraction, and characterisation stages is a challenging task. Determining which word references are appropriate for precisely planning terms from tweets to achieve accurate results is also challenging [4].

People are identifiable by their ability to express their sentiments and feelings orally. The most popular method for analyzing these feelings, which may be identified by the expressions that are feasible, is sentiment analysis. It is typically considered beneficial to share your experience with others regarding the problems you have encountered because it may help them make the right choice [5]. As a result, almost all businesses that work with clients these days have a section set apart for social gathering their intelligent data. These comments can be gathered in a variety of ways, such as through email, online apps, mobile apps, review systems, and more. A variety of audits and criticisms are compiled and made available to the public. When selecting a service or product, them can take these surveys into account, and organizations can use this contribution to provide clients with greater assistance [6].

Tests using an open-source, free data mining tool are presented in this research. Nevertheless, some people occasionally outperform others in terms of output. The analyses used benchmark datasets with various term-weighting strategies for extracting highlights and Chi-square for selecting highlights. Backing Vector Machines have been used to finish Backing Vector Machines (SVM). The following metrics were used to assess the outcomes and determine whether the suggested approach was adequate: accuracy, review, precision, F measure, and AUC.

II. RELATED WORK

For IoT systems, Ejaz Ahmed et al. (2017) conducted a review of the most recent advancements in the investigation of huge data. The primary concerns related to data collection expertise, data processing, security, and investigation were also mentioned. These problems result from the organization's expansion, which is encouraged by frequent processing of enormous amounts of data. Their analysis also demonstrates the potential openings in an IoT environment and the coordination of massive amounts of data. The main components of big data, research, and IoT are arranged with their challenges. The developers used large data sources, system components, big data enabling improvements, utilitarian viewpoints, and examination kind to categorize massive data and investigate solutions for the web of things. These configurations show the levels of big data handling that need to be addressed. Additionally, they focused on how to use data analysis in Internet of Things applications [7].

Abbreviations and emojis in light of the Dictionary M. Edison and A. Aloysius (2017) suggested a sentiment analysis (SA) approach on massive data sets, which creates vocabularies and looks at the wording in customer surveys, comparing them with word references that provide comments in categories like positive, negative, and nonpartisan [8].

By compiling previous tweets, Aldo Hernandez-Suarez et al. established a Friendly Conviction Sensor for Twitter. Following the calculation of three unique grouping calculations, the processed conclusion reveals that for negative, good, and security based tweets, Greatest Entropy conveys more powerful consequences than NB and SVM. The job of l 1 guideline assisted with working on the assessment of the Agreeable Conviction Sensor's huge conjecture of coefficients, which could reach around 80% precision among cyberattack and Twitter people group air. Since it efforts to recreate the offered strategy to foretell situations like pandemics, analyze, sacred game plans, and shop research, its usage isn't constrained to digital assaults [9].

Ankit Pradeep Patel, et al., presented how to preprocess tweets from sites and extract data from them. SA assumes an urgent part in calculating judgment in a specific region. Furthermore, it introduced an investigation of SVM, a text affiliation strategy that might be utilized to research the actual sequence of tweets. Enormous Layered Properties Space, which might dispose of the demand for highlight determination, is one of the factors that SVM produces. Accuracy and review can be applied to quantify SVM execution. This review suggests that SVM performs superior for text characterisation over ANN [10].

Pooja C. Sangvikar made the determination that there is space for progress in the data recovery domain because of the worldwide extension and continual improvement of the internet media. In addition, due to its varied workouts throughout web medias, rising SA and choice mining is chipped in as an exciting subject of study. The review subtleties choice mining procedures that experience the negative effects of diverse concerns, including rightness, expandability, configuration of data space dependability, and so forth. These difficulties should be settled, and those answers can be applied to work on the methodologies for sentiment analysis and SA[11].

Xing Tooth and Justin Zhan (2018) developed "Sentiment Analysis utilizing Item Survey," taking into account a dataset containing over 5.1 million item surveys from Amazon.com, separated into four categories: excellence, books, electronics, and household machines. The sentiment extremity order, which makes up most of the problems in sentiment analysis, is examined through data. The underlying abstract content is extracted for further research. Since client audits include the client's name, contact information, survey, and evaluations, they are used as data sources. Regarded as the ground truth for a more precise examination of the survey's sentiment is the survey and rating. Using a maximum entropy POS tagger, the words in the sentence are divided into 46 labels. 56 million action words, 25 million descriptions, and 22 million modifiers are produced as a result, providing unmistakable evidence of the sentiment that is consistently expressed. Invalidation terms such as actually no, not, and then some are included in modifiers; nevertheless, phrases are expressly indicated by the nullification of the descriptor and the denial of the action verb. Out of a total of 0.68 million expressions, 21,586 are recognized. In addition, the software generates an expression list considering recurrence. The following are some of the several characterization models that were chosen for arrangement: Irregular Woods, Straightforward Bayesian, and Backing Vector Machine [12].

III. PROPOSED METHOD

The sequential process and strategies that are regularly used in sentiment classification techniques are summarized in Figure 1. A methodology has also been studied that attempts to categorize liabilities based on conversation fragments. This model use data as a component of a collection of capabilities that enable sentiment categorization on a given dataset. The major markers of sentiment articulation are supposed to be the model grammatical elements, which revolve around subjectivity discovery and highlight areas of strength for the relationship between the existence of modifiers and sentence subjectivity. Still, some experimental findings show that using essentially descriptive terms as characteristics leads to poor performance [13].

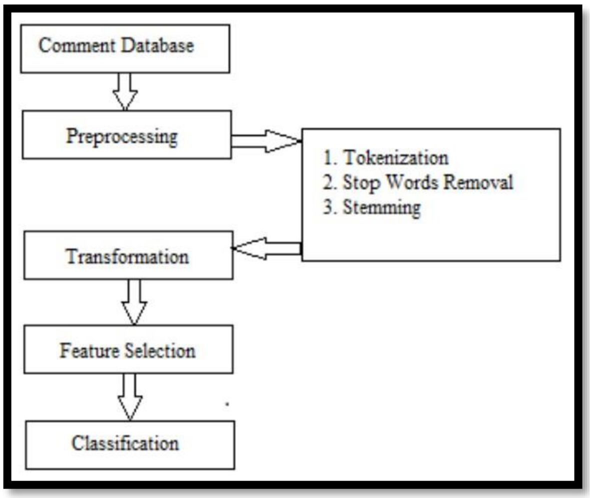


Figure: 1. Steps and methods for categorizing sentiment.

The goal of this study is to improve SVM by using benchmark datasets from Ache Corpus and Taboada Corpus. The steps that comprise the system are preprocessing, feature extraction, highlight choice, and organization. The achievement metric will likewise be briefly explained in the subsection that follows.

A. *Preprocessing Methods*

The datasets will undergo the pre-handling of the text reports, which includes stemming, tokenization, lowercase modification, and stop word expulsion. The process of tokenization involves breaking a text up into words, phrases, or other meaningful sections, or "tokens." Conjunctions, relational terms, and other words that are frequently used in texts but have no connection to a specific subject are known as stop words. Another stage of preprocessing is lowercase transformation. All uppercase characters are usually distinct from their lowercase counterparts prior to the characterization procedures. Obtaining the root and stem of identified words is the final step in the stemming system. The most well-known method for English stemming is the Doorman Stem.

B. Datasets Description

For the primer analyses, we used two mark datasets: 2000 positive and negative Film Audit Datasets and 400 positive and negative SFU Survey Corpus Datasets.

- **Ache Corpus:** The corpus was created to categorize movie reviews that were collected from IMDb.com. The assortment consists of 2000 audits total, 1000 positive and 1000 negative.

- **Taboada Corpus:** In this selection, 400 opinions were compiled from the website Epinions.com and divided into 200 "suggested" (positive) and 200 "not suggested" reviews. (negative). The databases include item and administration surveys for a wide range of products, including movies, books, cars, phones, and more.

C. Feature extraction

Highlight extraction is the process of turning the information data into a collection of items. Selecting the focal points is essential since the components of the machine learning cycle greatly influence how it is presented. Gathering and transforming the information data into a number of illustrative pieces (highlights vectors) that the classifier can actually use is the goal. Regardless, one of our main objectives is to use a few n-gram models—such as bigrams, trigrams, and unigrams—to examine the impact of using various n-gram plans [14].

- **Term-weighting Plan:** For the classifier's feedback, the most common highlights are primarily extracted based on the word weighting method calculation. The more conventional the highlights are, the better the classifier will work. A variety of term-weighting techniques, such as Binary Occurences (BO), Term Frequency Inverse Document Frequency (TFIDF), and Term Occurences (TO), were used in the preliminary stages to generate the word vectors.

D. Feature Selection

There are three ways for include determination: channel, covering, and inserting. Channel techniques were utilized in the testing because to the classifier's freedom and the channels' comparably quick calculation times.

- **Chi-Square:** The channel property appraisal of Chi Square weight highlights was employed to choose relevant elements, and the positioning methodology was also applied to exclude irrelevant elements. CHI2 stands out as a prominent approach for incorporate choice. A quantifiable method for determining how independent two occurrences are is the CHI2 test. In the unlikely event that X and Y are seen as independent events,

$$p(X,Y) = p(X)p(Y). \quad \textbf{(1)}$$

In terms of text highlight selection, these two occurrences are connected to the existence of a specific word and class, independently. The following recipe can be used to determine CHI2 values:

$$CHI2(t,C) = \sum_{t\in\{0,1\}} \sum_{C\in\{0,1\}} \frac{(N_{t,C} - E_{t,C})^2}{E_{t,C}} \quad \textbf{(2)}$$

N is the observed frequency and E is the expected frequency for each condition of term t and class C. The difference between expected counts E and observed counts N is measured by the chi2 statistic. It is suggested that the freedom speculation is false by a high CHI2 number. If the two events are connected, the event of the term increases the likelihood that the class will occur as well. As a result, the term from earlier is important as an ingredient. The CHI2 score for a term is calculated for each class. There are two approaches to handling the globalization of this score for every class. The first way involves calculating the weighted normal score for each class, and the second involves selecting the class with the highest score. This paper recommends the prior system to globalize CHI2 as an incentive for all classes, as in

$$\sum_{i=1}^{M} P(C_i).\,CHI2(t,C_i), \quad \textbf{(3)}$$

E. Selection of Text Classification Techniques

Support Vector Machine (SVM) has been selected as the arranging technique for the experiments. The assist vector machine is a learning tool for two-bunch characterisation problems. Sorting is effective when either positive or negative criteria are used. SVM works well for text categorization because of its benefits, such as its capacity to handle extreme highlights. Another benefit of SVM is its resilience in the face of rare events and its ability to separate most problems easily. Prior research on sentiment analysis has shown encouraging outcomes for Support Vector Machine.

F. Efficiency Metrics

Four effective measures have been employed in this study based on the confusion matrix's output: False Positive (FP), True Negative (TN), True Positive (TP), and false Negative (FN).

- Precision(P) = TP/(TP+FP)
- Recall(R) = TP/(TP+FN)
- Accuracy(A) = (TP+TN)/(TP + TN + FP + FN) • AUC (Area under the (ROC) Curve) = 1/2.((TP/(TP+FN))+(TN/(TN+FP))
- F-Measure(Micro-averaging) = 2.(P.R)/(P+R)

The F1, exactness, and AUC are commonly used to evaluate text order adequacy. Consolidated viability is addressed by the F1 measure, which is determined by accuracy and review. These days, the region under the ROC bend (AUC) is a widely used measure of how much supervised order approaches are capable of. However, only two classes are able to use the AUC in its simplest form [15].

IV. RESULT AND DISCUSSION

As in past text characterization tasks, the model will be trained on a training dataset and its precision will be evaluated on a different testing dataset. Every set was divided in half, using half for training and the other for testing. This resulted in a 70:30 ratio, or 70% for training and 30% for testing. Three optional weighting plans—binary occurrence (BO), term occurrence, and word frequency in documents and across the corpus (TFIDF)—were used to create the word vectors (TO). Support Vector Machine (SVM) was utilized in this investigation to categorize the testing datasets as positive or negative.

Tables 1 and 2 demonstrate the comparison of training and testing outcomes on the Taboada Corpus.

Table: 1. Result of 10-Fold Cross-Validation Training by Taboada Corpus

		Accuracy	Precision	Recall	AUC
TFIDF	Unigrams(%)	75.23	77.19	73.45	0.826
	Bigrams(%)	72.02	72.61	74.16	0.778
	Trigrams(%)	65.59	65.32	69.16	0.704
BO	Unigrams(%)	69.16	64.37	91.31	0.786
	Bigrams(%)	61.31	58.32	92.73	0.707
	Trigrams(%)	60.95	58.17	92.73	0.669
TO	Unigrams(%)	61.31	57.82	94.16	0.756
	Bigrams(%)	58.45	56.10	92.73	0.719
	Trigrams(%)	58.09	55.94	96.31	0.664

Table: 2. Results of the 10-Fold Cross-Validation Testing by Taboada Corpus

		Accuracy	Precision	Recall	AUC
TFIDF	Unigrams(%)	79.52	77.76	85.35	0.869
	Bigrams(%)	75.35	73.61	85.35	0.839
	Trigrams(%)	69.52	65.55	90.35	0.758
BO	Unigrams(%)	54.52	53.47	97.02	0.839
	Bigrams(%)	52.02	52.02	100	0.839
	Trigrams(%)	52.02	52.02	100	0.759
TO	Unigrams(%)	60.35	57.44	93.69	0.797
	Bigrams(%)	55.35	53.93	100	0.858
	Trigrams(%)	52.02	52.02	100	0.766

Tables 1 and 2 demonstrate that the unigram model has higher AUCs for SVM regardless of the weighting method used. The most notable AUC value from Table 1 is 0.826, and the highest AUC value from Table 2 is 0.807. In both tables, 75.23% and 73.07% deal with the most notable precision at the same time. Term occurrences (TO) and Binary occurrences (BO) are outperformed by TFIDF in these studies.

The results of the AUCs for the training process in one more experiment from the Pang Corpus are shown in Table 3.

Table: 3. Results of the Pang Corpus 10-Fold Cross-Validation Training

		Accuracy	Precision	Recall	AUC
TFIDF	Unigrams(%)	84.52	86.12	82.45	0.929
	Bigrams(%)	68.38	68.51	67.88	0.747
	Trigrams(%)	54.59	54.25	55.45	0.569
BO	Unigrams(%)	86.52	86.51	86.88	0.945
	Bigrams(%)	65.95	65.64	66.88	0.727
	Trigrams(%)	54.31	55.06	70.02	0.575
TO	Unigrams(%)	83.52	81.65	87.16	0.915
	Bigrams(%)	68.31	67.60	70.59	0.747
	Trigrams(%)	53.59	52.74	80.59	0.577

Table 3 demonstrates the superiority of unigram-based SVMs with AUCs of 0.945 and precision of 86.52% when employing the BO weighting scheme.

Table 4 presents the AUC results of the testing procedure.

Table: 4. Results of the 10-Fold Cross-Validation Testing by Pang Corpus

		Accuracy	Precision	Recall	AUC
TFIDF	Unigrams(%)	81.85	80.94	83.69	0.892
	Bigrams(%)	69.69	71.09	68.35	0.766
	Trigrams(%)	56.85	62.84	28.02	0.568
BO	Unigrams(%)	85.52	86.41	84.35	0.939
	Bigrams(%)	67.69	70.20	61.02	0.749
	Trigrams(%)	57.35	68.04	22.69	0.569
TO	Unigrams(%)	80.52	81.31	79.35	0.885
	Bigrams(%)	65.85	67.99	59	0.739
	Trigrams(%)	57.02	65.17	23.35	0.573

Furthermore, the results are consistent with the training data, which showed that unigram models using a BO weighting scheme had the highest accuracy and AUCs. Table 4's best AUC is 0.939. Overall, the results indicate that the unigram model outperforms the other ngram models. To acquire the maximum number of agent highlights for the Ache Corpus classifier, the Binary Occurences (BO) weighting mechanism is essential. In the interim, unigram models and the TFIDF weighting scheme are essential to the classifier's development in the Taboada Corpus.

Table 5 presents a comparison between the different feature selection methods and the F-measure on the Taboada Corpus.

Table: 5. After Chi-Square Feature-Selection With Respect To F Measure on Taboada Corpus from 10 Fold Cross Validation Results, the Classification Results

	Accuracy	Precision	Recall	F Measure
TFIDF	56.19%	0.783	0.564	0.44
BO	91.19%	0.933	0.914	0.91
TO	90.35%	0.916	0.905	0.905

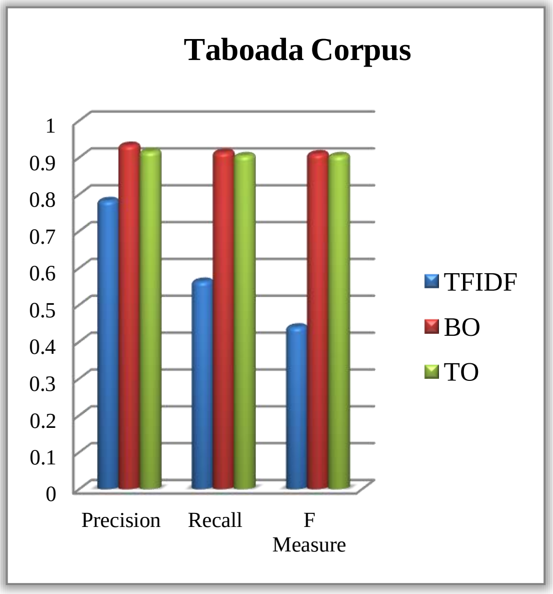


Figure: 2. a visual display of The Classification Results Following Chi-Square Feature-Selection With Respect To F Measure On Taboada Corpus

Table 5 shows how much better the SVM's F measure is in relation to Chi-square when compared to the prior preliminary without highlight selection. The accuracy obtained was

91.19%, and the F-measure's value was 0.91. These results are more significant than those from previous tests.

Chi-Square feature selection yields the F-measure of SVM on the Pang Corpus, as Table 6 illustrates.

Table: 6. After Chi-Square Feature-Selection With Respect To F-Measure on Pang Corpus from 10 Fold Cross Validation Results, the Classification Results

	Accuracy	Precision	Recall	F Measure
TFIDF	83.02%	0.876	0.83	0.826
BO	90.85%	0.909	0.908	0.908
TO	91.19%	0.914	0.914	0.914

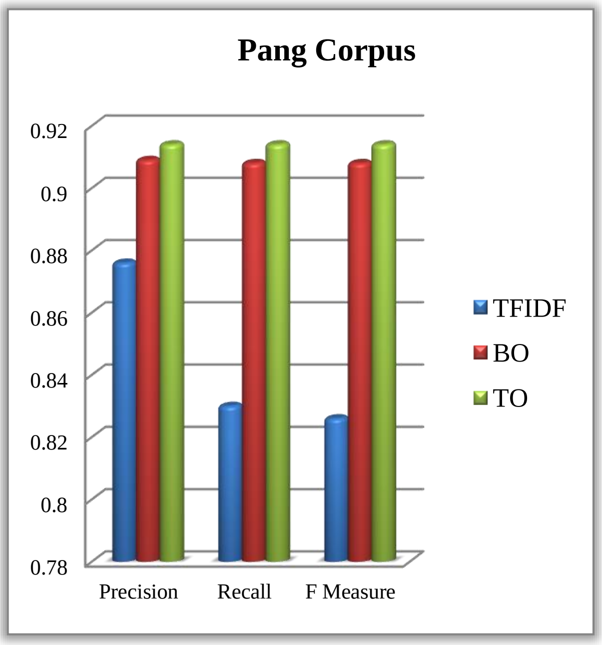


Figure: 3. Chi-Square Feature-Selection With Respect To F-Measure on Pang Corpus Classification Results Graphical Representation

The F-measure of SVM findings in Table 6 is superior than the results obtained in the absence of feature selection. The F-measure has a value of 0.914 and the accuracy is 91.19%.

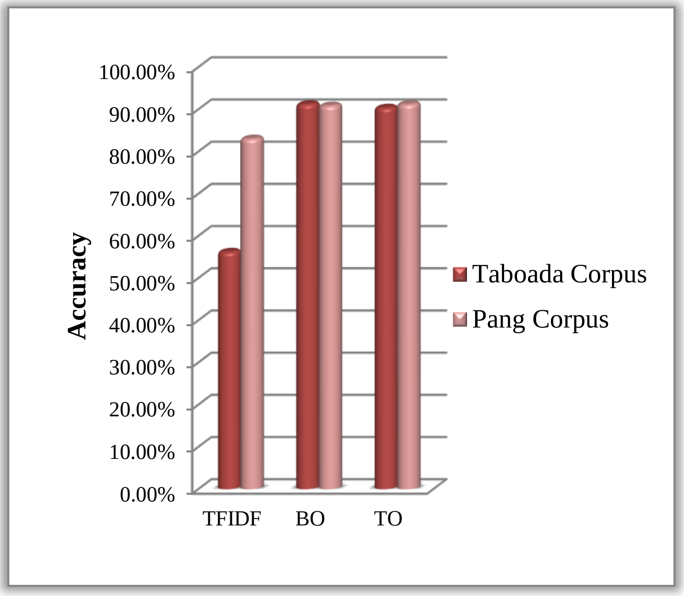


Figure: 4. Taboada and Pang Corpus Accuracy Comparison from 10 Fold Cross Validation Results

The overall results show that the approach of selecting the parts based on how well their chi-squared measurements suited them reduced the dimensionality and noise in the text, allowing the classifier to perform better and possibly even achieve topic order. It also demonstrates how selecting fewer highlights increases the expectation correctness of SVM predictions, as do other factors like corpus size and domains that may have an impact on the classifier's performance.

V. CONCLUSION

Sentiment analysis is one of the most well-known focus areas in the subject matter disclosure. A significant amount of internet content is often added, such as blog posts, comments from online entertainment, and reviews of movies and TV shows. These data sources can be used to obtain appropriate information by utilizing sentiment analysis techniques, for example, political race result expectation, client feedback on programming, brand notoriety analysis, popular opinion gathering previous item send off, and so on. Sentiment analysis should be achievable utilizing several methodologies, including vocabulary based, machine learning-based, and cross breed draws near. SVM is a well-known machine learning technique for removing extremities from text. Today, distinct specialized and integrated models for sentiment analysis and extremity identifiable proof have been introduced by scientists notwithstanding traditional machine learning order procedures. In this review, benchmark datasets were utilized to train an SVM-based sentiment classifier, which involves n-grams and an array of weighting schemes as input to the classifier. Unigrams perform better compared to other n-gram models for both datasets, according to the perceptions, while Binary Occurences (BO) and TFIDF weighting plan are basic for acquiring the most customary highlights for Ache Corpus and Taboada Corpus. The results also show that using chi-square element determination can significantly improve the arrangement accuracy for both datasets.

VI. FUTURE SCOPE

Another method will be used for sentiment analysis for cloud clients in the future. A novel, persuasive emotion characterization algorithm will be used to characterize online entertainment data. Deep learning and machine learning techniques will be applied to asset scheduling. To communicate outcomes, a respectable perception technique will be used. In order to get better results, a different classifier will be created and employed. We'll create a new lexicon to appropriately organize the emotions.

REFERENCES

1. [1] M. Abdul-Mageed, M. Diab, and S. Kubler, "Samar: Subjectivity ¨ and sentiment analysis for arabic social media," Computer Speech & Language, vol. 28, no. 1, pp. 20 – 37, 2019
2. [2] L. C. Borges, V. M. Marques, and J. Bernardino, "Comparison of data mining techniques and tools for data classification," in Proceedings of the International C* Conference on Computer Science and Software Engineering, ser. C3S2E '13. New York, NY, USA: ACM, 2017, pp. 113–116.
3. [3] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," Procedia Computer Science, vol. 17, no. 0, pp. 26 – 32, 2019, first International Conference on Information Technology and Quantitative Management.
4. [4] Rucha Jadhavar, Agastya Kumar Komarraju, "Sentiment Analysis of Netfix and Competitor Tweets to Classify Customer Opinions", Paper 2708-2018.
5. [5] Mamatha M, Thriveni J, Nenugopal K. R., "Techniques of Sentiment Classification, Emotion Detection, Feature Extraction and Sentiment Analysis:

A Comprehensive Review", International Journal of Computer Sciences and Engineering (IJCSE), E-ISSN: 2347-2693, 2018.

6. [6] E. Martinez Camara, V. Martin, M. Teresa, J. M. Perea Ortega, and L. A. Urena Lopez, "Tecnicas de clasificacion de opinionesaplicadas a un corpus enespanol," Procesamiento de Lenguaje Natural, vol. 47, pp. 163–170, 2020
7. [7] Ahmed Ejaz, Yaqoob, Ibrar Hashem, Ibrahim AbakerTargio Khan, Imran Ahmed, Abdelmuttlib Ibrahim Abdalla, Imran Muhammad &Vasilakos Athanasios, V 2017, 'The role of big data analytics in Internet of Things', in proceedings of Computer Networks, vol. 129, pp. 459-471.
8. [8] Edison, M & Aloysius, A 2017, 'Lexicon based Acronyms and Emoticons Classification of Sentiment Analysis (SA)', on Big Data International Journal of Database Theory and Application, vol. 10, no. 7 , pp. 41-54.
9. [9] Xing Fang & Justin Zhan 2018, 'Sentiment analysis using product review data', Journal of Big Data, https://doi.org/10.1186/s40537-015- 0015-2
10. [10] Aldo Hernandez-Suarez, Gabriel Sanchez-Perez, Karina Toscano-Medina, Victor Martinez-Hernandez, Hector Perez-Meana, Jesus OlivaresMrcaso and Victor Sanchez, "Social Sentiment Sensor in Twitter for Predicting Cyber-Attacks Using ℓ1 Regulation", March 2018 in Preprints(www.preprints.org), Instituto Politecnico Nacional, ESIME Culhuacan, Mexico City 04440, Mexico.
11. [11] Ankit Pradeep Patel, Ankit Vithalbhai Patel, SanjaykumarGhanshyambhaiButani and Prashant B. Sawant, "Literature Survey on Sentiment Analysis of Twitter Data using Machine Learning Approaches" in International Journal for Innovative Research in Science & Technology, ISSN:2349-6010, Volume 3, Issue 10, March 2017, Mumbai, India.
12. [12] Pooja C. Sangvikar, "A Survey on Sentiment Analysis and Opinion Mining", International Journal for Scientific Research & Development, Vol. 3, Issue 10, 2019, ISSN(online):2321-0613, Pune, India.
13. [13] N. Genc-Nayebi and A. Abran, "A systematic literature review: Opinion mining studies from mobile app store user reviews," J. Syst. Softw., vol. 125, no. November, pp. 207–219, 2017.

14. [14] S. Ashraf and S. Aftab, “Latest Transformations in Scrum: A State of the Art Review,” Int. J. Mod. Educ. Comput. Sci., vol. 9, no. 7, pp. 12–22, 2017.
15. [15] Sangeeta, Twitter Data Analysis Using FLUME & HIVE on HadoopFrameWork,Special Issue on International Journal of Recent Advances in Engineering & Technology (IJRAET) V-4 I-2 For National Conference on Recent Innovations in Science, Technology & Management (NCRISTM) ISSN (Online): 2347-2812, Gurgaon Institute of Technology and Management, Gurgaon 26th to 27th February 2019

Mechanism of Disease Detection via Artificial Neural Networks

Harpreet Kaur

Research Scholar

Abstract--Artificial neural networks (ANNs) are becoming increasingly popular due to their capacity to learn, correct mistakes, and transform a massive amount of unrefined data into practical clinical decisions for care and treatment. Understanding security and care quality has improved as a result of this. The organization viability of various Artificial Neural Network (ANN) models for the muscle disease osteoporosis is evaluated in this review. Osteoporosis risk prediction should be apparent as an example grouping difficulty based on different clinical features. Probabilistic neural networks (PNNs) and multi-layer perceptron's (MLPs) were used to handle the prediction of osteoporosis risk variables. In light of ANNs, this approach is the main computational insight tool for an osteoporosis risk assessment in the Greek population. The results demonstrate the superior performance of PNNs over MLPs, demonstrating its applicability as a computational insight tool for predicting osteoporosis risk variables. Also, the overfitting problem became more prevalent when the spread worth increased for MLPs rather than PNNs.

Keywords-- Artificial neural network, Disease Detection, Multi-Layer Perceptrons, Prediction, Probabilistic Neural Networks.

I. INTRODUCTION

Artificial neural networks provide physicians with a potent tool to help them analyze, model, and decipher complicated clinical data in a variety of clinical applications. Given the intentional information, the majority of artificial neural network applications in medicine involve grouping tasks where the patient should be assigned to an extremely rare instance of a class. Artificial intelligence (AI), with its countless uses, is transforming the world. AI stages are widely used, hastening the development of life-saving medications and lowering operational expenses. The main component of AI innovation is artificial neural networks (ANN), which mimic the human brain's inspection and processing capabilities to address complex problems [1]. A multitude of fields, such as the clinical sciences (particularly in the fields of cardiology, radiology, oncology, and urology), veterinary medicine, stock trading, law enforcement, human resource management, and network security, have benefited greatly from the remarkable characteristics of artificial neural networks (ANNs), which include efficient information handling, low complexity, reduced computation and capacity requirements, and productive information handling. The neural network (ANN) is mostly built on numerical models that take inspiration from natural sensory systems, such as the way the brain transmits information. The ANN functions essentially in the same way as a flexible framework, which modifies its configuration throughout the learning phase and can be illustrated for a particular application, such as the categorization of instances and the order of information. The input layer, stowed away layer, and outcome layer are the three main layers of a neural network. According to these layers, ANN can be divided into four distinct types: a single layer feed-forward neural network (FFNN), a multilayer repeating network, and a single hub with its own critique [2].

The equal creation of the human brain is the goal of an artificial neural network (ANN), a computational model. Neural nets, or artificial neural networks (ANNs), are collections of identical, densely networked handling units. These components are inspired by the anatomical nerve frameworks. Similar to natural systems, the relationships among the components of a network do not determine its overall performance. In a network, a layer refers to a portion of the handling components. The outcome layer is at the bottom, and the information layer is at the top. Between the information and result layer (s), there may be one or more additional levels of units, referred to as covered up layers [3]. A neural network can be trained to do a certain task by altering the upsides of the associations (loads) between components. Conventional neural networks Clinical decision support using artificial neural networks is currently a very active field of study in medicine, and it is anticipated that in the not-too-distant future, it will be applied even more frequently in biomedical contexts. The main reason for this is that the response isn't limited to a direct shape. Neural networks are good for identifying diseases from examines because no precise calculation to determine the ailment should be presented. Given that neural networks learn by doing, it is not necessary to be aware of every detail involved in differentiating across conditions. Given their mode of learning, all artificial neural networks can be classified as either directed learning or unassisted learning [4]. Information and outcome designs are used in administered learning to train the network. Throughout this phase, the neural network can adjust the association loads to align its output with the actual output in an iterative loop until the desired outcome is achieved. Similar to a self-sorting map, an unaided learning ANN receives only contributions and has no predetermined results. Before presenting the data inputs in its own way, the network should determine the appropriate association loads. By grouping the incoming data, the inherent characteristics of the problem are discovered. [5].

This work focuses on the creation and evaluation of ANN design acknowledgment models based on probabilistic neural networks (PNNs) and multi-layer perceptron's (MLPs), as well as their application to the prediction of osteoporosis risk factors. More specifically, a decision support tool has been developed to assist clinicians in determining whether patients require additional bone densitometry testing since their risk of osteoporosis is higher. This application location is thought to be of the utmost importance since early detection of osteoporosis is essential for preventing osteoporotic breaks, which are linked to increased morbidity and death as well as substantial financial costs. The proposed ANN architectures and their display on clinical data are shown in this examination [6].

II. RELATED WORK

The model is trained using the back-propogation procedure and the Caffe structure in a concentrate by Sue Han Lee. Deconvolutional Network was one of the methods employed to comprehend the CNN model's internal operations and visualize the selected channels. The Deconvolutional Network provides a capacity that enables us to visualize the element map at each layer by deconvoluting and unpooling down to the information image pixel. The models were trained using two datasets: the original MalayaKew Dataset and an expanded version of the first dataset that included neighboring leaf information. According to trial findings, the model trained on the modified dataset somewhat improved the model's accuracy when compared to the model trained on the original dataset [7].

Gil, Johnsson, Garicia, Paya, and Fernandez examined the development of two artificial neural network models as tools to aid in the clinical diagnosis of urological dysfunctions. Two different types of solo neural networks and one administered neural network were created [8].

Altunay, Telatar, Erogul, and Aydur evaluated the uroflowmetric data, and they also helped experts with their analysis. They therefore promoted a master pre-finding strategy for determining expected side effects from the uroflow signals. A pre-indicative result was produced by the framework using artificial neural networks (ANN) [9].

Moein, Monadjemi, and Moallem examined the actual clinical discovery process—which is often finished by specialists—and transformed it into a construction that could be used by a machine. Following the selection of a few adverse effects from eight distinct ailments, data from an information assortment containing the specifics of two or three hundred occasions was then built and put to an MLP neural network. Discussions also covered the trial's findings and the benefits of using a fluffy technique. The findings emphasize the importance of carefully selecting side effects and the benefits of information fluffing in a clinical finding framework that is programmed in light of neural networks [10].

Francisco, Juan Manuel, Antonio, and Daniel created a unique framework based on a multiagent framework model where each brain focus is compared to a specialist. This framework uses a heuristic to support it in the face of potential discrepancies. Symmetric affiliated memory is a neural network-based heuristic. Training has given information to the architecture by using as a base the legitimate urinary parcel behaviors and behaviors caused by dysfunctions of the two neural centers [11].

Monadjemi and Moallem looked on the analysis of normal diseases using artificial neural networks. The actual symptomatic cycle that experts practically used was examined and incorporated into a design that could be implemented by a machine. Talks also covered the findings of the inquiry and the benefits of using a fluffy technique [12].

The impact of class awkwardness in training data was examined by Mazurowski, Habas, Zurada, Lo, Bread cook, and Tourassi when developing neural network classifiers for PC assisted clinical conclusion. The request also has other characteristics that are typical of clinical information, such as a small training test size, lots of highlights, and linkages between its sections [13].

Higuchi, Sato, Makuuchi, Furuse, Takamoto, and Takeda examined a three-layered artificial neural network analysis of phonocardiogram accounts in order to subsequently and objectively review the state of the heart in patients with heart murmurs [14].

III. MATERIALS AND METHODS

A. Artificial Neural Networks

An artificial neural network (ANN) is a mathematical representation of the "learning" and "generalization" abilities of the human brain network. For this reason, ANNs are included under the artificial intelligence category. ANNs are widely utilized in scientific research because of their ability to simulate highly nonlinear systems with intricate or ill-defined variable interactions [15].

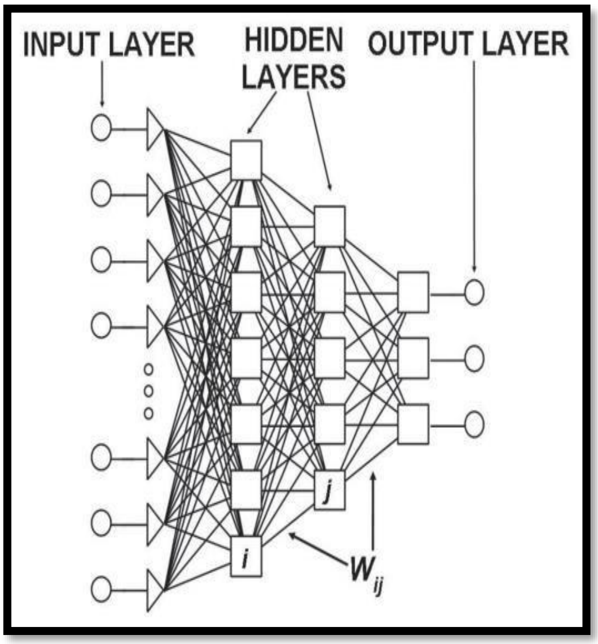


Figure: 1. a neural network having two hidden layers, in general.

- **Mathematical background**

A neural network is composed of layers of different "neurons" (also called "hubs"). A weighted association binds each neuron in one layer to each neuron in the layer above it. The weight wij esteem addresses the strength of the connection between the jth neuron in one layer and the ith neuron in the adjacent layer.

An "input" layer, at least one "stowed away" levels, and the "yield" layer make up the design of a neural network. The intricacy of the framework under consideration significantly affects both the quantity of neurons in a layer and the total number of layers. Subsequently, the best network engineering should be recognized.

Fig. 1 depicts the overall format of a conventional three-layered ANN engineering. The weighted links permit the info layer neurons to get the information and convey it to the principal stowed away layer neurons. The information are quantitatively analyzed here, and the outcome is shipped off the neurons in the accompanying layer. The result of the network is eventually formed by the neurons in the highest layer. The jth neuron in a hidden layer examinations the info information (xi) by I using Eq. 1's weighted total recipe and adding a "predisposition" component (j):

$$\boldsymbol{net_j} = \sum_{i-1}^{m} \boldsymbol{x_i} \times \boldsymbol{w_{ij}} + \boldsymbol{\theta_j} \qquad (\boldsymbol{j} = \boldsymbol{1}, \boldsymbol{2}, \dots, \boldsymbol{n}) \qquad \textbf{(1)}$$

(ii) adding a valid numerical "move capability" to the network, and (iii) transmitting the modified network result to neurons in the layer above. Although there are several other exchange capabilities, the most commonly used one is the sigmoid capability:

$$\boldsymbol{f(x)} = \frac{\boldsymbol{1}}{\boldsymbol{1+e^{-x}}} \qquad \textbf{(2)}$$

- **Network learning**

A definitive client can generally disregard the numerical approach by which the network performs "learning." The network can be viewed in this sense a "black box" that acknowledges a vector with m data sources and results a vector with n.. Here, we will essentially provide a brief overview of the educational process; more details are provided elsewhere. The "training data set" contains various "models" from which the network "learns." The purpose of the training system is to calculate the relationship between the Xim and Yin vectors, or capacity f:

$$\boldsymbol{Y_{i,n}} = \boldsymbol{f(X_{i,m})} \qquad \textbf{(3)}$$

This is accomplished by iteratively modifying the association loads' (wij) values in accordance with the training computation, a correct numerical formula. The steepest drop strategy is used to modify the weight values in order to limit an appropriate capability that serves as the training stopping models. The squared leftover provided by Equation 4 may be the most widely utilized capability.

$$\boldsymbol{E} = \frac{\boldsymbol{1}}{\boldsymbol{2}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \sum_{\boldsymbol{j=1}}^{\boldsymbol{n}} (\boldsymbol{y_{ij}} - \boldsymbol{y_{ij}^{*}})^{\boldsymbol{2}} \qquad \textbf{(4)}$$

where yij and yij* stand for the jth output of the network and the real ith input vector, respectively. The current weight change on a particular layer is given by Equation (5).

$$\Delta w_{ij} = -\eta \frac{dE}{dw_{ij}} \qquad \textbf{(5)}$$

in which the rate of learning is a positive constant (η).An additional term is added to Eq. 5 in order to speed up learning and avoid local minima.

$$\Delta w_{ij}^{k} = -\eta \frac{dE}{dw_{ij}} + \mu \Delta w_{ij}^{k-1} \qquad \textbf{(6)}$$

While keeping an eye on the previous weight change, the energy acts as a stabilizer, and the learning rate controls the weight update rate in accordance with the new weight change. The ability provided by Eq. 4 is also utilized as the model to improve network engineering since it depends on the number of secret layers and neurons within them. The most well-known method for determining the optimal scheme is to graph the value of E (Eq. 4) as a function of the number of hubs in the covert layer (q).

B. Osteoporosis

Osteoporosis is the most commonly known bone disease. It is characterized by decreased bone mass and changes to the minuscule architecture, which reduces the resistance of the bones and increases the risk of fracture. In addition to the obvious discomfort and anguish associated with a fracture, osteoporotic hip or spine cracks are a major cause of morbidity and mortality.

Studies show that two out of every five seniors who get a hip crack die shortly after the injury. Taking care of the immediate and long-term consequences of fractures, such as hospital stays, loss of independence, in-home or institutional help, and so on, is extremely expensive and essential to overall health. Osteoporosis causes someone in the European Association to break a bone like clockwork.

C. Osteoporosis is sometimes referred to as "the quiet crippler" because, although a fractured bone is typically the first obvious sign of the condition, it often goes unnoticed until it is too late to reverse. However, early detection and treatment of osteoporosis can significantly lower a person's risk of developing crack.

The primary cause of osteoporosis development in people between 45 and 50 years of age is a deficit of high osteal thickness mass. The consumption of osteal is greater than the production of osteal in women over 50. It follows that a decrease in osteal thickness mass is anticipated.

Osteal bone densitometry and research center testing are used to determine the existence of osteoporosis, regardless of whether the diagnosis is obtained before or after a bone fracture. This test uses Double Energy X-beam absorptiometry to target individual bones. Because this approach depends on the patient absorbing radiation, it is not recommended for situations involving the entire muscle.

The osteoporosis data used to plan the ANN models came from the Muscular Clinical Data Framework at the College Emergency clinic at Alexandroupolis, Greece. For each patient, four clinical boundaries were taken into account. The models included weight, age, level, and sex. The osteoporosis risk factor was assessed using the T-score esteem, which compares the patient's bone thickness to what is typically predicted in a healthy young adult of the same sex.

The foundation of the continuing investigation is formed by the 3628 cases in the informational collection. Out of this information assortment, a group of 2628 records and another arrangement of 1000 records were separated. The former was used to train MLP networks and construct PNNs, while the latter was used to evaluate how neural networks presented themselves.

D. Models of Neural **Networks for Osteoporosis Risk Assessment**

The non-representative computational intelligence technique known as artificial neural networks (ANN) is used to create the suggested design acknowledgment models for identifying osteoporosis risk variables. For the formation of such an ANN, it is crucial to decide on many boundaries, such as the type of ANN, the number of neurons in each layer, and the applicable learning computation.

MLPs, or feed-forward networks with a back-engendering learning algorithm, are used in the majority of ANN models. Given that MLPs are used in the processing of clinical data, the link between dependent and free components is an essential component of these models.

An MLP consists of two layers: a result layer that displays the findings of the reenactment and has the same number of hubs as the factors of the problem and an information layer that has the same number of hubs as the factors of the problem. To actualize the various components and hidden levels of the MPL design, experimentation would be used.

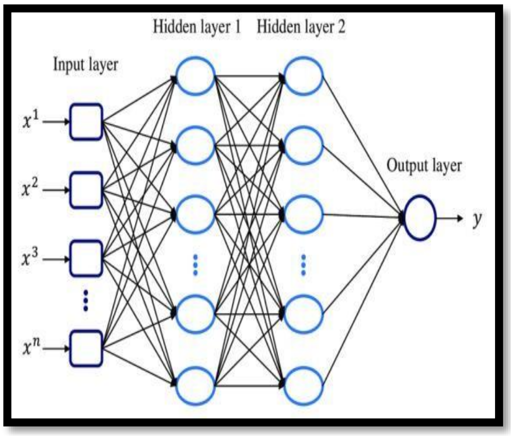


Figure: 2. two hidden layers and one prediction output are present in the multilayer perceptron (MLP) design.

Even while MLPs have been successfully applied in a variety of therapeutic settings, many scientists remain skeptical about them. This argument is prompted by MLPs' ability to find hidden relationships in data, or its "black-box" feature. Without the heuristic component of MLPs, PNNs mimic Bayesian factual hypothesis and identify designs using a known technique to human leaders.

PNNs are constructed using Parzen's Probabilistic Density Function (PDF) assessment. Three layers make up PNNs, which are feed-forward networks: an information layer, a serious layer, and a spiral premise layer. The spread premise layer calculates the separations between the training input and information vectors and builds a vector whose members represent how close each contribution is to each training input. These promises are included in the third layer for each classification of contributions, resulting in a net result of a vector of probabilities. The most notable of these probabilities is selected using a greedy exchange function on the third layer result, resulting in a value of 1 for that class and 0 for the other classes. PNNs can handle undoubtedly more data than MLP neural networks because they do not require an iterative learning process. The behavior of the Bayesian approach produces this PNN's component.

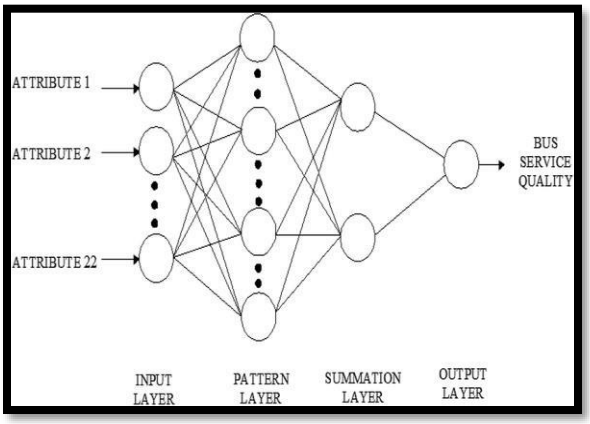


Figure: 3. Architecture of probabilistic neural network (PNN)

The number of neurons that should be used in the information layer depends on the issue's boundaries. Since the risk of osteoporosis is predicted using four variables, the information layer of the employed ANN models in this work consists of four neurons. The number of neurons in the result layer is determined by the type of ANN and the optimal number of parameters for the problem. In particular, the PNN engineering requires as many neurons as the phases of osteoporosis, which make up the quantity of classes of information, but the MLP requires a neuron in the result layer for assessment of the phases of the osteoporosis risk factor.

Information planning is essential to ANN development. In this review, the sex variable was coded as 1 for female members and 2 for male members. Age, level, and weight were acquired as maintained in the data set. The outcome variable of an ANN was the T-score boundary. Based on the T-score result, these bone densitometry findings were coded as a number and divided into 4 phases, as shown in Table 1.

Table: 1. T-Score Value Coding

T-score Value	Coding
≤ - 2.5	1
-2.5 – -1.5	2
-1.5 – 0	3
≥0	4

N is the number of samples, t(k), a(k), and e(k) are the ideal, the MLPs are found, and the error serves as a motivator for design k, respectively, in the numerical documentation. The execution standard of MLPs was assessed using the mean squared error (MSE).

$$\boldsymbol{MSE} = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{e}(\boldsymbol{k})^2 = \frac{1}{N}\sum_{k=1}^{N} [\boldsymbol{t}(\boldsymbol{k}) - \boldsymbol{a}(\boldsymbol{k})]^2 \qquad \textbf{(7)}$$

As recently said, the problem determines the amount of data and output layers for the neurons. This review has one outcome border and four information boundaries, as was made clear. As a result, the patients' osteoporosis risk factor is determined by one neuron in the outcome layer and four neurons in the info layer. The cycle of testing establishes the significance of a hidden layer. A computational scheme that modifies the amount of neurons in the secret layer and sorts the performance of each ANN geography has been devised.

The PNNs design is limited by the open elements of a particular problem, but it is still necessary to define the extent to which each calculated Gaussian bend of the likelihood density function is expansive. In the ongoing evaluation, this spread component ran from 0.1 to 50.

IV. EXPERIMENTAL RESULT

MATLAB Neural Network Tool compartment was used for the structure and execution assessment of ANN models due to its expertise and user-friendly interface. Below is a summary of the MLP execution results in Tables 2 and 3. The second and third sections of this table show, respectively, the engineering of the MLPs and the exchange functions of each ANN's model. The optimal MSE for every MLP configuration is displayed in the fourth section. The rates of accurate predictions for training, testing, and the entire informational collection are displayed individually in the fifth, sixth, and seventh segments. In the eighth, tenth, and tenth segments, respectively, are the rates of neurotic cases with a favorable vision for testing, training, and the entire informational gathering. The actual number of cases that successfully detected MLPs is indicated by the figures in brackets.

Additional processing was applied to the delivered MLPs in order to improve their affinity to gauge. As said, an osteoporosis assessment has four distinct stages: one for patients without osteoporosis and three for those with compulsive problems. The osteoporosis risk factor periods were then divided into two groups: one for individuals with neuroses and another for healthy individuals. To make the best discoveries usable by artificial neural networks, an attempt was made to parallel code them. For usual situations, these results were encoded as 0, and for neurotic cases, as 1.

The outcomes of the MLPs' two-step coding reproduction stage are summarized in Tables 2 and 3. The collected findings are presented in Tables 2 and 3 and demonstrate that MLP execution increases with the number of neurons in the secret layer. This behavior is expected since the loads and inclinations of the secret layer increase with the number of neurons in the system, enhancing the MLP's capacity to store new information. An exhibition correlation of the 4-5-1 and 4-3-1 geographies for MLPs supports this assertion.

There is a maximum number of neurons in a secret layer, and crossing it reduces MLP execution. The number of secret neurons is limited by the overfitting problem, which arises when a neural network remembers the training patterns but is unable to calculate how to sum up to new data.

Despite the necessity for MLPs to categorize cases into four stages, it is evident that their presentation of 2-stage osteoporosis coding has advanced. For instance, this outperforms Nr. 8 MLP when it comes to 2-stage osteoporosis coding. Compared to the original MLP, which was unable to identify individuals without osteoporosis and classified all records as obsessive cases, the preceding MLP correctly identified more instances. It is crucial for implemented MLPs that regular instances were precisely structured, even with regard to the exact full case and neurotic layout.

Table: 2. Experimental Results Using MLP for Coding Of the 4 Stages of Osteoporosis

Nr	Architecture of Artificial Neural Network	Goal MSE	Prognosis Success Rate as a Percentage			Prognosis Over Pathological Situations: Success Rate		
			Testing Set	Training Set	Overall Set	Testing Set	Training Set	Overall Set
1	4-3-1	0.7	30.32	29.99	30.09	35.42	36.33	35.35
2	4-3-1	0.11	32.02	32.22	32.86	37.66	39.25	38.79
3	4-5-1	0.7	32.12	32.76	32.57	37.78	38.71	38.44
4	4-5-1	0.11	31.72	31.89	31.84	37.31	37.58	37.50
5	4-7-1	0.7	32.62	33.05	32.92	38.02	38.61	38.44

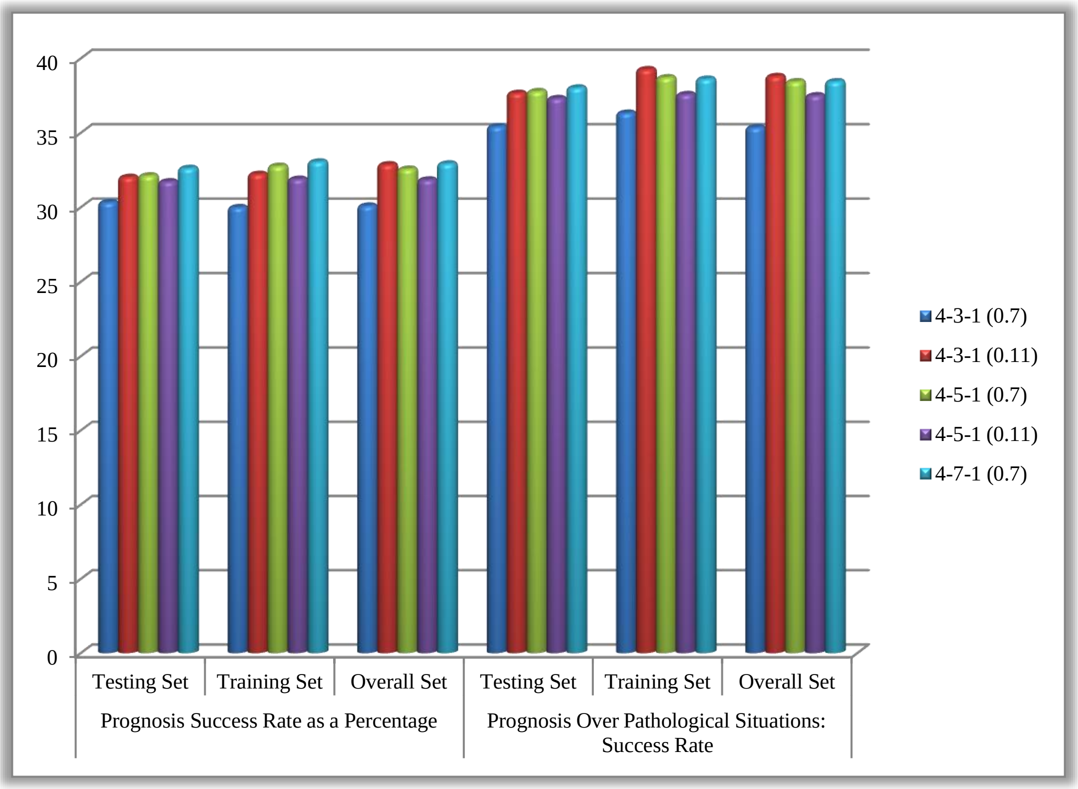


Figure: 4. Illustration of Experimental Results for Coding the Four Stages of Osteoporosis Using MLP

Table: 3. Results of the Experimental Use of MLP For 2-Stage Osteoporosis Coding

Nr	Architecture of Artificial Neural Network	Goal MSE	Prognosis Success Rate as a Percentage			Prognosis Over Pathological Situations: Success Rate		
			Testing Set	Training Set	Overall Set	Testing Set	Training Set	Overall Set
1	4-3-1	0.7	84.72	84.82	84.79	98.77	99.45	99.25
2	4-3-1	0.11	20.32	20.23	20.26	5.63	5.18	5.31
3	4-5-1	0.7	68.62	71.22	70.46	78.68	82.62	81.46
4	4-5-1	0.11	82.62	82.42	82.49	93.93	92.50	92.92
5	4-7-1	0.7	42.62	47.20	45.15	45.94	49.58	48.51

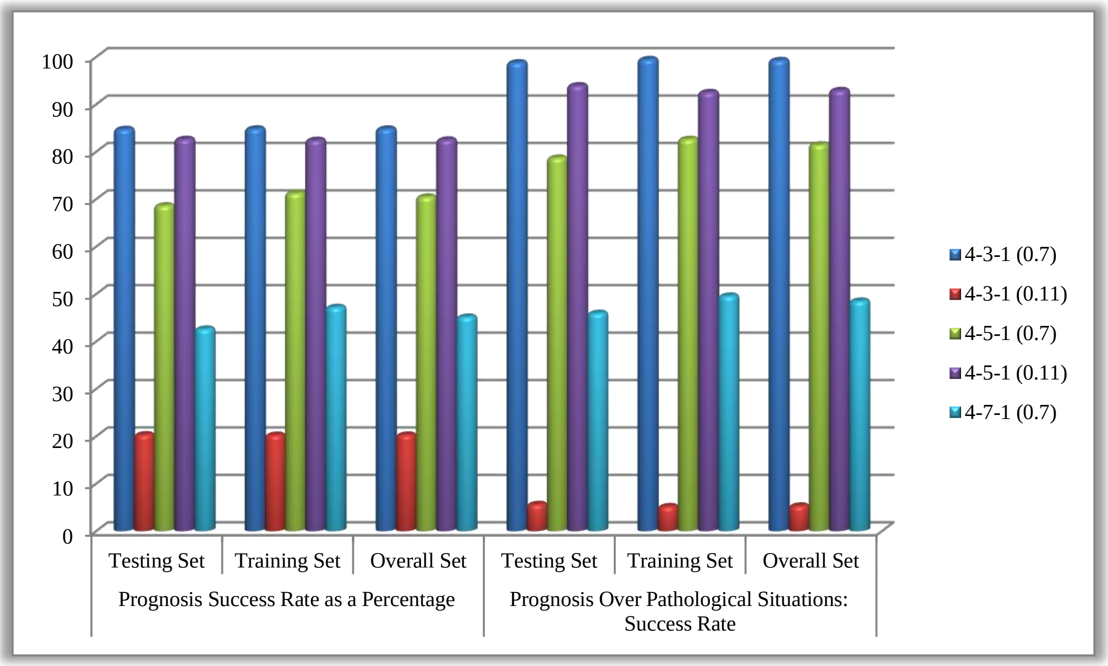


Figure: 5. Results of the Experimental Use of MLP for 2-Stage Osteoporosis Coding, presented graphically

This research examines the implementation of neural network models to consolidate them in clinical choice emotionally supporting networks. Consolidating various neural network types, geographies, and moving functions to create more viable neural networks is the driving force behind an MLP with a 4-5-1 geography and the use of a significant exchange function in the secret layer. In any case, the aforementioned MLP exhibits poorer execution as compared to other MLP arrangements. Furthermore, since this MLP is unable to distinguish between original ordinary and diseased patterns, it is evident that it has an overfitting problem.

The PNN discoveries are shown in Tables 4 and 5. Separately, the exchange functions for the yield and buried layers were compet and radbas. The number of neurons in PNNs' secret and feedback layers was constant, but the number in their result layers varied according to the coding of desired values. The spread of the outspread premise function, which is used in the next layer, is the primary boundary that can be changed. Because the outcome layer of PNNs relied on the four stages of osteoporosis risk factor prediction, it initially had four neurons.

Table: 4. Results Of Experiments Using PNN Architectures: 4 Neurons Make Up The Output Lay

Nr	Spread of Radius	Prognosis Success Rate as a Percentage			Prognosis Over Pathological Situations: Success Rate		
		Testing Set	Training Set	Overall Set	Testing Set	Training Set	Overall Set
1	0.7	31.42	99.06	75.07	33.76	94.12	76.44
2	0.8	31.42	99.06	75.07	33.76	94.12	76.44
3	3.9	35.82	59.18	53.35	39.43	62.60	55.81
4	4.8	37.22	51.64	47.43	41.69	55.21	51.24
5	8.5	41.32	42.65	42.26	47.12	49.23	48.61
6	14.3	39.22	40.76	40.31	46.94	47.86	47.30
7	20.4	38.52	39.15	38.97	45.23	46.30	45.98
8	41.2	33.12	33.30	33.24	38.96	39.28	39.23

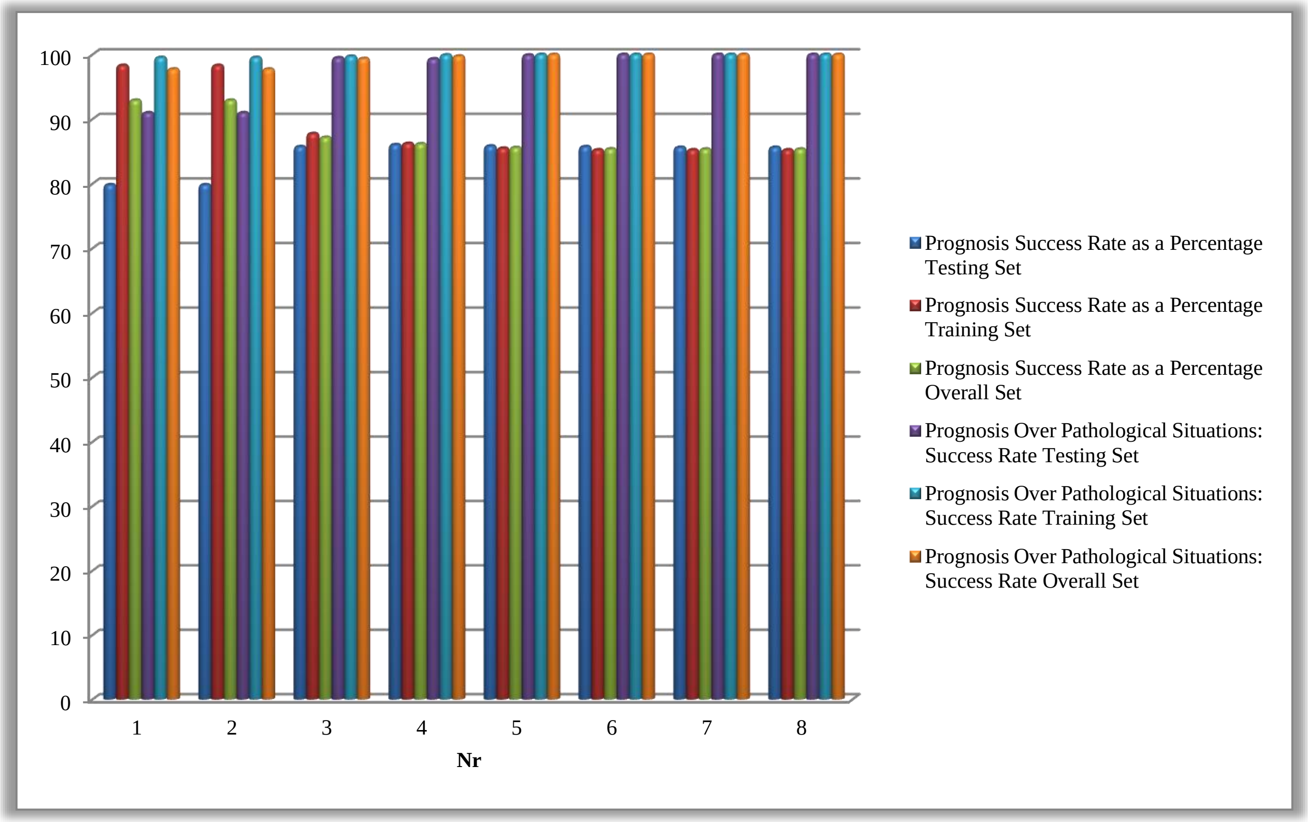


Figure: 6. Results of experiments using PNN architectures are graphically displayed:

Nr	Spread of Radius	Prognosis Success Rate as a Percentage			Prognosis Over Pathological Situations: Success Rate		
		Testing Set	Training Set	Overall Set	Testing Set	Training Set	Overall Set
1	0.7	79.82	98.30	92.91	90.97	99.53	97.73
2	0.8	79.82	98.30	92.91	90.97	99.53	97.73
3	3.9	85.72	87.75	87.16	99.48	99.72	99.36
4	4.8	86.02	86.22	86.16	99.31	99.92	99.74
5	8.5	85.82	85.48	85.58	99.90	100.00	99.99
6	14.3	85.72	85.23	85.37	100.00	100.00	100.00
7	20.4	85.62	85.23	85.35	100.00	100.00	100.00
8	41.2	85.62	85.23	85.35	100.00	100.00	100.00

The output layer is made up of 4 neurons.

Table: 5. Results of Experiments Using

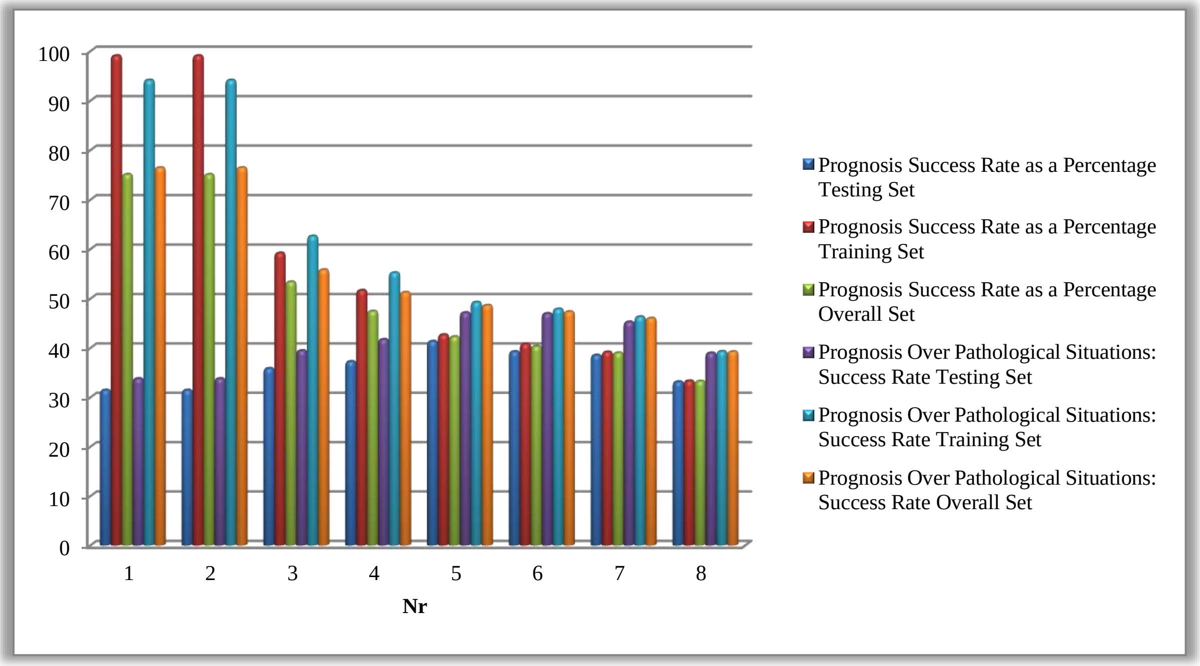


Figure: 7. Results of Experiments Using PNN Architectures Graphically Represented: 2 Neurons for the Output Layer

PNNs 1 and 2 in Tables 4 and 5 demonstrate that small variations in spiral premise spread have little effect on the display of PNNs. The results remained unchanged since it is impossible to detect a difference between 0.7 and 0.9. The value of the spread is increasing, indicating a decline in PNN's presentation in evaluating regular and compulsive events during the entire and training data collection. Number 3 PNN can, of course, describe more regular situations than number 4 PNN. As a result, Nr. 3 PNN outperforms Nr. 4 PNN in terms of speculative capacity. Instead of Nr. 4 PNN, Nr. 3 PNN successfully separates sick individuals from healthy examples; as a result, its prognostic range has increased. Still, its display has somewhat decreased.

ANNs are a powerful tool that can assist experts with various tasks and determination. ANNs provide a number of benefits in this way, like as:

- The ability to deal with a lot of information
- Less inclined to ignore significant data
- Shortening of the diagnosing system

V. CONCLUSION

In this paper, we provided an extensive analysis of the advancements in infectious disease diagnosis and identification that ANNs have made possible. Artificial neural networks (ANNs) have proven to be effective in accurately diagnosing a wide range of diseases. The advancement of outline and elaboration techniques has the potential to significantly contribute to the prompt, accurate, and efficient diagnosis of illnesses. Osteoporosis is a major clinical concern for overall health, which is why this work has investigated the possible application of artificial neural models in clinical navigation, and in particular, the assessment of osteoporosis risk factors. The primary driving forces for the intense interest in developing precise and practical treatments that shield patients from radiation exposure are its prevalence and the severe consequences it has on patients. PNNs and MLPs, which are feed-forward neural networks and have a back-proliferation computation, served as the foundation for the development of artificial neural methods. Because the internal relationships of MLPs are highly asymmetric and not amenable to conventional quantifiable analysis, they have been referred to as "secret elements". Instead, PNNs closely resemble the Bayesian function, but as it takes several stages to predict the risk variables for osteoporosis, their outcome is obviously not a likelihood. PNNs were found to outperform MLPs in terms of case visualization achievement pace. Therefore, the suggested approach revealed that PNN artificial models perform significantly better than MLP artificial networks, or, to put it another way, that PNNs' predictive capability is enhanced relative to MLPs' organizing execution.

VI. FUTURE SCOPE

Given that the primary focus of this work is on ANN-based disease detection models, it will be interesting to explore the ANN's actual potential for more cautious approaches and targeted drug delivery in the future. We also recommend evaluating various cutting-edge approaches, including as exchange learning and crossover plans, to further improve detecting exactness and accuracy in clinical determination.

REFERENCES

[1] D. Howard and B. Mark (2018), “Neural network toolbox User’s Guide”, The Math Works, Inc.

[2] L. Lanzarini, M. V. Camacho, A. Badran, and I. D. G. Armando, “Images compression for medical diagnosis using neural networks,” J. Computer Science and Technology, Vol.2, No. 1, pp. 78-80.

[3] D. Mantzaris, G. Anastassopoulos, A. Adamopouos, I. Stephanakis, K. Kambouri, and S. Gardikis, “Abdominal pain estimation in childhood based on artificial neural network classification,” in Proc. the 10th Int. Conf. on Engineering Applications of Neural Networks, Thessaloniki, 2017, pp. 129-134.

[4] T. Goel, R. Murugan, S. Mirjalili, and D. K. Chakrabartty, “Optconet: an optimized convolutional neural network for an automatic diagnosis of covid-19,” Applied Intelligence, vol. 51, no. 3, pp. 1351–1366, 2021

[5] L. Wang, Z. Q. Lin, and A. Wong, “Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images,” Scientific Reports, vol. 10, no. 1, pp. 1–12, 2020.

[6] S. Toraman, T. B. Alakus, and I. Turkoglu, “Convolutional capsnet: A novel artificial neural network approach to detect covid19 disease from x-ray images using capsule networks,” Chaos, Solitons & Fractals, vol. 140, p. 110122, 2020.

[7] Lee, Sue Han, et al. "How deep learning extracts and learns leaf features for plant classification." Pattern Recognition 71 (2017): 1-13.

[8] D. Gil, M. Johnsson, J. M. GariciaChemizo, A. S. Paya and D. R. Fernandez, "Application of Artificial Neural Networks in the Diagnosis of Urological Dysfunctions", Expert Systems with Applications, Vol.36, No.3, 2019, pp. 5754-5760.

[9] S. Altunay, Z. Telatar, O. Erogul and E. Aydur, "A New Approach to urinary system dynamics problems: Evaluation and classification of uroflowmeter signals using artificial neural networks", Expert Systems with Applications, Vol.36, No.3, 2019, pp. 4891-4895.

[10] S. Moein, S. A. Monadjemi and P. Moallem, "A Novel Fuzzy-Neural Based Medical Diagnosis System", International Journal of Biological & Medical Sciences, Vol.4, No.3, 2019, pp. 146-150.

[11] M. P. Francisco, G. C. Juan Manuel, S. P. Antonio and R. F. Daniel, "A robust model of the neuronal regulator of the lower urinary tract based on artificial neural networks", Neurocomputing, Vol.71, No.4-6, 2018, pp. 743-754.

[12] S. A. Monadjemi and P. Moallem, "Automatic Diagnosis of Particular Diseases Using a Fuzzy-Neural Approach", International Review on Computers & Software, Vol.3, No.4, 2018, pp. 406-411.

[13] M. A. Mazurowski, P. A. Habas, J. M. Zurada, J. Y. Lo, J. A. Baker and G. D. Tourassi, "Training Neural Network Classifiers for Medical Decision Making: The Effects of Imbalanced Datasets on Classification Performance", Special Issue on Advances in Neural Networks Research: IJCNN ’07, International Joint Conference on Neural Networks IJCNN’07 Orlando, Florida, USA, April, 2018, Vol.21, No.3, pp. 427-436,

[14] K. Higuchi, K. Sato, H. Makuuchi, A. Furuse, S. Takamoto and H. Takeda, "Automated diagnosis of heart disease in patients with heart murmurs: application of a neural network technique", Journal of Medical Engineering & Technology, Vol.30, No.2, 2017, pp. 61- 68.

[15] P. Saha, M. S. Sadi, and M. M. Islam, “Emcnet: Automated covid19 diagnosis from x-ray images using convolutional neural network and ensemble of machine learning classifiers,” Informatics in medicine unlocked, vol. 22, p. 100505, 2021.

The Heart Disease Prediction Using Technique of Classification in Machine Learning Using the Concepts of Data Mining

Ravi Agrawal
Research Scholar

Abstract--- heart disease is one of the main factors contributing to problems in life and, finally, death. Data mining (KDD) is the most used technique for information disclosure separation. One of the most important and fundamental methods for identifying important data, sifting through massive databases to unearth hidden examples, is data mining. Data mining and machine learning techniques are used in the clinical sciences to address real wellbeing-related issues by predicting and diagnosing various disorders. Machine learning computations and data mining methods are essential in the pharmaceutical industry. We use eight algorithms to measure cardiac illnesses: Guileless Bayes, KNN, Choice Tree, J48, Strategic Model Tree, Irregular Woodland, Backing Vector Machine, and Closest Neighbor. Prediction accuracy is flawless while employing a greater number of attributes. Using these data mining and machine learning computations, we hope to conduct a predictive analysis of cardiac diseases, analyze various mining and machine learning techniques, and determine which are the most effective.

Keywords--- Classification, Data mining, Heart disease, Machine learning, Prediction.

I. INTRODUCTION

The heart is one of the most vital human organs. It is a kind of muscular organ that pumps blood into the body and acts as the circulatory system's heart. All blood vessels, including arteries, veins, and capillaries, are part of the cardiovascular system. These blood vessels act as a complex network to circulate blood throughout the body [1].

Heart illness can exacerbate a number of serious conditions if the heart's normal blood flow is obstructed or otherwise abnormal. Often referred to as cardiovascular diseases (CVDs), these are arguably the worst diseases on the globe. The term "CVDs" refers to diseases affecting the heart, veins, and brain. According to the World Health Organization's (WHO) Worldwide Chart book on Cardiovascular Disease Anticipation and Control, CVDs are the leading cause of mortality and disability worldwide. Even while adopting a healthier lifestyle and doing other relevant actions can help prevent CVDs, all available data point to a continued daily rise in CVDs, as confirmed by multiple WHO examinations [2]. Heart disease refers to a range of conditions that affect the heart. The term "heart disease" refers to a variety of illnesses, such as intrinsic heart attacks, vein disorders such as coronary course disease, irregular heartbeats (arrhythmias), and more. In several instances, heart disease and cardiovascular disease are used interchangeably. The phrase "cardiovascular disease" (CVD) refers to a group of illnesses that include angina, strokes, and constricted or dilated veins that can result in dead tissue patches in the heart. Illnesses that affect the heart's muscles, valves, or emotions also represent heart illnesses. A total of 17.9 million people die from CVDs each year, accounting for 31% of all deaths [3].

The current medical services sector generates enormous amounts of data about patients, diseases, and other topics, but specialists and scientists are unable to effectively use this data. Currently, the medical care business is very concerned about the nature of administration (QoS). The type of administration suggests the appropriate diagnosis and course of treatment for the patient. Unfortunate findings can lead to unacceptable outcomes. Heart disease is associated with a variety of risk factors. Certain elements that influence gambling are unchangeable, such as age, country, family history, and handling business. However, several risks can be prevented or controlled, such as diabetes, smoking, high blood pressure, high cholesterol, inactivity, and being overweight or obese [4].

The most popular method for identifying previously unknown secret examples (knowledge) is data mining, which makes use of database structures, calculations, measurements, and machine learning techniques. The data that has been discovered can be used to create intelligent, dynamic vision frameworks for many endeavors, such as healthcare, to provide timely diagnosis and treatment to preserve vital life. Machine learning enables PC initiatives to benefit from determined data, improve execution from encounters, and then apply what they have learned to employ solid judgment without the need for human collaboration. With each wise choice, a machine learning programming's presentation improves [5].

We use the Waikato Environment for Knowledge Analysis (WEKA). Information from the UCI store is routinely given in a database or bookkeeping sheet. To be used with the WEKA apparatus, the data sets must be in the ATTRIBUTE-RELATED FILE FORMAT (ARFF). The WEKA device is used to pre-handle the dataset. Through careful examination of various decision tree computations within the WEKA apparatus and the subsequent generation of decisions, the framework will be able to predict the likelihood that the patient has heart disease, which will undoubtedly aid in early diagnosis and timely treatment [6].

II. RELATED WORK

Strike Alzubi et al.: In this review, a subset of single nucleotide polymorphisms (SNPs) has been identified using conditional mutual information boost (CMIM). These SNPs are then used in conjunction with the group technique for analysis in various machine learning computations, such as KNN, LDA, SVM, NB, and ANN. The SVM, 5-NN, and NB classifier troupe technique produced the highest notable precision (93.21%) and f1-score (91.27%) among the previously described computations. The dataset's source is the 2001 instances of the Welcome Trust Case Control Consortium (WTCCC) [7].

A model that combines KNN with Hereditary Calculation was suggested in the publication by M. Akhil Jabbar et al. The UCI machine learning store provided the dataset, which has five attributes and 14 occurrences. While real success was shown for necessary growths and bosom illness, the recommended model performed poorly overall due to excess and immaterial features included in the dataset. The evaluation reveals a 60% KNN get over rate for hereditary calculations, indicating an improvement in the accuracy of heart disease prediction [8].

Gandhi Monika, among others The main objective of this project is to automate a system for identifying heart illness by taking historical data and information into account. It aims to illustrate some data mining-based knowledge reflection systems, along with their benefits and drawbacks. In data mining, the Decision Tree Calculation, Brain Organization, and Credulous Bayes computations are employed. The ID3 Calculation uses Shanon Entropy to create decision trees. The data is streamed using Quinlan's C4.5 and C5.O computations because to the limitations of trying to create a short tree decision tree using set off learning data. Nave Bayes performs the best among the previously stated algorithms because of its attribute freedom. Different calculations can then be investigated and applied to address problems such as overfitting, lengthy handling times, improved preparation, and many other concerns [9].

The main goal of this investigation, according to Salam Ismaeel et al., is to replace the costly clinical evaluation with a pre-notification system for people who are predisposed to develop cardiovascular problems. The dataset, comprising 300 examples of continuous data collected by the Cut Land Clinical Establishment, is presented through an amazing machine learning process. ANN is less accurate than Outrageous Machine Learning (87%), which is the better of the two machine learning computations used in the test. Applying this model makes it possible to examine a different patient using data from the past. The Outrageous machine learning computation has five (0-4) outcomes. Eventually, by developing the model and determining how features affect the ultimate decision, the problem of missing characteristics can be resolved [10].

Using data mining techniques including neural networks, decision trees, and credulous bayes, an insightful heart disease prediction framework model was created. The findings reveal that every process occasionally succeeds in achieving the mining objectives that have been specified. IHDPS can provide confused "what if" answers that traditional decision emotionally supporting networks are unable to, as demonstrated by Sellappan Palaniappan et al. The outcomes demonstrated each technique's unique strength in understanding the motivation behind the mining targets that were defined. IHDPS was able to respond to demands that traditional decision-making emotional support networks were unable to handle. Important information, such as linkages and examples within clinical boundaries relevant to heart disease, was easier to incorporate. IHDPS continues to be a secure, user-friendly, reliable, adaptable, and scalable web framework [11].

In 2017, Zeinab Arabasadi and colleagues suggested a cross-breed diagnosis model for coronary artery disease utilizing artificial neural network (ANN) and hereditary calculation. Only 22 urgent features were considered in the trial, out of the 303 patient records with 54 highlights in the Z-Alizadeh Sani dataset used in this investigation, 216 of whom had coronary artery disease (Cad). Hereditary calculation was used to first determine the artificial neural network (ANN) weights, and then preparation data was used to create the ANN model. An ANN with one information layer, one result layer, and one secret layer with five neurons is used in this trial's feed-forward setup.

The 10-crease cross approval approach is used in this inquiry to assess the framework. The results showed that, in terms of exactness, our suggested model performed better than the current simple ANN model. Four more well-realized heart disease data sets are used to test our model, and the results are examined [12].

In order to determine the most reliable method, N. Shirwalkar and T. Tak led a logical focus in 2018 and investigated a few data mining and machine learning techniques used in heart disease prediction. The suggested prediction for heart disease makes use of innocent Bayes and K-implies work. 303 records of individuals with cardiac illness were extracted from the Cleveland database maintained by the UCI repository. The technique used to change the first dataset table from one form to the next is called discretization. A better version of the k-implies method is used to create groups from the dataset table. To predict people with heart disease, the model is created using the Credulous Bayes technique [13].

III. PROPOSED APPROACH

A. Proposed System

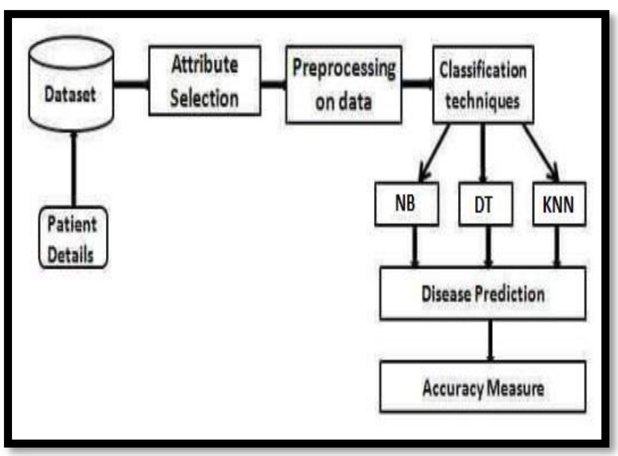


Figure: 1. Block diagram of proposed system

B. This system's activity is separated bit by bit:

- A variety of datasets containing patient information.
- The most popular method of selecting traits selects those that are helpful in predicting heart disease.

- After that, the available data assets are further selected, cleaned up, and formatted as needed.
- Using preprocessed data, a few classification techniques will be used to accurately predict cardiovascular disease.
- A few classifiers' precision is differentiated by the exactness metric.

C. Heart Disease

Heart and vascular diseases are implied by the term "heart infirmity." Heart disease is a general term that encompasses a variety of conditions affecting different parts of the heart. Since the word "cardio" means "heart," all cardiovascular ailments are included in this category [14].

- **Symptoms**

The most frequent adverse impact of coronary artery disease is angina, or chest pain. Chest pain or discomfort, often known as angina, can be described as heavy, constricted, pounding, copying, full, pressing, or severe. It could very well be confused for acid reflux or heartburn. Angina can also affect the neck, throat, jaw, shoulders, arms, or back.

- **Types of Heart Diseases**

There are numerous varieties of heart disease, and they all impact the organ differently and present in different ways.

- Coronary artery disease (Cad) is the most common type of heart disease. The veins that provide blood to the heart muscle are lined with plaque, which is made up of materials like fat and cholesterol, in coronary artery disease (CAD). Plaque development, also known as atherosclerosis, causes the courses to choke by giving the heart muscle less oxygen than necessary for it to function normally. Heart attacks or angina (chest pain) can occur when the heart muscle does not receive enough oxygen.
- An arrhythmia is a haphazard or irregular heartbeat. This may manifest as irregular heartbeats, fast heartbeats (tachycardia), or slow heartbeats (brachycardia). The most common arrhythmias are likely bradyarrhythmias, which are slow heart rhythms brought on by circumstances affecting the heart's conduction system, untimely ventricular compressions,

which are extra pulsates that originate from the ventricles, and atrial fibrillation, which occurs when the atria, or upper heart chambers, contract erratically.

- Congestive heart failure, or CHF, is the medical term for the situation where the heart is unable to pump enough oxygen-rich blood to meet the body's needs. This could be due to a lack of blood flow or the heart's inability to siphon off enough blood. A few people struggle with both problems.
- Heart valve disease develops when at least one of the heart's four valves malfunctions. Heart valves aid in allowing blood to continue moving forward as the heart pumps it out. It is testing for conditions such as stenosis and mitral valve prolapse, which affect the heart valves.
- Cardiomyopathy—a disease affecting the heart's muscles—can cause the heart to grow or acquire thick heart walls. This reduces the heart's readiness to pump blood throughout the body, which frequently results in heart disappointment.
- About one in every 100 live births are affected by a type of birth abnormality called intrinsic heart disease, which causes problems with the heart from the moment the baby is born. Two of the most prevalent types of intrinsic heart disease are ventricular septal deformities (VSD) and atrial septal defects (ASD), which result from incomplete closure of the walls separating the right and left chambers of the heart.

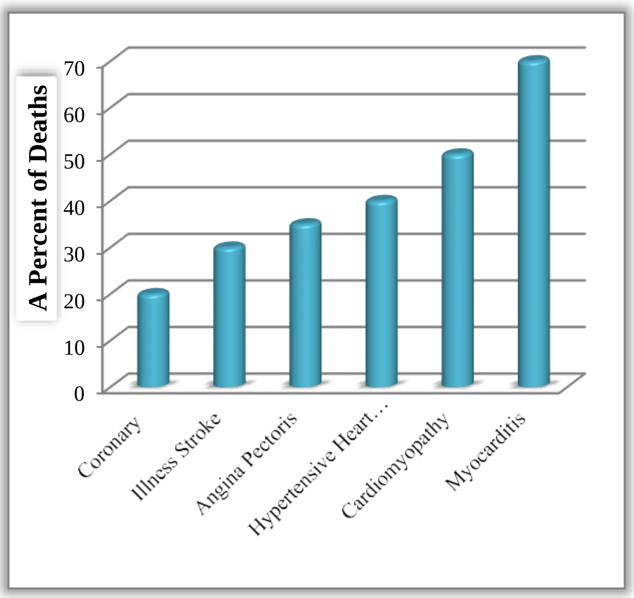


Figure: 2. Deaths % from various heart disease causes

D. Tool Used

Data can be broken down using a variety of data mining techniques and tools such as WEKA Device (Waikato Environment for Knowledge Analysis). WEKA is created with Java. Algorithmic examination of data sets stored in the arff format is made possible by WEKA. This review applies the Decision Tree, Guileless Bayes, and K-NN computations to the cardiovascular data set and shows the results.

E. Decision Tree

At each hub, the best portion is chosen to create the decision tree. To select the best trait for the split, the information gain is calculated at each hub and the characteristics are arranged accordingly. Purchase the Proportion Attribute In this instance, the attribute evaluator is Eval, and the hunt mechanism is the WEKA Device's Ranker operation. A summary of the positional attributes is provided in Table 1.

Table: 1. Information-based attribute ranking

Info gain	Rank	Attribute
0.19	14	thal
0.18	15	ca
0.17	11	exang
0.15	10	thalach
0.13	5	cp
0.12	12	oldpeak
0.011	13	slope
0.067	4	sex
0.062	3	age
0.024	9	restecg
0	8	fbs
0	7	chol
0	6	trestbps

Table: 2. Decision tree algorithm results

	No of instances	Percentage (%)
Correctly classified instances	299	94.0994
Incorrectly classified instances	26	9.9410
Total instances	325	

The following are the qualities: 12, 13, 9, 8, 3, 10, 11, 2, 1, 7, 6, 5, 4. The qualities are chosen in a particular sequence. The decision tree is then created by subjecting the cardiovascular data set to the Decision Tree calculation J48. With this option tree, a prediction can be made. The findings are shown in Table 2.

F. Naïve Bayes

Get Proportion Attribute Acquisition The Ranker technique is the hunt process, and Eval is the attribute evaluator that is used. The decision tree makes use of traits that are positioned similarly. The cardiovascular data set is subjected to the Gullible Bayes algorithm, and Table 3 shows the outcomes of several features.

The outcomes are displayed in Table 4.

Table: 3. Results of utilizing the Naive Bayes algorithm for a few attributes

Attribute	**<50 (0.54)**	**>50_1 (0.45)**	**>50_2 (0)**	**>50_3 (0)**	**>50_4 (0)**
cp					

typ_angina	19.0	10.0	3.0	3.0	3.0
asymp	42.0	127.0	3.0	3.0	3.0
non_anginal	72.0	21.0	3.0	3.0	3.0
atyp_angina	44.0	12.0	3.0	3.0	3.0
[total]	171.0	164.0	6.0	6.0	6.0
Restecg					
left_vent_hyper	71.0	82.0	3.0	3.0	3.0
Normal	99.0	59.0	3.0	3.0	3.0
st_t_wave_abnormality	4.0	6.0	3.0	3.0	3.0
[total]	189.0	163.0	5.0	5.0	5.0

Table: 4. Results of the Naive Bayes method

	No of instances	**Percentage (%)**
Correctly classified instances	277	86.1786
Incorrectly classified instances	50	17.8618
Total instances	325	

G. K-Nearest Neighbor

The KNN algorithm is applied to the cardiac data set, and Table 5 shows the results.

Table: 5. K-nearest neighbor method results

	No of instances	**Percentage (%)**
Correctly classified instances	325	100
Incorrectly classified instances	0	0
Total instances	325	

IV. EXPERIMENTAL RESULT

These findings show that, although most analysts use different classifier techniques, such as KNN, neural networks, and paired discretization with Gain Proportion Decision Trees, to analyze heart disease, using Nave Bayes and Decision trees with information gain estimations produces better outcomes and more notable precision than using different classifiers. We conjecture that the improved properties are what drove the increase in accuracy. Furthermore, decision trees have been shown to outperform Credulous Bayes. The decision tree classifier performs even more precisely than the Guileless Bayes classifier.

The evaluation parameters used include accuracy, specificity, and sensitivity.

(i) Sensitivity = TP/P
(ii) Specificity = TN/N
(iii) Accuracy = (TP + TN)/(P + N)

Where P and T stand for actual positives and negatives, respectively, and TP stands for true positives. For a convincing indicator, high exactness, low particularity, and high awareness are anticipated. Table 6 summarizes the ways in which these activities and the three prediction techniques are compared.

Table: 6. An overview of prediction methods and their results

	Prediction technique	Sensitivity	Specificity Accuracy
Decision tree	0.943	0.087	0.944
Naïve bayes	0.864	0.187	0.864
KNN	3	2	3

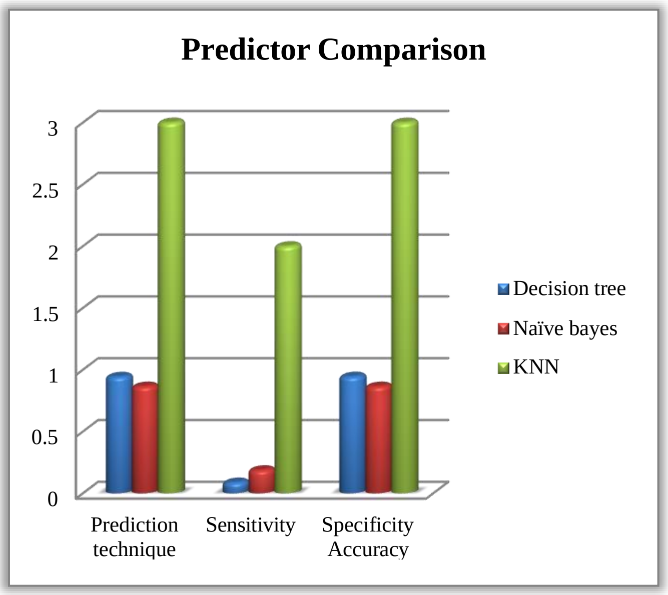


Figure: 3. Comparison of prediction approaches represented graphically

The tests, which are finished using the WEKA device, use the computations on the cardiac dataset. According to the curve in Figure 3, specificity is low while accuracy and responsiveness are high. As a result, when applied functionally, the signs are strong. Based on its consistent ability to be remembered every time, the results show that KNN has the most remarkable accuracy in terms of display development. For the provided cardiovascular dataset, the Decision Tree technique performs better than the other two when it comes to prediction.

In order to identify the most reliable computation for predicting the likelihood that a patient would develop heart disease, this review evaluates the practicality of several classification formulas. Here is a comparison of the results of the various classification calculations along with a table that shows their precision ratings.

Table: 7. Comparison of Classification Algorithms

Algorithm	Accuracy
Logistic Regression	77.43%
Naïve Bayes	79.07%
Support Vector Machine	75.79%
K-Nearest Neighbor	59.85%
Decision Tree	79.07%
Random Forest	88.91%
XGBoost	80.71%

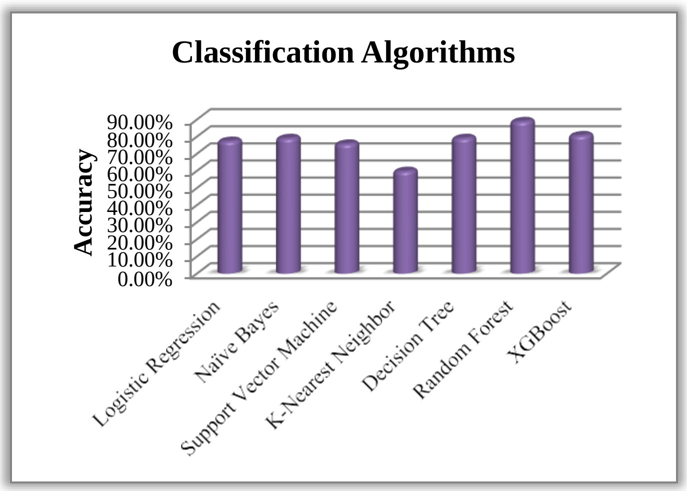


Figure: 4. Comparison of classification algorithms represented graphically

XG Boost (80.71%) and Arbitrary Forest (88.91%) have the best calculations. However, K-Closest Neighbor, who scored 59.85%, displayed the most glaringly poor exactness. Additional data mining techniques, such as time series, bunching and affiliation criteria, and other group techniques, could be included to expand this analysis.

V. CONCLUSION

To determine the predictive accuracy of several classifiers, we conducted a trial. Based on their subjective performance, we select four well-realized classifiers for the analysis. When examining the classification performance of four machine learning computations, it is generally assumed that the Nave base classifier performs better than the Decision Tree and K-Closest Neighbor, as demonstrated by trial results in Table 6. To analyze classification performance, classifiers are applied to similar data, and the misclassification and correct classification rates of the results are examined. After looking at the quantitative data that the programming experiences provide, it is also found that their performance is almost exactly the same, with very little difference. Further preliminary work on many other datasets should be taken into consideration in order to get a more comprehensive conclusion regarding the overall viability of the classifiers. As evidenced by the writing survey, which led us to conclude that the creation of predictive models for patients with heart disease is a negligible accomplishment, a combinational and more complex model is thus anticipated to improve on the accuracy of predicting the early stage of heart disease.

VI. FUTURE SCOPE

There is a lot of work that is expected of you in this course. Future attribute choice, attribute decrease, and line decrease have all been completed in accordance with business standards. Subsequent research on this topic will concentrate on using various datasets to construct the prediction model in order to raise the exactness rate from its current level to an exceptional precision rate.

REFERENCES

[1] Tan, P.N., Steinbach, M., Kumar, V.: Introduction to Data Mining, 4th edn. Pearson Publications, Boston.

[2] Ferdousy, E.Z., Islam, M.M, Matin, M.A.: Combination of Naïve Bayes classifier and Knearest neighbor in the classification based predictive models. J. Comput. Inf. Sci. 6(3), 48–56. ISSN: 1913-8989

[3] Suneetha, N., Hari, V.M.K., Kumar, V.S.: Modified gini index classification: a case study of heart disease dataset. Int. J. Comput. Sci. Eng. 2(6), Patka, S., et al.: Recent trends and rapid development of applications in data mining. IOSR J. Comput. Sci. (IOSR-JCE) 73–78. e-ISSN: 2278-0661, p-ISSN: 2278-8727

[4] M.A.Nishara Banu, B.Gomathy, "Disease Forecasting System Using Data Mining Methods," International Conference on Intelligent Computing Applications

[5] Feixiang Huang, Shengyong Wang, and ChienChung Chan, "Predicting Disease By Using Data Mining Based on Healthcare Information System", IEEE International Conference on Granular Computing.

[6] Zeinab Arabasadi et al. , " Computer aided decision making for heart disease detection using hybrid neural network- Genetic algorithm" , Computer Methods and Programs in Biomedicine- ELSEVIER, Vol. 141, pp.19- 26, April- 2017.

[7] R. Alzubi, N. Ramzan, H. Alzoubi, and S. Katsigiannis, "SNPs-based hypertension disease detection via machine learning techniques," ICAC 2018 - 2018 24th IEEE Int. Conf. Autom. Comput. Improv. Product. through Autom. Comput., pp. 1–6, 2018.

[8] H. Sharma et al., "A prediction of heart disease using machine learning algorithms," Adv. Intell. Syst. Comput., vol. 7, no. 1, pp. 497–504, 2018, doi: 10.1007/978-3-030-51859-2.

[9] M. Gandhi and S. N. Singh, "Predictions in heart disease using techniques of data mining," 2019 1st Int. Conf. Futur. Trends Comput. Anal. Knowl. Manag. ABLAZE 2019, pp. 520–525, 2019, doi: 10.1109/ABLAZE.2018.7154917.

[10] S. Ismaeel, A. Miri, and D. Chourishi, "Using the Extreme Learning Machine (ELM) technique for heart disease diagnosis," 2019 IEEE Canada Int. Humanit. Technol. Conf. IHTC 2015, no. 1, pp. 1–3, doi: 10.1109/IHTC.2015.7238043.

[11] SellappanPalaniappan and Rafiah Awang, "Intelligent heart disease prediction system using data mining techniques", International Journal of Computer Science and Network Security, vol.8, no.8, pp. 343-350,2018.

[12] Zeinab Arabasadi et al. , " Computer aided decision making for heart disease detection using hybrid neural network- Genetic algorithm" , Computer Methods and Programs in Biomedicine-ELSEVIER, Vol. 141, pp.19- 26, April- 2017.

[13] Nikita Shirwalkar, and Tushar Tak, " Human Heart Disease Prediction System Using Data Mining Techniques", International Journal of Innovations & Advancement in Computer Science, Vol. 7, Issue 3, Mar.2018.

[14] AnimeshHazra, S.Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Advances in Computational Sciences and Technology, Vol. 10, No.7, July-2017.

Principles and Advances in Deep Reinforcement Learning for Information Extraction

Anushka Garg

Research Scholar

***Abstract*—** Data recovery (IR) techniques are essential to reducing the problem of data over-burden. These tactics include pursuit, proposal, and web-based advertising, which meet clients' data needs by introducing personalized objects (data or services) with perfect timing and place. The basic ideas and intuitions behind brain-in-the-rest (IR) models are introduced in this course and contextualized within the framework of earlier non-brain approaches to IR. First, we will review important concepts in recovery as well as a variety of brain and non-brain methods for individual message vector portrayal learning. Consequently, a fundamental AI worldview for assisting learning-based intelligent IR has emerged: benefiting from cooperation. Deep reinforcement learning (DRL) has gained notoriety recently due to continuous improvements. Therefore, we suggest a novel promoting strategy for the rec/promotions trade off in this study. Trial findings in the context of verified data demonstrate the suggested system's viability. Finally, we discuss potential future directions for neuronal IR.

***Keywords* —** *Advertising Method, Deep Reinforcement Learning, Information Retrieval (IR), non-neural techniques.*

I. INTRODUCTION

Massive amounts of information have been created as a result of the Internet's recent incredible advancement and increasing ubiquity, which has led to an unquestionably critical data over-burden problem. Thus, it has become increasingly important to understand how to segregate data (things or administrations) that meet clients' needs for data in a legal general framework. This promotes a wide range of data recovery techniques, including web-based marketing, suggestion, and search. Because of the renown of deep learning breakthroughs and the continuous advancements in support learning theories, profound learning techniques like Go, computer games, and mechanical technology stand out sufficiently to be observed by many people [1]. DRL is appropriate for not only continuing to observe and determine how to behave by integrating deep learning into support learning, but also for adding social event problematic examples with the strength of profound learning [2]. The DRL framework is used to solve difficult problems by developing expertise through teamwork in a strong environment [3]. Therefore, in the event that clear instructions are lacking, an optimal strategy may provide dynamic solutions for intricate problems. Consequently, we will host this studio to enable academic scientists and industry professionals to analyze the fundamental guidelines, specific and practical needs, and insights and illustrations obtained from DRL implementations for (intelligent) data recovery. Furthermore, we would like to promote research on innovative methods and algorithms for information retrieval as well as DRL applications **[4]**.

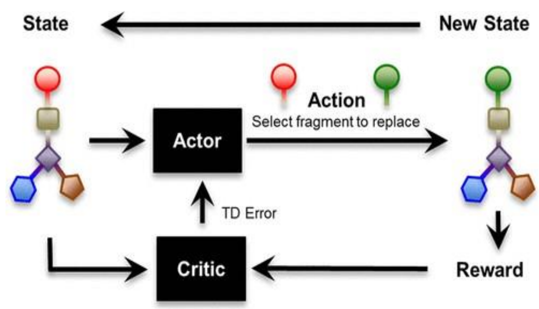


Fig. 1 Training using Deep Reinforcement

Supportive science programs use fundamental planning and heightening techniques to safeguard areas of strength. This multiparameter smoothing out procedure is laborious, intricate, and uncomfortable; it integrates a few properties with balanced designs as much as possible [5]. Additionally, new techniques for the automatic matching of mixes to profiles of different attributes are quite valuable. In this section, we introduce a performance savant model-based section set up help learning strategy for the era of unique compounds with perfect features.

A. *Reinforcement Learning Models*

Generally speaking, there are two types of support realizing:

- **Positive Reinforcement**
- **Negative Reinforcement**

✓ **Positive Reinforcement:**

Reassuring input learning entails providing something to increase the likelihood that usualbehavior will persist. It increases the conduct's feasibility and has a definitive effect on the expert'sbehavior [6].

This kind of assistance may promote advancements for a considerable amount of time, yet too much constructive criticism could lead to too many states, which could diminish outcomes.

✓ **Negative Reinforcement:**

unfavourable support learning works in opposition to positive feedback because it increases the likelihood that the specific behavior will occur in the future by avoiding the unfavourable situation.

A. Reinforcement Learning Applications

- **Robotics:** Robot route, robot soccer, walking, shuffling, and other tasks all use reinforcement learning.
- **Control:** A helicopter pilot is an example of support learning. RL can be used for diverse control, such as manufacturing plant processes and affirmation control in telecom.
- **Game Playing:** RL can be used when playing games like chess, spasm tac toe, and others.
- **Chemistry:** It is possible to advance the complex responses with RL.
- **Business:** Currently, business system arrangement uses RL.
- **Manufacturing:** Some auto manufacturers use sophisticated artificial intelligence to teach robots how to choose products and place them in designated areas.
- **Finance Sector:** Currently, the money domain uses the RL to evaluate exchange systems **[7]**.

Fig.2 Support Learning Applications

II. RELATED WORKS

(Richard S 2018) When we think of learning, the concept that we progress by interacting with our existing situation probably comes to us quickly. A newborn has no clear teacher when it plays, waves its arms, or looks about, but it has a direct sensorimotor relationship with its environment at that moment **[8].**

(Chen, M.; Beutel 2018) Many items to propose are managed by modern recommender systems across extraordinarily broad activity domains. In addition, they have a complex client state space due to the billions of unique clients they service every day. Making sense of the logged criticism is in any case subject to biases resulting from just taking note of comments made on recommendations selected by the recommender's previous incarnations. In this study, we provide a general recipe for managing these predispositions in a YouTube creation top-K recommender framework, using, for instance, a strategy slope-based computation **[9].**

(Li He, Liang 2018) Supported search is an essential strategy and a major source of revenue for almost all online search engines. From the promoters' perspective, aligning their business goal with the positioning of the list items through the purchase of the supported pursuit advertisement attracts greater mindfulness and purchases. From the perspective of the customers, offering tailored advertisements that correspond with their preferences will improve their online search experience **[10].**

(Nachum, O.; Gu 2018) Open-space exchange age is a challenging problem: maximum likelihood training can lead to gloomy outcomes, models struggle to meet long-term conversational goals, and training on standard film or internet datasets can lead to the age of incorrect, biased, or hostile text. One robust technique that could potentially address these problems is support learning (RL), which allows an exchange model to adjust for decreasing redundancy and poisonousness. This work presents a novel approach to progressive support learning, called VHRL, that tunes the expression level implanting of a variation grouping model using strategy slopes **[11].**

(Rohde, D.; Bonner 2018) Recommender frameworks are becoming commonplace in a variety of contexts and take different forms, such as item recommendations in online commercial establishments, search engine query suggestions, and companion recommendations in unofficial groups. Many experts report an error between improvements in disconnected measurements for managed learning and the web-based execution of the recently proposed models. Flow research headings, which are typically based on directed learning from verifiable information, appear to be displaying inevitable losses. A plausible argument could be that we are applying an unsuitable worldview: looking at the methodical process of obtaining authentic execution data, creating a second version of the suggested model, A/B testing it, and then implementing it **[12].**

(Feng, J.; Li, H.; Huang 2018) This study aims to provide an optimal and comprehensive overview of the new patterns of profound support learning in recommender frameworks, taking into account the emergence of profound support learning (DRL) in recommender frameworks research and some recent fruitful outcomes. We begin by drawing inspiration from the use of DRL in recommender systems. Next, we provide an overview of current methods and a scientific classification of DRL-based recommender frameworks. This overview identifies important open doors for further research and serves as a starting point for readings from the academic community and industry **[13].**

III. METHODS

In this section, we will provide a deep reinforcement learning framework for web-based recommendation in recommender systems. To be more precise, we will first suggest an ingenious DQN design that could simultaneously address the three tasks listed above. After that, we look at how to get the system ready using disconnected clients' behavior logs.

A. Online Advertising DQN Architecture

According to a recent statement, the web-based advertising problem in recommender systems is problematic because: I the promoting specialist's (AA) activity is complex, involving three sub-activities, specifically deciding whether to add an advertisement to the current rec-list, if so, which promotion is best, and where to put it; and (ii) the three sub-activities are interconnected, specifically when the AA decides to add a promotion, the up-and-coming promotions and To address these challenges, we introduce a unique deep Q-network design deep assist learning architecture. The handling of state and activity viewpoints is then depicted before the suggested DQN engineering with an enhancement technique is shown **[14].**

A. State and Action Feature Processing

The requirement includes a client's progress and accomplishments, a review of their past, logical data, and a summary of their most recent requests. A collection of suggestions (or advertisements) that the customer has read makes up the proposal (or advancement) reading history. Two RNNs with Gated Redundant Units (GRU) are used to identify the successive inclinations of clients for advances and offers.

The connected pieces of L proposed items that will be introduced in the ongoing request are then addressed in the ongoing sales relist, and they are transformed into a low-layered thick vector, represented as rect = tanh (Wrecconcat (rec1, recL) + brec). It should be noted that other models, such as CNN for NLP, may also be used. Finally, by combining, we have a low-level representation of state st:

$$s_t = concat\ (P_t^{rec}, P_t^{ad}, c_t\ and\ rec_t) \qquad \ldots 1$$

A. Problem in Applying DRL

While working on a robot that is controlled by a pi camera and an ultrasonic sensor, the viability of the dimensionality bother problem is identified. The information provided by non-visionary sensors, such as ultrasonic sensors, may generate the loud value that astounds the robot's actions. These boisterous characteristics resemble instances where dimensionality is irritated. In the unlikely event that the robot has a significant number of these kinds of sensors installed, even little disturbances will have a significant impact. The following challenges could arise when directly using DRL on the robot (equipment) to resolve such problems.

B. DQN Architectural Design

AA's action, given the state st, consists of three sub-exercises: (i) where is the optimal district, (ii) whether to embed an advancement, assuming this is true, and (iii) anything advancement is fantastic.

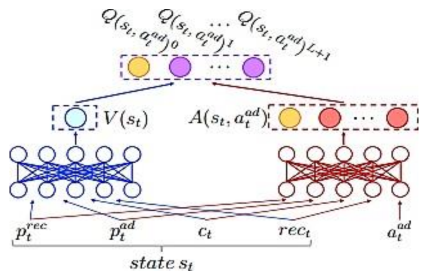


Fig .3 The nitty gritty design of the proposed DQN

First, we address the sub-activities (ii) and (iii) simultaneously. Ultimately, our goal is to determine the Q-upsides of every possible location and lock down for a random competitive promotion or advertisement. as seen in Figure 2. The state's representations and any application promotions or advertisements are the data sources, and the activity esteem (Q-esteem) in comparison to L + 1 areas is the outcome.

Algorithm 1 Off-policy Training of DEAR Framework.

```
Initialize the capacity of replay buffer D
Initialize action-value function Q with random weights
for session = 1, M do
    Initialize state s_0 from previous sessions
    for t = 1, T do
        Observe state s_t = concat(p_t^rec, p_t^ad, c_t, rec_t)
        Execute action a_t following off-policy b(s_t)
        Calculate reward r_t = r_t^ad + α r_t^ex from offline log
        Update state s_t to s_{t+1}
        Store transition (s_t, a_t, r_t, s_{t+1}) into D
        Sample mini-batch of transitions (s, a, r, s') from D
        Set
            y = { r                          terminal s'
                { r + γ max_a' Q(s', a'; θ)  non-terminal s'
        Minimize (y − Q(s, a; θ))^2 according to Eq.(6)
    end for
end for
```

IV. RESULTS AND DISCUSSION

In this section, we conduct extensive analyses on a real short video website to evaluate the feasibility of the suggested solution. The main questions we focus on are (I) how the DEAR performs in comparison to delegate baselines; (ii) how the system components enhance the presentation; and (iii) what the hyper-boundaries imply for the display. First, we introduce trial configurations. Next, we search for answers to these questions.

A. Dataset

We use the Walk 1–30, 2019 dataset, which was collected on the short-form video website Doujins, to train our model. Standard accounts (recommended things) and limited time accounts (pitched things) are the two types of accounts. The following components are typically included with a typical video: id, equivalent to score, finish score, remark score, follow score, and assembling score, with scores that are normal according to stage. The fact that (I) the standard characteristics are efficiently used in various applications, such as idea and publicity in the stage, (ii) we discretize each component as a one-hot vector, and (iii) baselines use similar features for an equitable evaluation, are all highly crucial. In response to global demand, we aggregate 1,000,000 gatherings, using the first 70% as a planning/endorsement set and the remaining 30% as a test set **[15].**

B. Implementation Details

The replay cushion size is set to 10,000 and the limited element is set to 0.95. For instance, we select the proposed system's hyper-boundaries using cross-approve.

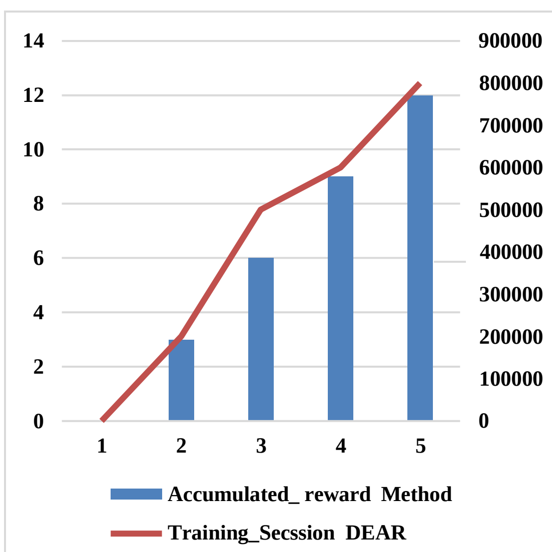


Fig. 4 Model execution with preparing information

Generally speaking, Execution Examination Table 1 shows the exhibitions for the general public. We highlight the objective facts that go with it: Figure 4 shows the suggested model's preparation cycle.

Table 1: In general execution correlation

Method	Reward	Improvement	P-value
W&D	2.66	56.3	0.000
DFM	26.3	46.3	0.000
GRU	25.3	53.6	0.000
HDQN	46.3	25.6	-
DEAR	5.69	-	0.002

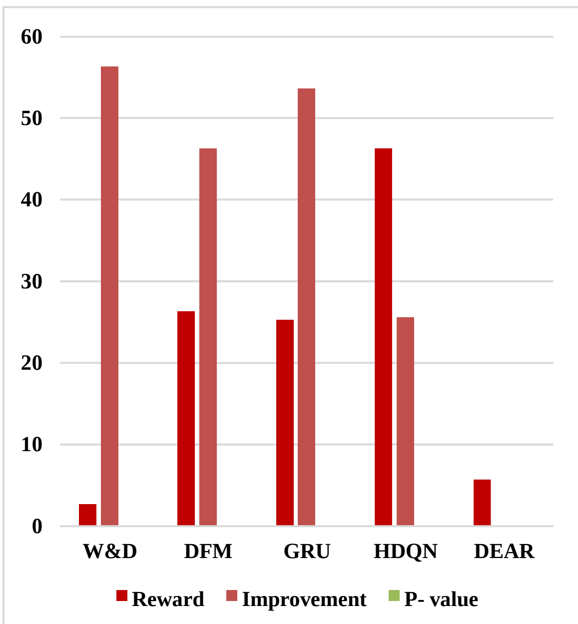


Fig .5 By and large execution examination

C. Component Study

In order to address the following question, we systematically eliminate the related components of DEAR by describing the corresponding variants: DEAR-1: These alternative uses directed learning to prepare the system and shares structures with the proposed model. DEAR 2 in addition to other.

Table 2: Part concentrate on results

Variant	Reward	Improvement	p-value
DEAR-1	6.33	12.1%	0.000
DEAR-2	2.365	2.36%	0.1000
DEAR-3	2.16	4.36%	0.002
DEAR-4	0.236	46.3%	0.2100
DEAR-5	2.369	-	-

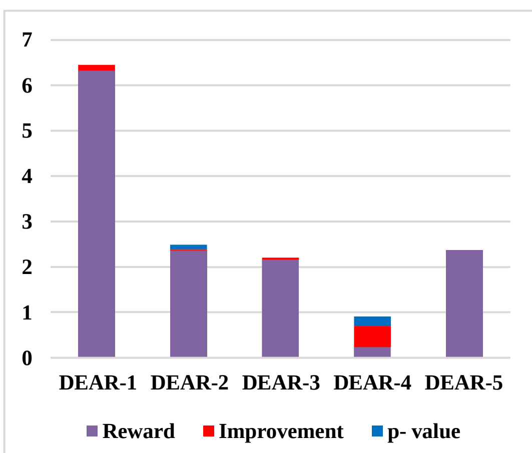


Fig. 6 Part concentrate on results

V. CONCLUSION

For electronic publicizing in recommender systems, we provide in this review a deep learning framework DEAR with a unique Significant Q-network design. It could simultaneously decide on three interconnected tasks, such as whether to remember an advancement for a rec-list and, if so, which advancement and region to include; (ii) increase the income of advancements while reducing the negative impact of advancements on the customer experience. In light of a quick video site, we evaluate our method using extensive research. Results show how our structure may effectively advance web-based promoting implementation in recommender systems.

VI. FUTURE WORKS

In closing, it is worthwhile to revisit the overarching goal of this entire investigation: the creation of broadly applicable artificial intelligence frameworks that are able to interact with and learn from their environment. Linking to the climate is both an advantage and a disadvantage of real-time learning. Although there are numerous challenges in attempting to understand our intricate and dynamic reality, RL gives us the freedom to choose how we do our research.

VII. REFERENCES

[1] Youhai Tan, Lingxue Dai, Weifeng Huang, Yinfeng Guo, Shuangjia Zheng, Jinping Lei, Hongming Chen, Yuedong Yang. DRlinker: Deep Reinforcement Learning for Optimization in Fragment Linking Design. Journal of Chemical Information and Modeling 2022, 62 (23) , 5907-5917.

[2] Isaac M. Jackson, E. William Webb, Peter J.H. Scott, Michelle L. James. In Silico Approaches for Addressing Challenges in CNS Radiopharmaceutical Design. ACS Chemical Neuroscience 2022, 13 (12) , 1675-1683.

[3] Jon Paul Janet . Data-Driven Mapping of Inorganic Chemical Space for the Design of Transition Metal Complexes and Metal-Organic Frameworks., 127-179.

[4] Navneet Bung, Sowmya Ramaswamy Krishnan, Arijit Roy. An In Silico Explainable Multiparameter Optimization Approach for De Novo Drug Design against Proteins from the Central Nervous System. Journal of Chemical Information and Modeling 2022, 62 (11) , 2685-2695.

[5] Cheng-Hao Liu, Maksym Korablyov, Stanisław Jastrzębski, Paweł Włodarczyk-Pruszyński, Yoshua Bengio, Marwin Segler. RetroGNN: Fast Estimation of Synthesizability for Virtual Screening and De Novo Design by Learning from Slow Retrosynthesis Software. Journal of Chemical Information and Modeling 2022, 62 (10) , 2293-2300.

[6] Giovanni Bolcato, Esther Heid, Jonas Boström. On the Value of Using 3D Shape and Electrostatic Similarities in Deep Generative Methods. Journal of Chemical Information and Modeling 2022, 62 (6) ,

[7] Nicolae C. Iovanac, Robert MacKnight, Brett M. Savoie. Actively Searching: Inverse Design of Novel Molecules with Simultaneously Optimized Properties. The Journal of Physical Chemistry A 2022, 126 (2),

[8] Richard S Sutton and Andrew G Barto. 2018. Reinforcement learning: An introduction. MIT press.

[9] Chen, M.; Beutel, A.; Covington, P.; Jain, S.; Belletti, F.; and Chi, E. 2018b. Top-K Off-Policy Correction for a REINFORCE Recommender System. arXiv preprint arXiv:1812.02353 .

[10] Li He, Liang Wang, Kaipeng Liu, Bo Wu, and Weinan Zhang. 2018. Optimizing Sponsored Search Ranking Strategy by Deep Reinforcement Learning. arXiv preprint arXiv:1803.07347 (2018).

[11] Nachum, O.; Gu, S. S.; Lee, H.; and Levine, S. 2018. Dataefficient hierarchical reinforcement learning. In Advances in Neural Information Processing Systems, 3303–3313.

[12] Rohde, D.; Bonner, S.; Dunlop, T.; Vasile, F.; and Karatzoglou, A. 2018. RecoGym: A Reinforcement Learning Environment for the problem of Product Recommendation in Online Advertising. arXiv preprint arXiv:1808.00720 .

[13] Feng, J.; Li, H.; Huang, M.; Liu, S.; Ou, W.; Wang, Z.; and Zhu, X. 2018. Learning to Collaborate: Multi-Scenario Ranking via Multi-Agent Reinforcement Learning. In Proceedings of the 2018 World Wide Web Conference on World Wide Web, 1939–1948. International World Wide Web Conferences Steering Committee.

[14] Ghazvininejad, M., C. Brockett, M.-W. Chang, B. Dolan, J. Gao, W.-t. Yih, and M. Galley. 2017. “A Knowledge-Grounded Neural Conversation Model”. arXiv preprint arXiv:1702.01932.

[15] Fergus Imrie, Anthony R. Bradley, Mihaela van der Schaar, Charlotte M. Deane. Deep Generative Models for 3D Linker Design. Journal of Chemical Information and Modeling 2020, 60 (4) , 1983-1995.

The capacity of private information data from diverse uncoded storage databases

Ayush Shrivastava
Research Scholar

Abstract— Our primary purpose is to retrieve a single record from N non-plotting informational sets with m = (m1/Mn) heterogeneous limit objectives. In order to reduce the expenses associated with downloading information during the PIR stage, this examination is being completed with the intention of concurrently nurturing the substance area stage and the information recuperation stage. The best approach to describe the cost of the perfect PIR download is with an instant program. It maintains the data on a massive number of servers and restores lost data while maintaining the data's covert nature. Still, the processes for the infinite storage problem require splitting each message into an incredible number of sub-messages in comparison to the number of Dbs. The information related to our plan is encrypted before being sent off the cloud servers. Additionally, it is divided into many pieces of varying lengths. The preliminary evaluation of different configurations confirmed the productivity and rationality of the suggested approach, which demonstrated remarkable execution and consistency throughout the testing.

Keywords— linear program, non-colluding, private information retrieval (PIR), storage databases, un coded.

I. INTRODUCTION

In the data hypothesis local domain, the offered issue of private data recovery (PIR) has generated a lot of curiosity with driving endeavors. A client must locate a record in the traditional PIR environment by searching through data sets of documents that store identical information about entire records to the point that no single data set is able to identify the characteristics of the perfect record [1].

Different from FS-PIR, SC-PIR has an additional plan aspect that requires a carefully thought-out capacity structure. As an example, the initial homogenous SC-PIR plot reached the limit. The capacity position strategy for the conventional coded storage common connection problem. This approach has several limitations because of the capacity arrangement, which requires that each message be divided into O (expN) sub-messages [2]. For many different information bases, the suggested PIR plot may therefore be nonsensical. The decentralized stockpiling scenario summed up this workable plan. Also examined is the direct coded stockpile configuration at the data sets. In addition, the optimal straight plan was suggested, and the SC-PIR problem was investigated using the Shannon hypothetical approach for the approved cases of K = 2 and N = 2. To make matters even more intriguing, they also illustrated how a non-direct path can use less capacity than the optimal straight approach.

The following are our commitments in this paper:

- By establishing a nonexclusive relationship between the FS-PIR and SC-PIR concerns, we provide a general plan system for the SC-PIR issue. Given this correlation, a SC-PIR plan can be quickly created from an arbitrary FS-PIR plot.
- For homogeneous SC-PIR, we provide two capacity arrangement schemes that assume N sub-messages per message overall without taking into account the number of sub-messages necessary for question age.
- To address the heterogeneous SC-PIR position issue for general N and number t, which assumes N cycles and N sub messages per message, we offer an iterative stockpile arrangement calculation.
- We expand our results to take into account non-number t.

A. *Problem Formulation*

There are K autonomous messages, W_1…, W_K every one of size L pieces.

$$\boldsymbol{H}\ (\boldsymbol{W_1} \ldots, \boldsymbol{W_K}) = \boldsymbol{H}\ (\boldsymbol{W_1}) + \cdots + \boldsymbol{H}(\boldsymbol{W_k}) \qquad \ldots 1$$

$$\boldsymbol{H}\ (\boldsymbol{W_1} \ldots, \boldsymbol{W_K}) = \ldots = \boldsymbol{H}(\boldsymbol{W_k}) \qquad \ldots 2$$

B. *Sc-Pir Plan Engineering and Feasible Rat*

We provide a comprehensive engineering strategy for the SC-PIR system in this section. We divided the SC-PIR problem into stages related to location and conveyance, and then we suggested a particular scenario and choice of conveyance strategies. The suggested plan's workable pace is outlined.

$$\boldsymbol{m_f} \hat{=} \bigcup_{\boldsymbol{k} \in} \boldsymbol{W_{k,f}} \qquad \ldots 3$$

C. *Translating Heterogeneous Sc-Pir To A Filling Issue*

Here, we translate the heterogeneous PIR stockpile position problem into a similar filling problem (FP). Firstly, we express a set of sufficient conditions to satisfy the SC-PIR limit. Next, we demonstrate an FP solution for the achieve limit in a similar heterogeneous SC-PIR stockpile position problem [3]. Next, we formally describe the FP and elucidate its connection to the sufficient conditions in Lemma 1. Furthermore, we specify a number of requirements that guarantee an FP arrangement and show that, for any t ∈ Z+, a heterogeneous SC-PIR stockpile position arrangement exists in general. This section motivates the remainder of the research, which seeks to provide a solution to the heterogeneous SC-PIR scenario problem by addressing an identical FP.

D. Adequate Circumstances to Accomplish Limit with respect to SC-PIR

$$\sum_{f|\,N\,f=[t]} af = [t] - t \qquad \ldots 4$$

$$\sum_{f|\,N\,f=[t]} af = (t) - [t] \qquad \ldots 5$$

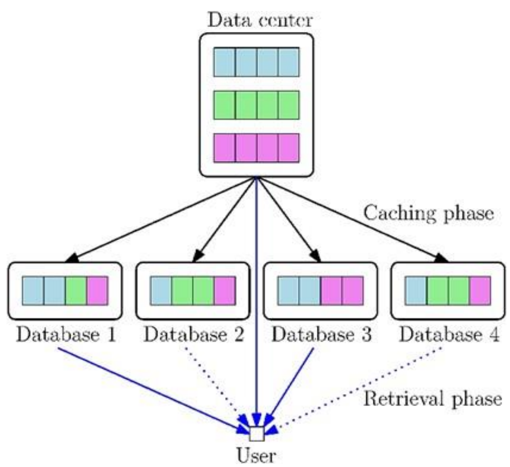


Fig. 1 Adequate Circumstances to Accomplish Limit with respect to SC-PIR

Coded reserving refers to the practice of early and appropriately placing documents in clients' neighborhood stockpiling stores and skillfully organizing conveyance plans at the hour of specified client requests in a way that minimizes traffic throughout the conveyance stage [4]. In the first configuration (see to Figure 1), N clients connect to a server with K documents via a mistake-free shared connect, and each client has a neighborhood memory that can store up to M data. There are two stages to the framework's operation: the conveyance stage and the situation stage. The server stores the records in each client's local memory throughout the scenario stage. During the delivery phase, each client requests a document from the server, and the server anticipates fulfilling each request with the least amount of traffic [5]. Under the assumption that the client arrangements in the two stages are identical, the server can more effectively orchestrate the content in each client's local memory, a feature known as unified coded reserving. suggests a symmetric bunch reserving scheme that is shown to be optimal in the case of an integrated uncoded position [6]. If there is a change in the configuration of clients in the two phases, the server will not be able to coordinate the documents in that state uniformly. All things considered, decentralized coded reserving refers to the server's treatment of each client independently and indistinguishably. suggests a consistent and erratic reservation strategy, which is shown to be optimal in the case of a dispersed, unseeded position **[8]**.

II. RELATED WORKS

(H. Sunand S.A 2018) Another version of the confidential data recovery (PIR) problem that we intend to tackle is flexible confidential data recovery (PPIR), where the client is pliable and interested in any message from a perfect subset of the available dataset. We first derive bounds on the M-PPIR rate, defined as the ratio of the ideal data measure to the total amount of downloaded data, followed by the corresponding achievable plans. As a result, we demonstrate that the limit of PIR with n data sets and Γ messages is coordinated by the PPIR limit, or the most extreme achievable PPIR rate, for n noncolliding data sets **[8]**.

(K. Banawanand 2018) Meanwhile, the problem of substance task for heterogeneous data sets (reserves) is investigated in the context of coded storing. The goal of the coded reservation issue is to cooperatively organize the circumstances and gradually relax the conveyance stage in order to reduce the amount of traffic during peak hours. suggests a strategy of streamlining that improves situation and conveyance plans by addressing a direct program. investigates the effects of reserve size heterogeneity on the conveyance load memory compromise with encoded arrangement using this streamlined structure **[9]**.

(Ibrahim, A.M 2019) We focus on a network supported by two jump stores, where a blend organization is connected to a server through a layer of transfer hubs. We take into account the scenario in which the end clients and the hand-off hubs are both capable of making reservations. We provide upper and lower bounds that apply to any mix organization, keeping in mind that previous research had focused on models in which the transfers lack reserves and on plans that made sense for a special kind of mix organizations **[10]**.

(Hassanzadeh, P.; Tulino 2018) Content delivery across a Gaussian transmission channel (BC) between clients and a server containing a data set of correlated contents is examined in a multi-client store backed broadcast network. When customers have stores of the same size, the base transmission power required to satisfy all possible interest mixes is evaluated. A lower limitation on the necessary transmit power as a component of the reserve limit is established, anticipating an unrecorded store scenario **[11]**.

(Wan, K.; Tuninetti 2018) One interesting solution for reducing heavy traffic loads and improving the Web of Things' (IoT) client inertness experience is the reservation approach. In this paper, two novel coded reserving plans that significantly reduce transmission dormancy for the unified and decentralized storing settings, separately, are proposed by leveraging edge store assets and correspondence amazing opportunities in gadget to-gadget (D2D) organizations and broadcast organizations **[12]**.

III. METHOD

In the accompanying hypothesis, we characterize the optimal standardized download cost for PIR from decentralized uncorded reserving data sets.

Hypothesis 1. The optimal standardized download cost for PIR from decentralized unrecorded reserving data sets with K records, where each document is of size L pieces, N data sets despite a server farm available throughout the recovery stage, and a stockpiling size need of μKL, $0<\mu<1$, bits for each data set, is:

$$D^* \sum_{n=1}^{N+1} \binom{N}{n-1} \mu^{n-1} (1-\mu)^{N+1-n} \frac{1}{(1+\frac{1}{n}+..+\frac{1}{n^{k-1}})}. \quad \ldots 6$$

Inspiring Model: $K = 3 \; and \; N = 2$

The scenario where the server farm keeps K=3 free records with the names A, B, and C and each document is composed of size L parts is taken into consideration in this model.

A. *Achievability Scheme*

During the reserve phase, each data set randomly and reliably stores 3μL pieces from a total of 3L pieces from the server farm in order to meet the capacity size need. Through a comparable likelihood dispersion, each data set functions independently, resulting in decentralized reserving.

Assume that during the recovery stage, the client has access to N=2 data sets, designated as DB1 and DB2, even though the server farm, designated as DB0, is not. The client also wants to retrieve document A covertly. Let's focus on a single record, let's say A, initially. Document An can be divided into four smaller files **[13]**.

$$A = (A_0, A_0, 1, A_{,2,0}, A_{,1,2,}), \quad \ldots 7$$

B. *Converse Proof*

Here, we demonstrate that the most decreased standardized download cost for N=2 information base exists among all decentralized storing likelihood appropriations PH. For a decentralized storage likelihood conveyance PH, the recovery stage has a following H.

$$D_\delta \geq L + \frac{4}{27} \sum_{k=1}^{3} H(W_k) + \frac{11}{108} \sum_{i=0}^{2} \sum_{k=1}^{3} H(W_k \,|Z_{|0:2|fi}) +$$

$$\frac{17}{54} \sum_{i=0}^{2} \sum_{k=1}^{3} H(W_k \,|Z_{|0:2|fi}) + o(L) \quad \ldots 8$$

C. *Algorithm for Veiling the Code of Put away Projects*

We examine the implementation aspects of this strategy with the example of hiding the source codes of projects that have been shelved for Prophet DBMS. While the general layouts of the content below apply to any DBMS of the class under consideration that maintains the ability to work with stored projects, the main points of this paper relate to explicit framework tables, views, programming executions of native cryptographic functions, and a few other objects of a specific DBMS.

D. Implementation

Our model is run on the highest level of our OpenStack-based classified cloud environment. By showing numerous virtual examples that are divided into three characterizations according on their design, we have developed our suggested contrive:

Class 1: Filling in as limit servers are virtual occurrences. Those servers have a 4 VCPUS, 4 GB Hammer, and 80 GB circular layout. We have taken into account the following, therefore those servers' IP addresses are: with subnet cover. Managing the mixed data slices that are provided to them by the Limit Chairman is one of those limit servers' two primary responsibilities. The task that results is returning the encoded data scales to the Entry Chairman server as answers so that the referring client can receive them.

Class 2: two examples that are virtual. Four VCPUS, eight GB Crush, and a twenty GB plate are the plans for each of them. They are filling in as the IP address-based PIR Show server and the IP address-based Key Box server. To ensure the confidentiality of the data, the client shouldn't be supplied so many data slices after the region obtained by the PIR Show from the Entry Manager server. Similarly, the customer shouldn't receive fewer data cuts than the typical cuts necessary to provide the relevant data. In the same way, the PIR Show server executes the PIR show to guarantee that the referenced data is covered by the cut characters from the circulated stockpiling servers, in addition to guaranteeing that the client is given the appropriate data slices so they can choose to get the relevant data. The keys must be consigned for the reference clients as well as the data owners by the Key Boss server.

Class 3: two digital representations. They are all set up with eight VCPUS, sixteen GB Hammer, and forty GB circle. They function as IP-assigned Storing Boss servers and IP-assigned Access Overseer servers. The encoded data cuts are received by the Limit Manager, who then transmits them to the relevant circulating stockpiling events. The Entry Manager delivers the encoded data cuts to the PIR Show after obtaining them from the proper stockpiling events.

IV. RESULTS AND DISCUSSION

We provide the exhibition investigation of our proposed plot in the context of two necessary circumstances.

A. Client Situation

As part of our agreement, the customer must use a standard encryption method in addition to a more flexible unscrambling method that differs from the Trademark Encryption Plan (AES). Because the unscrambling in our arrangement is parallelized on a few focal VCPUs, the client handling is quite convincing. We have considered five distinct size classes for our tests:

Table 1 Throughput shifting the quantity of thought about cuts for four different size classes.

Through put (MBPS)	Number of slices
50	32.127
100	512.2047
150	81.23444
200	124.22347
250	25.44137
300	27.23566
350	42.56644

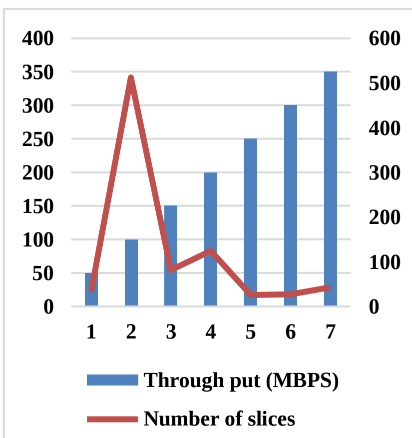


Fig.1 Throughput shifting the quantity of thought about cuts for four different size classes

B. Overlay Situation

The Fast (it gathers sections inside compartments) organization is used as a form of perspective to separate the overlay condition in our proposed layout. To complete the acquire and put piece methods that address our strategy, we have enlisted Fast's DLO assistance. We have developed a Python Fast client application that executes the get and additionally put part strategies that handle our interaction based on our tests at the overlay scenario. We used two methods: first, we divided an object into its component parts, and then we applied Fast's DLO.

Table 2 The execution time for the get demands on Quick

Times (S)	Object Size	
KB	MB	GB
22	45	77
41	42	41
25	23	85
45	70	25

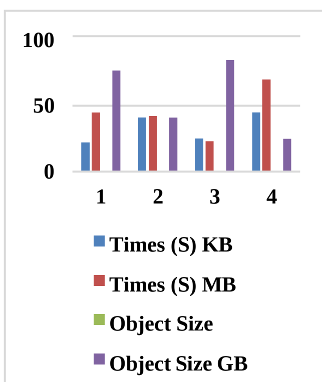


Fig. 2 The execution time for the get demands on Quick

The above-mentioned strategy of handling one piece for each section works well in cases of low costs, but as items get bigger, the framework's ability to convey data is drained, creating a significant bottleneck. Examining a 1 GB object, the anticipated execution time for the obtain requests is approximately 92 seconds for a single concept about piece and 1000 seconds for 1024 idea about parts for an equivalent thing size. Owing to a 64 KB object, the estimated execution time for the get requests is typically 0.08 seconds for a single thought segment and 50 seconds for 1024 thought segments for a nearly equivalent thing size [14].

Coded storing is suggested to satisfy PIR in late deals with SC-PIR; that is, the document position arrangements of coded reserving are beneficial for the SC-PIR sub-message situation issue. We demonstrate in this study that coded reserving position techniques are too much for SC-PIR by putting forth two innovative sub-message arrangement designs that achieve the limit. Relegating different records to a very large number of covering client bunches is crucial in the coded storage problem in order to set up multicasting open doors so that a client can drop "impedance" from a received coded transmission that also serves other clients. The fact that only one client is being served makes the SC-PIR issue less perplexing **[15]**.

V. CONCLUSION

In order to effectively and securely store and retrieve sensitive data from untrusted cloud servers, we unveiled a potent PIR conspire. Our design enables the owners of the information to segregate and encode their own data into small segments that fall into five distinct size categories. We demonstrate through our implementation and trial analysis the feasibility and suitability of our suggested architecture, which provides notable throughput gains in terms of source security and decoding. An approximate period of 92 seconds is expected to complete the obtain requests for a 1 GB item; for an equivalent thing size, the span is 1000 seconds for 1024 thought about parts. In general, it takes 0.08 seconds to handle a single idea about segment and 50 seconds to handle 1024 idea about segments for an equivalent thing size when handling get requests for a 64 KB object. Additionally, the suggested scheme maintains its similarity to all current distributed stockpiling circumstances, making it applicable to a wide range of application areas.

VI. FUTURE WORKS

Without any prior planning, the SC-PIR plans were developed, and surprisingly, we discover that our plans only call for a polynomial number of sub-messages for each message. Additionally, we provided sufficient and significant conditions to meet the SC-PIR limit, which can support the development of further SC-PIR plans. These results not only showed that there is a solution to the general stockpiling problem for heterogeneous SC-PIR, but also that a simple iterative stockpiling scenario computation exists to the extent that the conditions are satisfied following each emphasis.

VII. REFERENCES

[1] J. Shen, C. Wang, T. Li, X. Chen, X. Huang, and Z.-H. Zhan, "Secure data uploading scheme for a smart home system," Information Sciences, vol. 453, pp. 186–197, 2018.

[2] H. Sun and S. A. Jafar, "The capacity of robust private information retrieval with colluding databases," Institute of Electrical and Electronics Engineers Transactions on Information Theory, vol. 64, no. 4, part 1, pp. 2361–2370, 2018.

[3] C. Gao, S. Lv, Y. Wei, Z. Wang, Z. Liu, and X. Cheng, "M-SSE: an effective searchable symmetric encryption with enhanced security for mobile devices," IEEE Access, vol. 6, pp. 38860–38869, 2018.

[4] J. Xu, L. Wei, Y. Zhang, A. Wang, F. Zhou, and C. Gao, "Dynamic Fully Homomorphic encryption-based Merkle Tree for lightweight streaming authenticated data structures," Journal of Network and Computer Applications, vol. 107, pp. 113–124, 2018.

[5] Banawan, K.; Ulukus, S. Noisy private information retrieval: On separability of channel coding and information retrieval. IEEE Trans. Inf. Theory 2019.

[6] Jia, Z.; Sun, H.; Jafar, S.A. Cross Subspace Alignment and the Asymptotic Capacity of X-Secure T-Private Information Retrieval. IEEE Trans. Inf. Theory 2019, 65, 5783–5798.

[7] Tian, C.; Sun, H.; Chen, J. Capacity-Achieving Private Information Retrieval Codes with Optimal Message Size and Upload Cost. IEEE Trans. Inf. Theory 2019, 65, 7613–7627.

[8] H. Sunand S.A. Jafar, "The capacity of robust private information retrieval with colluding databases," IEEETrans.Inf. Theory, vol.64, no.4, pp.2361–2370,Apr.2018.

[9] K. Banawanand S.Ulukus, "The capacity of private information retrieval with colluding databases" IEEETrans.Inf.Theory,vol.64,no.3, pp.1945–1956,Mar.2018.

[10] Ibrahim, A.M.; Zewail, A.A.; Yener, A. Coded Caching for Heterogeneous Systems: An Optimization Perspective. IEEE Trans. Commun. 2019, 67, 5321–5335.

[11] Hassanzadeh, P.; Tulino, A.M.; Llorca, J.; Erkip, E. Rate-memory trade-off for caching and delivery of correlated sources. arXiv 2018, arXiv:1806.07333.

[12] Wan, K.; Tuninetti, D.; Piantanida, P. On the optimality of uncoded cache placement. In Proceedings of the 2016 IEEE Information Theory Workshop (ITW), Cambridge, UK, 11–14 September 2018.

[13] Kumar, S.; i Amat, A.G.; Rosnes, E.; Senigagliesi, L. Private Information Retrieval From a Cellular Network With Caching at the Edge. IEEE Trans. Commun. 2019, 67, 4900–4912.

[14] Ibrahim, A.M.; Zewail, A.A.; Yener, A. Coded Caching for Heterogeneous Systems: An Optimization Perspective. IEEE Trans. Commun. 2019, 67, 5321–5335.

[15] R. Zhou, C. Tian, H. Sun, and T. Liu, “Capacity-achieving private information retrieval codes from MDS-coded databases with minimum message size,” Mar. 2019, arXiv: 1903.08229.

Blockchain-Based Distribution and Storage of Government Documents

Sonali Bhajwa
Research Scholar

***Abstract*—** In the previous ten years, electronic taxpayer-supported organizations have rapidly transitioned from paper-based regulation architecture to computerized administrations. Increases in environment development and appropriated record innovation are crucial to achieving blockchain's problematic capabilities. The evaluation of non-mechanical barriers, such as the incompatibility of blockchain-based agreements with the existing legal and regulatoryframeworks, should be included in the approach strategy. It is not possible to achieve this importantstrategy point by bringing innovation to outdated frameworks. The writing research in this paper focuses on how blockchain innovation applies to e-government applications. This study looks at the fundamental levels of innovation and development of the suggested arrangements, as well as e-taxpayer driven organizations that may benefit from the use of blockchains. The objective is to demonstrate blockchain's genuine potential and dedication to the industry, as well as to provide states considering investing in this innovative technology with useful information. Additionally, the goal is to assist scholars in their future endeavors, such as developing blockchain-powered e- taxpayer driven organizations. Outlining the value and potential of blockchain is also an aim.

***Keywords*—** *blockchain technology, e-Government, digital services, e-government services, organizational systems, public service.*

I. INTRODUCTION

The early uses of blockchain innovation were in digital money. Blockchains are simple, secure digital records that operate in a distributed network of dispersed hubs, allowing transactions to be completed securely and typically without the approval of a central authority. The data from each exchange is hashed and stored in blocks, each of which also contains the hash from the previous block. This sequence of blocks is referred to as the record. This makes it possible for a group of clients to keep track of exchanges in a shared ledger such that, once added to the blockchain, they cannot be altered. Blockchains thus enable distributed centers without a trust association to exchange trade data without the need for intermediaries or delegates [1]. This information may pertain to money, contracts, real estate titles, clinical and educational records, offers, the purchase and sale of goods and services, or other transactions or assets that may be converted to digital form. Like in countless other domains, e-government has explored blockchain technology to forward the transformation of arrangement execution and collaborate with the core of fundamentally sound public organizations. The primary rationale for implementing this innovativesystem is to avoid using a central point of contact for residents' or associations' interactions with government-trained specialists, to decentralize the collection, storage, and management of data, and to guarantee the accuracy and consistency of data.

A. Blockchain For Government

In essence, blockchain is a public record where all transactions are stored in a chain of information bundles, or impedes, and shared throughout a common organization. Every complex hub within the company has a copy of the blocks. The majority of the participating hubs in the organization must accept each trade or computational event that is recorded in the public record using an agreement component. The exchange is stored in a separate block in the unlikely event that it was agreed upon. In order to establish a connection to the previous block, a timestamp is appended to the new block along with a nonce—an erratic number—and a hash pointer. The new block is then appropriated throughout the organization and added to the previous block chain. Therefore, blockchain provides a stable, decentralized, consistent, flexible, and auditable exchange platform that enables a decentralized exchange to take place without the requirement for a central delegate **[2]**.

- **Decentralization -** Unlike traditional trades, which require approval from a centralized authority, each centre point within the association possesses an identical copy of the record and can facilitate exchanges. This method oversees the benefits of blockchain-based exchanges in terms of information consistency, enhanced client control, attack resistance, transparency, and adaptability to internal setbacks. It also facilitates the introduction of external delegates, such as a governmental body or financial institution.
- **Consistency -** The use of a timestamp, a cryptographic seal, and an understanding framework ensures that incorrect transactions won't be accepted and makes it challenging to alter, remove,

or duplicate transactions that have already been stored in the blockchain. These characteristics of blockchain include proprietorship verification, data consistency, distortion prevention, and immutable exchange records.

- **Secrecy -** Blockchain-enabled communications use public-key cryptography and are directed between two groups whose identities are concealed by pseudonyms. Client security will thus be prioritized in exceptional electronic collaborations.
- Every exchange in a blockchain is stored in a subsequent request, together with the hash of the previous block and the maximum hash of the current block, which is meant to relate the subsequent block when it is added. Exchanges may be verified and properly followed with this section. These core characteristics of blockchain innovation offer public enterprises a number of potential benefits. Some of the benefits, such as distributed creativity, uniformity, and transparency, can be used to eliminate deceit and corruption in public areas. This innovation allows for the oversight-free recording of all transactions, including those involving open administrations. This leads to increased transparency and, ultimately, public confidence in open organizations. If the progress in blockchain technology could prove its promised benefits, it is possible that the technology would soon reach a tipping point and be widely accepted by governments worldwide. Thus, additional inter-disciplinary research is necessary for larger-scale blockchain components like effect and chance, plan components, organization models, and effect. Therefore, in order to design the challenges of blockchain gathering in the public authority area, we employed a writing study. However, in the next section, we first provide the methodology we used to locate the writing**[4]**.

B. Blockchain and advanced state-run administrations

The state of craftsmanship worldview in policy management science is advanced government. An informational or empowering influence of modernizing policy management was acknowledged as the role of digitalization in the earlier, much narrower concept of e-government. The arrangement of client-driven, lean, and creative public administrations is at the heart of computerized government, which advances significantly. These services and administration conveyance models should make use of state-of-the-art technologies as well as resident and legislative data resources. In the context of the new paradigm of administrative plan formulation and administration conveyance, blockchain is arguably one of the most inventive technological innovations. It is claimed that the primary benefits of implementing blockchain technology in state-run administrations include:

- ✓ Decreased financial expenses, time and complication in between legislative and public confidential data transactions that boost the managerial capability of state-run administrations.
- ✓ The use of appropriated records and programmed brilliant agreements has led to a decrease in administration, optional power, and defilement.
- ✓ Enhanced computerization, transparency, auditability, and accountability of data in administrative libraries to assist inhabitants.
- ✓ Increased confidence among citizens and institutions in administrative procedures and documentation due to the use of computations that are now not exclusively governed by the government.

C. e-Government Services

In order to motorize public organizations and work with their usage by citizens (G2C), associations (G2B), and intergovernmental bodies (G2G), e-Government is an official endeavor that makes use of information and communication technology (ICT). Examples of this work include secure data transfer, e-obtaining, filling government structures, character the board, electronic majority rule, etc. The goal of this effort is to electronically organize public organizations so that residents and organizations can provide basic, trustworthy, and safe assistance to one another in a decentralized manner without the involvement of public professionals. It also aims to remove organization. In general, electronic organizations are easily usable, accessible from anywhere, and readily available [5]. Requests from individuals, groups, and legislators to transition to e-citizen supported associations have prompted the European Commission to work on an e-government activity plan.

II. RELATED WORKS

(Payeras-Capella, M.M, 2019) Electronic naming may be beneficial to many different types of transportation frameworks (ET). ET frameworks are being dynamically integrated into the transportation foundation, resulting in reduced associated costs and lead times, along with improved framework management. However, as client confidentiality isn't usually guaranteed and customers might be watched and their profiles of typical upgrades could be built, the work of ET frameworks takes into account a variety of security abuses in a powerful and comprehensive manner. We look at and demonstrate the centre counsel in this survey piece, with a special emphasis on client protection **[6]**.

(Oliveira, T. 2020) The way people interact with their urban neighborhoods affects the overall quality of life. Their involvement in amicable decisions is crucial for supporting open choices that impact policy, administration, and training. In the context of intelligent and technologically advanced metropolitan communities, this collaboration could be aided, especially by the continuous advancements in blockchain technology. This work shares experiences regarding how creative innovations and innovative ideas for metropolitan areas can help society face daily challenges and improve the mindfulness of its citizens. Through the application of Information and Communication Technology (ICT) and data, technological innovations can propel economic and social progress. In this particular case, e-administration, associated with problematic conceptslike blockchain, is emerging as a key component for a decentralized vote-based system **[7]**.

(European Commission 2018) The concept of #distributed records has entered conventional analysis and strategy plans in less than ten years after it first surfaced in 2008. Motivated by the success of Bitcoin and the flurry of potential applications, there was a lively gathering that raised expectations about the remarkable potential of blockchain technology for both the public and commercial sectors—assuming that it wasn't overhyped. The latest JRC report examines states' ongoing exploration of blockchain innovation. Examining seven blockchain-based services developed in Europe, the study is supported by public experts at both local and state levels. Blockchain innovation, according to the assessment, can reduce bureaucracy, improve the efficacy of regulatory cycles, and raise public trust in publicly accessible record keeping. However, unlike what is sometimes said, blockchain has not yet shown to be an amazing or even troublesome breakthrough for governments **[8]**.

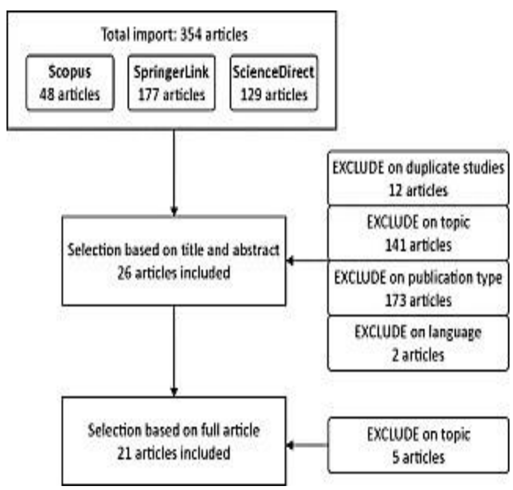
(Baldacci, E.; Frade 2021) Like many other domains, blockchain technology has been studied by e-government to facilitate changes in policy management and assist in setting up simple, safe public administrations. The main motivation for adopting this automated approach is to avoid using a central point of contact for citizens' or organizations' interactions with government experts,to decentralize the gathering, storing, and processing of data, and to ensure the integrity and constancy of the data **[9]**.

III. METHODS

The audit paper's exploration technique echoed an earlier Scopus mission to identify distributed works related to the application of blockchain innovation in e-taxpayer supported organizations [10]. Looking at the titles, catchphrases, and unique aspects of the articles, the search query that was employed for this purpose is as follows:

$$\begin{array}{l} \boldsymbol{TITLE - ABS\ KEY'e - governance','} \\ \boldsymbol{- governance',' or\ smart\ governnc} \\ \boldsymbol{'' or\ e\ governance} \\ \boldsymbol{- governance \qquad AB\ blockchain} \\ \boldsymbol{, \qquad and} \\ \boldsymbol{OR\ distribution} \end{array} \quad \ldots\ldots.1$$

To locate research publications, we used the following three electronic data sets: Springer Link, Science Direct, and Scopus. To guide a search for distributed diaries articles, meeting methods, studios, and conferences, the title, theory, and watchwords were used. Up to December 30, 2017, the writing search produced roughly 354 articles (see Figure 1).



A. Authentication

Electronic naming may be beneficial to many different types of transportation frameworks (ET). ET frameworks are being dynamically integrated into the transportation foundation, resulting in reduced associated costs and lead times, along with improved framework management. However, as client confidentiality isn't usually guaranteed and customers might be watched and their profiles of typical upgrades could be built, the work of ET frameworks takes into account a variety of security abuses in a powerful and comprehensive manner. We look at and demonstrate the center counsel in this survey piece, with a special emphasis on client security.

B. Data Sharing

an information sharing framework that illustrates how to prevent information breaches during exchange and highlights information security lapses when dispersing information among different government departments. Private blockchains are used to divide data across hubs; this verifies the hubs' membership in the organization and gives them mutual trust. It simultaneously reduces the disarray in the information structure and sets out the important points of the content. In accordance with the requirements, the framework may also query the data, compile the names of the departments that possess the data, and exchange the request **[11]**.

C. e- Voting

a balloting mechanism that uses blockchain technology with permissions. In order to ful fill their security and insurance requirements, they have implemented a private Proof of Authority (PoA) blockchain on Go-Ethereum. The arranging tool is reliant on the way of life under threat, which facilitates the interchange of messages.

D. Land Property Services

land in Dubai as a context-oriented study by identifying the confusing viewpoints, examining how Dubai's land cycle functions, and identifying the challenges and effects of implementing blockchain innovation on the real estate market. They have sent a permissioned blockchain engineering in this evaluation to improve exchange transparency, reduce costs, and work offshore cycles. They have considered the Hyperledger Surface stage and its ingenious connections during their investigation.

E. e-Delivery Services

In contrast to present tactics, the e-Transport organization will reduce the assistance of third-party trusted parties while maintaining EU requirements for e-Movement structures. The suggestions were predicated on an e-Move framework that used smart contracts and blockchain for both private and public use **[12]**.

F. Human Resources Management

presented a design that is applicable to all. A blockchain network powered by clever contracts. Moreover, they provide an HR [13]. The executives (HRM) have a decentralized correspondence design that guarantees openness and protects people from retaliatory actions. Three types of customers are included in the model, whose public keys are verifiable: applicants, spectators, and organizations.

G. Government Contracting

A blockchain-based architecture that can be applied to any method of handling government contracting and takes into account continuous verification of e-citizen driven associations. This system provides the board with accountability, clarity, and enhanced support. They provided a non-exclusive blockchain solution for implementing any administration contracting strategy, using the US Free Endeavor Association concept as a background analysis. The blockchain-based system manages all transactions and disseminates the results to a wider audience.

IV. RESULT AND DISCUSSION

The blockchain was first limited to financial transactions, such as advanced cash. Long-term, blockchain features like transparency, information security—which makes changes to the data unnecessary when they are added to the blockchain—and its decentralized nature have sparked a real concern among certain experts to implement blockchain development in industries other than finance. The e-association for more capable and safe organizations for residents, affiliations, and assemblies is one of these domains.

Table 1 Level of development gave in the included e-taxpayer driven organizations

Proposal	10%
Prototype	21%
System evaluation	5%
Conceptual	32%
Experimental	32%

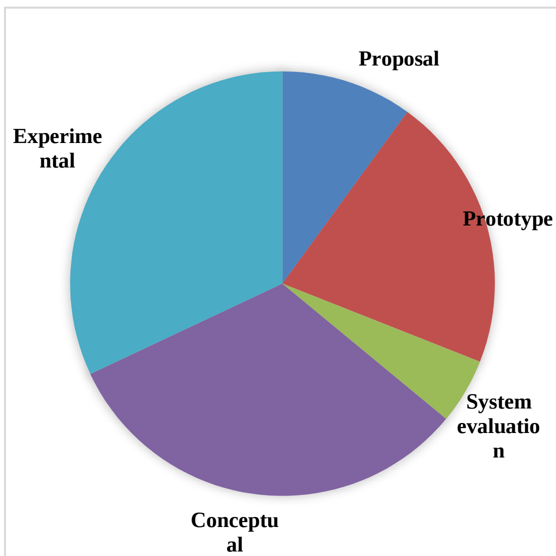


Fig 2 Level of development gave in the included e-taxpayer driven organizations

Table 2 Normal degree of development per e-taxpayer driven organization

Authentication	45
Data Sharing	2.36
e- Voting	14

Land Property Services	25.3
Human Resources Management	16
e-Delivery Services	23.4
Government Contracting	55.3

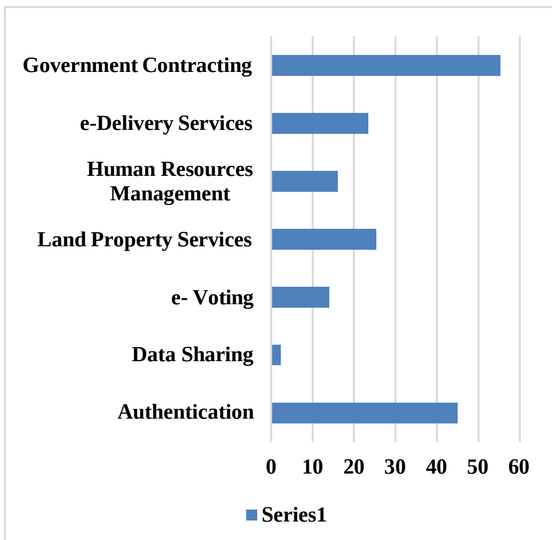


Fig. 3 Normal degree of development per e-taxpayer driven organization

While fewer findings were found in relation to other e-citizen sponsored associations, the majority of findings in the coordinated review for this study were connected to affirmation, data sharing, and e-projecting a polling form organization. Furthermore, only 4 out of the 21 proposed plans were finished as models, and most of the improvements made to all courses of action were at the appraised and exploratory level (2.89); for further details, see Figure 3.

We predict that more research will be done on citizen-driven businesses using blockchain development in the coming years when changes are made to introduce blockchain stages or new ones are delivered. Rather than other arrangement components, the trades are carried out using the basic working out resources through clever contracts **[14]**.

V. CONCLUSIONS

Since the advent of blockchain technology in fields other than virtual currencies, research efforts have expanded significantly in the preceding years. In the field of public administration and e-administration, blockchain innovation is regarded as a progressive approach. It serves as an empowering force that enables citizens, organizations, and legislators to work together in an easy and problem-free manner. Precisely organized, its growth stems from the combination of honesty, openness, categorization, and duty [15]. In addition, a distributed blockchain network increases trust among all participants because transactions are conducted securely without the approval of a central authority. The current literature demonstrates that the primary barriers to blockchain adoption are rooted in mechanical aspects such as security, flexibility, and versatility. Additionally, from a natural perspective, the most pressing issue that needs to be addressed is support for laws and conventions.

VI. FUTURE WORKS

To narrow down the results and gain insight into our assessment questions about the use of blockchain in e-citizen driven initiatives, we implemented a two-level screening system in this review. In conclusion, this study's review provides a good starting point for further research into the development of blockchain-based e-Government systems for experts and qualified professionals.

VII. REFERENCES

[1]. , D.; Mell, P.; Roby, N.; Scarfone, K. Blockchain Technology Overview; Technical Report NIST IR 8202; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2018.

[2]. Nguyen, C.T.; Hoang, D.T.; Nguyen, D.N.; Niyato, D.; Nguyen, H.T.; Dutkiewicz, E. Proof-of-stake consensus mechanisms for future blockchain networks: Fundamentals, applications and opportunities. IEEE Access 2019, 7, 85727–85745.

[3]. Tasca, P.; Tessone, C.J. A Taxonomy of Blockchain Technologies: Principles of Identification and Classification. Ledger 2019, 4

[4]. Khan, S.N.; Shael, M.; Majdalawieh, M. Blockchain Technology as a Support Infrastructure in E-Government Evolution at Dubai Economic Department. In Proceedings of the 2019 International Electronics Communication Conference, IECC'19, Okinawa, Japan, 7–9 July 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 124–130.

[5]. Gartner. (2018a). Hype Cycle for Blockchain Business. Retrieved from

[6]. Payeras-Capella, M.M.; Mut-Puigserver, M.; Cabot-Nadal, M.A. Blockchain-Based System for Multiparty Electronic Registered Delivery Services. IEEE Access 2019, 7, 95825–95843.

[7]. Oliveira, T.; Oliver, M.; Ramalhinho, H. Challenges for connecting citizens and smart cities: ICT, e-governance and blockchain. Sustainability 2020, 12, 2926.

[8]. European Commission. (2018b). Blockchain Technologies. Retrieved January 18, 2019,

[9]. Baldacci, E.; Frade, J.R. Advancing Digital Transformation in the Public Sector with Blockchain: A View from the European Union. In Disintermediation Economics: The Impact of Blockchain on Markets and Policies; Kaili, E., Psarrakis, D., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 281–295.

[10]. Páez, R.; Pérez, M.; Ramírez, G.; Montes, J.; Bouvarel, L. An architecture for biometric electronic identification document system based on blockchain. Future Internet 2020, 12, 10.

[11]. Khan, K.; Arshad, J.; Khan, M. Investigating performance constraints for blockchain based secure e-voting system. Future Gener. Comput. Syst. 2020, 105, 13–26.

[12]. Alketbi, A.; Nasir, Q.; Abu Talib, M. Novel blockchain reference model for government services: Dubai government case study. Int. J. Syst. Assur. Eng. Manag. 2020, 11, 1170–1191.

[13]. Pinter, K.; Schmelz, D.; Lamber, R.; Strobl, S.; Grechenig, T. Towards a Multi-party, Blockchain-Based Identity Verification Solution to Implement Clear Name Laws for Online MediaPlatforms. In Proceedings of the Business Process Management: Blockchain and Central and Eastern Europe Forum, Vienna, Austria, 1–6 September 2019;

[14]. Di Ciccio, C., Gabryelczyk, R., García-Bañuelos, L., Hernaus, T., Hull, R., Indihar Štemberger,M., K˝o, A., Staples, M., Eds.; Springer International Publishing: Cham, Switzerland, 2019; Volume 361, pp. 151–165.

[15]. Páez, R.; Pérez, M.; Ramírez, G.; Montes, J.; Bouvarel, L. An architecture for biometric electronic identification document system based on blockchain. Future Internet 2020, 12, 10.

Using blockchain technology in the global payroll system

Mamta Ahuja

Research Scholar

Abstract— Payroll fraud can be completely eliminated and payroll system efficiency can be raised with the help of blockchain technology. Since payroll administration makes sure that employees are paid accurately and on time, it is an essential role for all enterprises, governments, and other organizations. Payroll systems in poor nations encounter difficulties from human manipulation, cybercrime, and ghost work. This essay examines the issues that developing nations' payroll systems face and how blockchain technology could be able to help. The paper's remaining sections describe how blockchain technology might allay worries about payroll systems. In order to evaluate the importance of blockchain technology generally and its applicability to payrollsystems, we examine the body of available material. This essay examines the consequences of thisfinding for the payroll tax compliance sector while acknowledging the technical evaluation that blockchain represents a new fundamental technology. It tries to predict and plan for the kinds of changes that this industry will see. It will evaluate the move's effects on industry and government.

Keywords— *Blockchain, Permissioned Blockchain, Payroll, blockchain technology, Decentralization, Public Universities.*

I. INTRODUCTION

A fundamental and basic connection for all administrations, businesses, and associations is finance the executives, as it impacts staff instalment, which ought to be accurate and timely. When the internal control of an association's financial framework is weak, monetary control and cheating become the norm. It is also suggested that the primary focus should be on the supervision and security of electronic instalment frameworks. In today's technological environment, a multitude of financial matters are handled on a daily basis. In New Zealand, the Ministry of Business, Development, and Enterprise (MBIE) reported that over 24,000 specialists are underqualified. There are theories that 2,000,000 additional staff members may be present even if the investigation is not yet complete. Non-industrialized countries confront additional challenges with instalment stages, such as control, cybercrime, phantom labourers, and centralization [1].

Organizations get the most from using blockchain innovation when they collaborate and create a shared stage, but at that moment, they are unable to define rules without asking other members to participate in their creation. If a company is building a blockchain only for itself, it will have challenges related to scale, information harmonization, and inward buying. In the unlikely event that additional businesses band together to build a blockchain, things are becoming a lot more intricate. Blockchain is a fundamental breakthrough that will take time to seep into our financial and social foundation; it is by no means a problem that can provide a low-cost solution for customary courses of action. The pace of inventive, mechanical, and institutional change will accelerate, but the rate of acceptance will remain slow.

A. Foundational Technology

Fundamental (groundbreaking) applications are supported by primary innovations. Business (and social) relationships inside financial contexts are practically characterized and then re-characterized by persistent and compelling applications. Blockchain is a crucial invention that allows for apps like Bitcoin, upending traditional worth-moving models in the process. It is very likely that Bitcoin (and other cryptocurrencies) will directly undermine the global business banking system. All the way down to the individual buyer, National Banks are feeling the weight of this challenge. Blockchain developments will extend far beyond financial transactions as new blockchain applications are created.

Blockchain controls for business archives [2]. On August 10, 2016, a blockchain application was created by Bank of America, HSBC, and the Singaporean government's IT department to replicate the procedures involved in exchanging letters of credit between banks, exporters, and shippers;

- ***Blockchain medical care record controls*** *On September1,2016, the Department of Health and Human Services received more than 70 proposals in response to the Exploration Challenge, Utilizing Blockchain in Health IT and Health Related. Patient records were successfully uploaded to the blockchain by a few papers. It was admired that the Ethereum blockchain balanced access to research with patient protection.*

- ***Electoral cheating counteraction Follow** My Vote is the first blockchain startup in the democratic sphere, having launched on July 4, 2012.43 On November 1, 2016, Spread Borderless joined44. Shortly after, the organization launched a spoof democratic blockchain (a proof of concept) in which people were asked to select a president (as if the democratic process was crucial to the current US official election).45 But the real money in a democratic blockchain application is in business applications (voting through an intermediary andservices that allow investors to vote at financial backer meetings they can't attend in person).The financial services company Broadridge began actively pursuing blockchain technology on October 6, 2016, with the goal of being "quick to showcase a total intermediary blockchain arrangement.*

B. Payroll on the Blockchain

Why do we not see any notable blockchain-based financial applications? All aspects of finance are digital. Finance collaborates with a number of administrative departments, each of which compiles copy data, holds it in progress, and conducts covering consistency reviews. AML and KYC regulations also apply to all exchanges. This is the ideal environment for distributed records in terms of production [3].

While it would be ideal to see a plethora of new businesses and traditional specialized organizations utilizing blockchain solutions for finance, this is not the case at the moment. By all accounts, a void exists. There are two plausible reasons for the void:

- Significant barriers to entry, namely the heightened lack of a fiat digital currency to enable seamless worldwide financial exchanges without crucially important new trade opportunities.
- Barriers that prevent new and innovative businesses from entering the market while also giving ongoing finance specialist cooperatives—who might believe they are immune to this innovation—the false impression that all is well.

C. Blockchain Technology

Blockchain is a distributed, decentralized electronic data set that can store transactions, events, and records, and it can also provide guidelines for updating this data. Without the need for delegates, it enables executing groups to share accountability for addressed resources in a consistent and dynamic distributed structure. The six phases of resource exchange between two financial performers employing blockchain innovation are depicted in figure 1 below **[4]**.

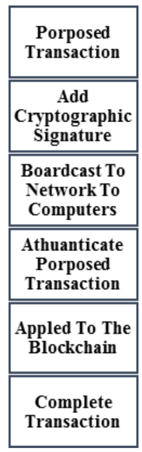


Fig.1 the six stages of resource trade between two financial entertainers utilizing blockchain innovation.

II. RELATED WORKS

(Aluko. 2020) Blockchain technology was initially applied to a public organization that does not require central authority; however, permissioned and private frameworks are gaining popularity and are better suited for financial frameworks. perceived permissioned blockchain as "Recent proposals for alleged permissioned blockchains aim to only authorize a limited group of readers and academics. Here, a central component selects and grants friends the ability to participate in the compose or read tasks on the blockchain. Readers and authors could also operate in separate, equal blockchains that are coupled to provide embodiment and security **[5]**.

(Ante, L. 2021) This study examines the ways in which the lucidity of logical discourse varies over time and the extent to which logical effect may be understood in terms of reference counts. Delegate datasets of 135,502 modified compositions from academic research publications related to twelve advancements of different development are the basis for the idea. Using three distinct measures of meaningfulness, it is discovered that over time, the digests' phrasing has grown increasingly cryptic **[6]**.

(Antwi, M., Adnane 2021) The issue of information trustworthiness in clinical preliminary care was reinforced by the recent introduction of the forceful assent concept. It may make sense to combine blockchain technology into a potent assent step. Owing to varying clinical initial environments, an interest-driven improvement process is necessary **[7]**.

(Baker, C., & Werbach, K. 2019) Because of the anticipated benefits, blockchain innovation has received a lot of attention from the media and is gaining traction among organizations inside financial administrations. Since blockchain-based frameworks are still in their infancy, it is important to understand their requirements in order to enable blockchain frameworks to be efficiently coordinated within financial assistance organizations. There are gaps in academic research regarding how CEOs evaluate the value of a blockchain-based system **[8]**.

III. METHODS

The price of bitcoin is the key factor that has affected everything related to the cryptocurrency market in the past and is also applicable to business. When Satoshi Nakamoto first introduced bitcoin in 2009, interest in blockchain technology began to rapidly grow. From that point on, the income growth was gradual until 2017, when the price of bitcoin reached around $20,000. As we approach the end of 2021, we can see that the market is once again positive and that the price of bitcoin has just split $42,000 while continuing to rise. Everyone is talking about bitcoin these days, therefore nobody has to miss the opportunity to make a truly outstanding contribution **[6]**.

Table 1 blockchain advancements (2009-2021).

2009	2.36
2010	2.01
2011	2.26
2012	2.34
2013	1.26
2014	2.24
2015	2.6
2016	2.2

2017	4.3
2018	3.22
2019	5.63
2020	8.3
2021	6.2

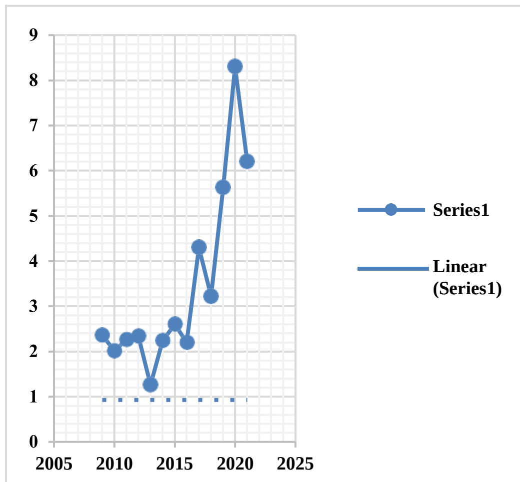


Fig. 1 blockchain advancements (2009-2021)

The invention of blockchain technology offers several advantages, including speed, simplicity, reach, and lower prices. There are many good guesses, as Gartner predicts that by 2030, the coin market capitalization would reach $3 trillion. It is anticipated that overall blockchain spending would increase from $1.5 billion in 2018 to $15.9 billion by 2023.

A. Permissioned Blockchain in Finance

The first public organization to use blockchain technology was one that did not require central authority; nevertheless, private and permissioned frameworks are gaining popularity and are better suited for financial frameworks. perceived permissioned blockchain as "Recent proposals for purported permissioned blockchains aim to only authorize a limited group of readers and academics. In this case, a focal substance selects and grants pals the ability to participate in the compose or read tasks on the blockchain [7]. Readers and essayists could also operate in separate, equal blockchains that are joined to provide illustration and security." Permissioned blockchain technology verifies and authorizes users and exchanges through the Part Specialist organization (MSP), as opposed to unidentified excavators. Some of the two permissioned blockchain platforms are Majority, R3 Corda, Ethereum, Hyperledger, and so on. Permissioned blockchain offers several advantages for financial frameworks. Firstly, it eliminates the need for the notorious Proof of Work, since all members of the organization are authorized. Secondly, it fits in well with real hierarchical structures, allowing government agencies, educational institutions, and corporate divisions to manage their own finances with full appointed power, even in the event of a network outage, while maintaining complete central oversight and auditability [8].

A blockchain arrangement for a finance specialist organization: developing. The ancillary elements of a blockchain agreement for a finance specialist co-op are notably under development at this time:

- Futurize employees track overtime on the Ethereum blockchain and are paid for it in the regular course of business;
- Managers pay their global workforce in bitcoins, which can be stored in bitcoin or automatically converted to local government currency in 60 countries worldwide with Spot wage through clever agreements that combine the first bitcoin blockchain and the Ethereum stage [9];
- J.P. Morgan In light of the Ethereum convention, Pursue is developing Majority, a private/permissioned blockchain that would arrange payments among financial institutions in a way that is entirely visible to controllers and totally private to members.

IV. RESULTS AND DISCUSSION

These are some of the crucial indicators that blockchain is changing the economy's destiny. Blockchain-enabled venture programming stages will help firms improve the reliability of their information, facilitate information exchange, and streamline procedures. Blockchain-based frameworks are being used, or will be used, by numerous large corporations to streamline operations. A few of them that started out as a combo are SAP, Salesforce, Microsoft, Prophet, and Salesforce **[9]**.

Table 2 Far Along Are Organizations with Blockchain

Paused	10.23%
Live	12.22%
Pilot	4.366%
Research	1.26%
None	46.3%
Development	4.23%

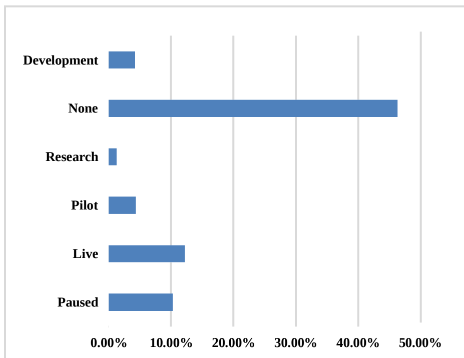


Fig.2 Far Along Are Organizations with Blockchain

Table 3 Top Subjects Organizers Tackle with Blockchain, % of Replies

Cost reduction	11%
Immutability	12%
Adoptability	12%
Trust	45%
Digitalized	55%
Integrity	16%
Certification	13%
Consumer experience	43%
Shaving	43%

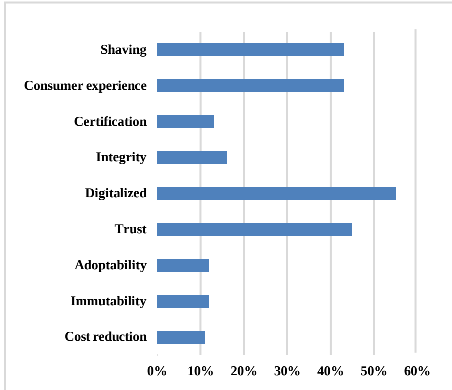


Fig.3 Top Subjects Organizers Tackle with Blockchain, % of Replies

Table 4 Enterprises Are Viewed as Pioneers in Blockchain

Financial services	**46%**
Industrial products and manufacturing	44%
Energy and utilizes	12%
Healthcare	47%
Government	13%
Retail and consumers	43%
Environment and media	26%

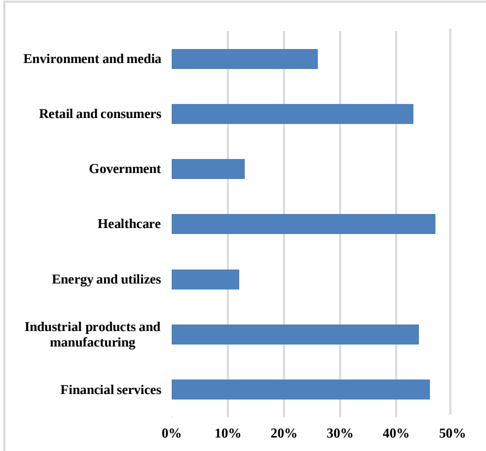


Fig. 4 Enterprises Are Viewed as Pioneers in Blockchain

The Internet of Things, or IoT, is becoming more and more popular in the blockchain industry. According to Statista [10], there are currently 50 billion IoT devices worldwide, which will pique engineers' curiosity even more.

There is a lack of understanding of the rules around digital currencies and blockchain. Many countries are still in the process of studying and discussing all things related to blockchain, especially in the field of finance. This led to a situation where regulations are still broken in the majority of the countries [11]. Organizations get the most from using blockchain innovation when they collaborate and create a shared stage, but at that moment, they are unable to define rules without asking other members to participate in their creation. If a company is building a blockchain only for itself, it will face challenges related to internal integration, information standardization, and scalability [12]. Everything is getting more and more complicated if more firms band together to create a blockchain. Blockchain is a fundamental innovation that will take a long time to integrate into our financial and social framework; it is by no means a problem invention that can provide a low-cost solution for customary plans of action. The pace of inventive, mechanical, and institutional change will accelerate, but the rate of acceptance will remain slow **[13]**.

V. CONCLUSION

Permissioned blockchain has evolved into a cutting-edge, practical solution for managing businesses effectively. The invention has the ability and potential to address problems related to the financial condition, such as centralization, cybercrimes, cyber control and irregularities, phantom laborers, seamless evaluation, and so forth. Decentralization, information integrity, accessibility, clarity, and room for information review will all be outlined by permissioned blockchain. We outlined the potential applications of blockchain technology in this study [14], as well as how it may be used to address financial problems. It is anticipated that future research will plan and evaluate a permissioned blockchain architecture for financial systems.

As this application is developed, enormous cost-cutting efficiencies—which are not included in this article—will outperform traditional finance expert businesses. The question will be whether the blockchain with this application will be given by the public authority, a group of industry players like IPPA, or at least one of the traditional finance expert co-ops. Financial transactions exhibit a distinct bias towards blockchain technology [15]. The financial industry is firmly headed in this direction, with a partnership between J.P. Morgan The result of pursuit and a sizable financial expert organization is almost a certain.

VI. FUTURE WORKS

A subset of representatives (perhaps chips) deprived to step forward and permit their finance to be put on a private blockchain, maybe in participation with J.P. Morgan Pursue.

- It would be fantastic to try writing the clever contracts for financial deductions, to work with the blockchain's immutable, incredibly durable record, and to approach public authority organizations to see if a workable blockchain finance model could be developed.
- After reading this piece about the benefits of working in the blockchain industry, it wouldn't be strange if you were thinking about making a permanent change. This emerging breakthrough is a remarkable choice if you want to upskill, become a blockchain engineer, or change your current career. Later on, blockchain technology will be a part of almost every industry, and successful professionals are constantly trying to stay up to date on the newest developments.

VII. REFERENCES

[1]. Baliga, A., Solanki, N., Verekar, S., Pednekar, A., Kamat, P., & Chatterjee, S. (2018).

[2]. Parizi, R. M., Dehghantanha, A., Azmoodeh, A., & Choo, K.-K. R. (2020). Blockchain in Cybersecurity Realm: An Overview. In Blockchain Cybersecurity, Trust and Privacy (pp. 1–5). Springer.

[3]. Polge, J., Robert, J., & Le Traon, Y. (2021). Permissioned blockchain frameworks in the industry: A comparison. ICT Express, 7(2), 229–233.

[4]. Quarm, R. S., Sam-Quarm, R., & Sam-Quarm, R. (2020). Exorcising the'Ghosts' from the Government Payroll in Developing Countries in the Wake of the COVID-19 Pandemic: Ghana's Empirical Example. Journal of Economics and Business, 3(4).

[5]. Aluko. (2020). IPPIS: ASUU kicks as FG sacks contract lecturers. Punch Newspapers.

[6]. Ante, L. (2021). Smart contracts on the blockchain – A bibliometric analysis and review. Telematics and Informatics, 57, 101519.

[7]. Antwi, M., Adnane, A., Ahmad, F., Hussain, R., Habib ur Rehman, M., & Kerrache, C. A. (2021). The Case of HyperLedger Fabric as a Blockchain Solution for Healthcare Applications. Blockchain: Research and Applications, 100012.

[8]. Baker, C., & Werbach, K. (2019). BLOCKCHAIN IN FINANCIAL SERVICES. FINTECH.

[9]. Sikorski, J. J., Haughton, J., & Kraft, M. (2017). Blockchain technology in the chemical industry: Machine-to-machine electricity market. Applied Energy, 195, 234–246.

[10]. Singh, A., Parizi, R. M., Zhang, Q., Choo, K.-K. R., & Dehghantanha, A. (2020). Blockchain smart contracts formalization: Approaches and challenges to address vulnerabilities. Computers & Security, 88, 101654.

[11]. Smits, M., & Hulstijn, J. (2020). Blockchain Applications and Institutional Trust. Front. Blockchain 3: 5.

[12]. Uzoh, B. C. (2020). ACADEMIC STAFF UNION OF UNIVERSITIES (ASUU) AND THE POLITICS OF INTEGRATED PAYROLL AND PERSONNEL INFORMATION SYSTEM (IPPIS) IN NIGERIAN FEDERAL UNIVERSITIES.

[13]. Wüst, K., & Gervais, A. (2018). Do you Need a Blockchain? 2018 Crypto Valley Conference on Blockchain Technology (CVCBT), 45–54.

[14]. Yaga, D., Mell, P., Roby, N., & Scarfone, K. (2018). Blockchain technology overview (NIST IR 8202; p. NIST IR 8202). National Institute of Standards and Technology.

[15]. Yoo, C. S. (2018). Paul Baran, Network Theory, and the Past, Present, and Future of the Internet. Colorado Technology Law Journal, 17(1), 161–186.

A Secure Blockchain System for Electronic Voting and Its Output

Swati Tripathi

Research Scholar

***Abstract*—** Voting electronically is also referred to as e-voting or electronic voting. With huge data, real-time, and high security qualities, it's an efficient and economical method of conducting voting. But in the context of electronic voting, worries about communication privacy and network security have emerged. In the domains of networking and communications, secure electronic voting has emerged as a major issue. We talk about ways to enhance the security of electronic voting by incorporating blockchain technology into a peer-to-peer network. We initially use distributed ledger technology (DLT) to create a synchronized model of voting data in order to prevent vote tampering. This study outlines an attempt to build a practical voting system using blockchain features like transparency and cryptographic underpinnings. The suggested solution provides end-to-end verifiability and complies with the essential standards for electronic voting systems. This article describes the proposed electronic voting mechanism and how it was implemented on the Multichain platform. The study provides a detailed analysis of the methodology, demonstrating its application in creating an electronic voting process that is verifiable from start to finish.

Keywords— *Electronic voting (e-voting), distributed ledger technology, network security.*

I. INTRODUCTION

A voting form projected onto a screen is a method for reaching a consensus or providing an alternative viewpoint to attendees of a gathering or social event. A voting form is projected after talks, debates, and political contests on a regular basis. When projecting a voting form, the resident projects a surveying structure for their selected opponent, and the person to be selected is the promising newcomer of a political choice. Generally speaking, the voter can assign a vote based on the list of newcomers or select the candidates he or she prefers. Casting ballots in a poll should be unsigned and done separately by each voter in a classified area to prevent other parties from discovering who is projecting the ballot. Projecting a polling form has been the standard tool used by current representatives to implement a vote-based framework since the seventeenth century. Similar to organizations, endeavors, and intentional associations, projecting a vote form is also utilized in a number of other private and social gatherings [1].

They could be caught on camera projecting that a voting form system shouldn't be noticeable. The majority rule structure ought to guarantee that each voter's vote was counted and that voting documentation was provided. It shouldn't be possible for one element to dominate structures by projecting a voting form system. Projecting a polling form is restricted to qualified personnel only. The cost of the political choice system shouldn't be high. Relying on the gig political choice system, individuals should only receive restricted induction. Blockchain Development's Application in E-As a blockchain is fixed and nonliteral, projecting a voting form structure can satisfy all of the aforementioned requirements.

A. *Key features of Blockchain*

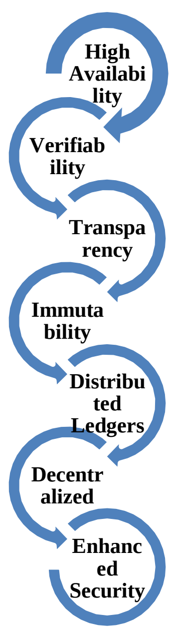


Fig. 1 Key features of Blockchain

Modules of projecting a voting form Structure Consistently it involves two models:

1. **The Administrator Module**

The chairman of the affiliation or the supported person is the intended user of the Association module. The executive is capable of overseeing all aspects of the balloter and the promising newcomer to the community. The Chairman can perform tasks such as evaluating the name, eliminating the name, revitalizing the name, and implementing a framework based on popularity. The manager is going to be ready to investigate whether there has been any vote tampering, and he will act decisively to prevent it **[2]**.

2. **The Client or Resident Module In this module**

A client or resident should be ready to see the names of all the candidates and cast their vote for the newcomer to the neighbourhood.

B. Security in Blockchain

To provide structure security, we incorporated ECC cryptography into the proposed blockchain architecture. The fundamental components of ECC cryptography are the public key and the secret key. Everybody has a split copy of the public key, which allows them to view all of the association's public information. It is, for security reasons, inextricably linked to the private key. These keys are useful for the processes of encryption and decryption. A public key is linked to a classified key, which has an appearance similar to a secret phrase. To provide security to a record, a classifiedkey is not given to anyone or is not transparently accessible to all. It is only known to the record holder and is kept confidential. Accounts are given believability exercises using the private key. In contrast to a regular record, in order to access and comprehend the details of the record or to doany significant actions, the beneficiary end's private key—known as the gatherer—must be used in order to provide security to a record through ECC computation. When source is connecting to gatherer, one can observe how public key and private key matching functions in the figure below. Initially, the source encrypts the message using the recipient's public key before securely sending the mixed message to the gatherer. On the other hand, when the encoded message is sent to the beneficiary, the recipient receives it and uses the transporter's secret key—which is referred to as the recipient in this context—to unscramble it. ECC cryptography is used to secure the communication as a result **[3]**.

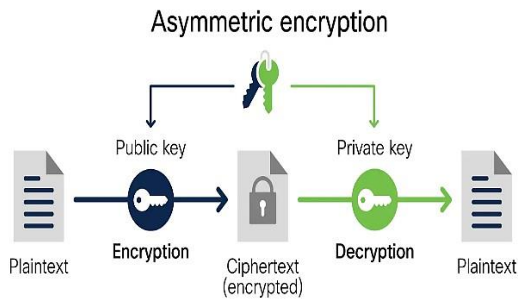


Fig. 2 Public and Private key Encryption and Translating process

C. Core Parts of Blockchain Design

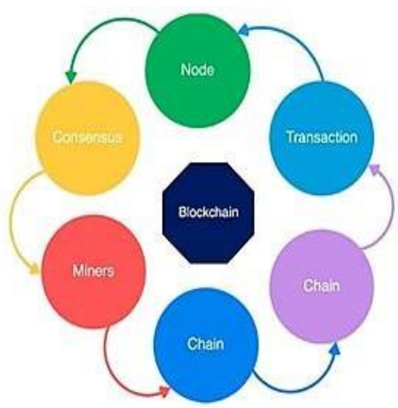


Fig. 3 Center Pieces of Blockchain Plan

- **Center point:** In a blockchain plan, clients' laptops (each device contains a backup copy of an all-out record from the blockchain);
- **Trade**: Records and subtleties are the most basic design block used by the blockchain system;
- **Block:** Trades across the association handed on to all center points are managed using a variety of data structures called blocks.
- **Chain:** A sequence of blocks in a specific request for proposals;
- **Backhoes:** Provide support for the trade and include that block in the blockchain architecture;
- **Understanding:** A set of guidelines and connections to complete blockchain operations.

D. How Blockchain Can Change the Electronic Popularity based System

Blockchain creation filled up gaps in the current system by making the reviewing section clear and understandable, ending illegal popularity-based marketing, bolstering data security, and thoroughly examining the study's outcomes. The implementation of the blockchain-based electronic vote-based approach is enormous. However, there are some fundamental hazards associated with electronic majority rule. For example, if the system is compromised, all votes cast are likely to be manipulated and mismanaged. Despite all of its anticipated benefits, electronic popularity-based has not yet been accepted on a large scale by the general public. Blockchain development is a practical way to overcome the hazards associated with electronic voting nowadays **[4]**.

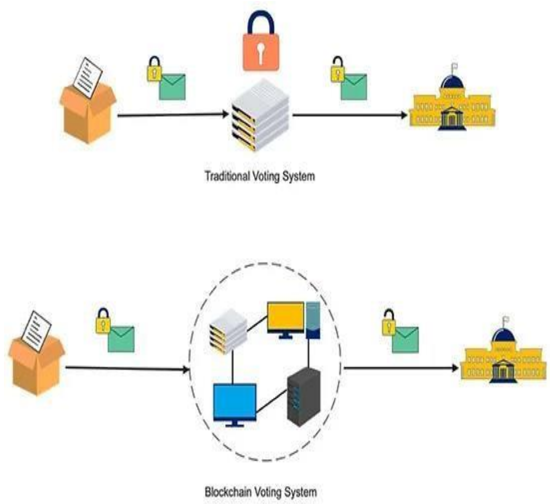


Fig. 4 Change the Electronic Popularity based System

II. RELATED WORKS

(Jaffal, R.; Mohd 2021) The various types of supply networks that exist now have developed into intricate relationships via late progress. Numerous obstacles confront the leader's structures in the creation network. These include the following: the downstream party (Client) not seeing the upstream party (Provider) as having discernible quality; the lack of adaptability despite sudden changes in the market and cost control; the appearance of dependence on security partners; and the improper arrangement of stock organization options **[5]**.

(Zhang, L.; Peng, M 2021) Clinical consideration information is vital for both expert associations and patients. Further safe exchange and backing of Electronic Clinical consideration Records (EHR) are necessary. EHR structures in clinical benefits have traditionally relied upon a brought together system (e.g., cloud) to interchange prosperity data between clinical benefits accomplices, which could reveal private and sensitive patient information. EHR has challenged to satisfy the needs of a handful of accomplices and systems with regards to security, detachment, and other regulatory aims. Blockchain is a ringed, decentralized record development that can give obtained, supported, and constant data sharing workplaces **[6]**.

(Kim, T.; Ochoa 2020) Supporting electric vehicles (EVs) is a strong way to combat the climate catastrophe and develop carbon non-partisanship. Drives in electric vehicle batteries and battery chiefs enthusiastically relate to government goals and customer experiences. This essay examines the challenges and developments surrounding (I) cutting-edge battery technology advancements and (ii) premium battery board advancements for hybrid and all-electric vehicles. Finding the essential elements, advantages and disadvantages, fresh, creative forward leaps, upcoming challenges, and entryways for promoting electric compactness is crucial **[7]**.

(Ometov, A.; Bardinova 2020) This document serves as a data descriptor for many of the battery yield data assessments made possible by the current diverse blockchain initiatives being implemented on Android devices throughout the turned-on display discharge process. The evaluations were carried out for arrangement estimations for confirmation of work (PoW) and confirmation of development (PoA). We include instances of Samsung World S9 movement in this description, however a larger portion of the evaluations are available within the collection. Models provide information on the temperature, condition, yield voltage, and current of batteries. We also present the evaluations obtained using cell (LTE) and short-range (IEEE 802.11n) relationships **[8]**.

(Hakak, S.; Khan 2020) The components that take into account the dispersed environment have become a popular option for strategies, even though they are no longer limited to a single area (i.e., crypto cash). Instead, it has influenced a number of bold attempts to support expansive insurance and security strategies, such as magnificent homes, insightful electrical systems, prudent landscaping, astute medical care, astute transportation, etc. These Smart Real Structures energetically rely on Internet of Things (IoT)-based marvels that entail a synchronized strategy of interdependent components for the seamless operation of the entire system **[9]**.

III. METHOD

The analysis of this article starts with the requirement to create an e-projecting voting form structure that is more practical and secure, since it has become a significant topic in the fields of information security and industry. DLT is the foundation of blockchain, which Satoshi Nakamoto created in 2008. Blocks are created once and stored on a blockchain. The hash regard of each block, with the exception of the chief block, is stored. Using neighbourhood, which is accepted to service the public trade record of the digital currency Bitcoin, it synchronizes the replicated records among several center points. In order to address the security of e-projecting a polling form, we describe approaches to leverage blockchain [10]. Having being unique and the first blockchain, the enhancements are as follows:

(1) To prevent vote spoofing, we intend to implement a synchronized model for projecting polling form data that takes DLT into account.

(2) Taking into account that ECC provides affirmation and non-repudiation, we design a client certificate model.

(3) We intend to implement a withdrawal model that permits voters to quickly amend their ballots. We present a blockchain-based e-projecting a voting form approach that satisfies the majority of e-projecting a voting form cycle requirements by arranging the aforementioned plans.

The following is how we approach the blockchain-based electronic voting form plot a role:

(1) The voting form system that is e-projected using blockchain technology is decentralized, open, and circular. Votes cast by voters from different computers and cells can be recorded by it.

(2) The inhabitants can survey and validate the votes at a reasonable cost thanks to the blockchain-based e-projecting polling form design.

(3) The vote database is openly monitored and uses a coursed timestamp server on a cherry on top sidekick association.

(4) The process of settling on blockchain eliminates the large repeatability of e-projecting a polling form and places residents' concerns about data security to the background.

Considering the above demarcation, the plan is shown in Fig. 1 and is organized as follows:

(1) Ballot casting blockchain: this is an evolving list of ballot casting blocks.

(2) Citizens: An elector is a person who projects a voting form for their chosen candidate. The voter has two options: pull out their ballot or cast one.

(3) Ballot casting: this refers to the act of casting a ballot. It has the ability to verify votes, inquire about votes, and challenge a citizen's public key.

The Public Key Infrastructure (PKI) is a collection of systems that manage public-key encryption.

(5) The vote data set is a collection of votes that are updated via the voting process.

(6) Diggers: their responsibility is to record votes that have been acknowledged and upload them to the public democratic blockchain **[11]**.

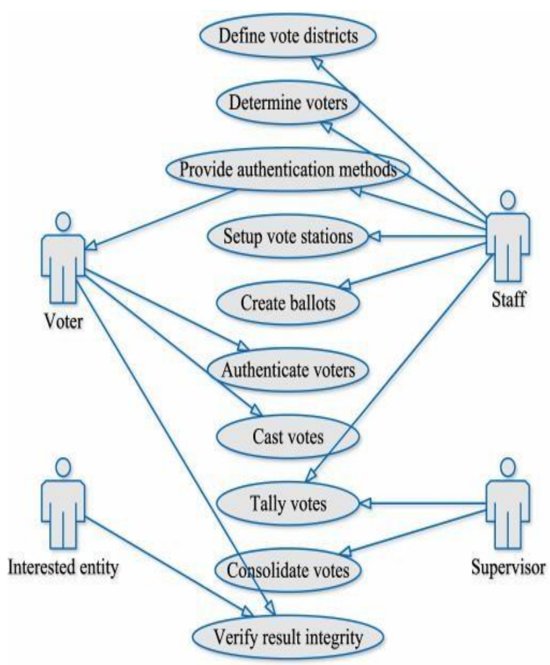


Fig. 5 The Use of the Blockchain Development in projecting a polling form Structures

IV. RESULT AND DISCUSSION

The continuing associations and affiliations that have been established but mostly defined in recent years are fostering the popularity-based region. All of the proposals are good starting points for the blockchain association to combine transparency.

Extending the block size or reducing the hash difficulty are insufficient methods for scaling a blockchain. [12] Every system has a limit before it can complete the trades required to match associations. Take Visa, for example, which processes 150 million trades a day on average. According to research presented by Farewell Correspondences in 2018, 44% of the associations employed blockchain in their review, which suggests that there are universal problems resulting from the application of new technologies.

Table 1 Assessment of current blockchain-based electronic majority rule structures.

Online Voting Platforms	Framework	Language	Cryptographic Algorithm
Follow My Vote	Bitcoin	12.1%	4.22%
Voatz	Hyperledger Fabric	14.2%	7.33%
Polyas	Private/local Blockchains	4.23%	12.4%
Luxoft	Hyperledger Fabric	4.23%	4.24%
Polys	Ethereum	4.2%	1.23%

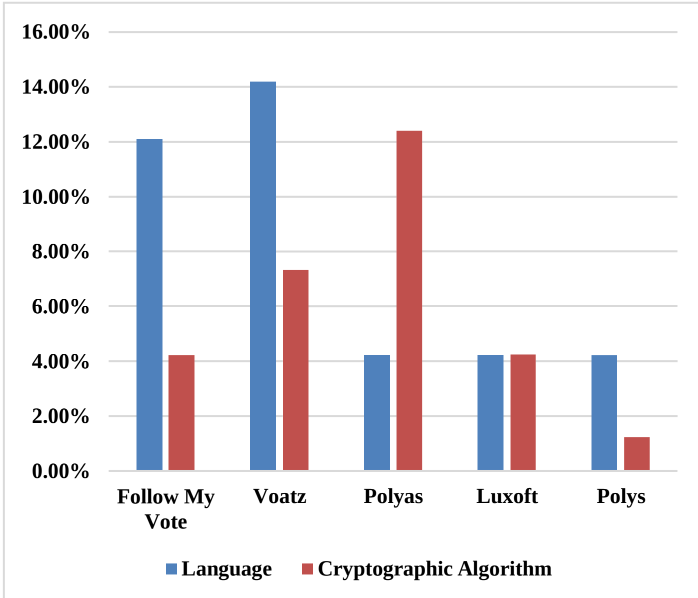


Fig. 6 Assessment of current blockchain-based electronic majority rule structures

Table 2 Flexibility examination of well-known blockchain stages.

Framework	Year Release	Generation Time	Hash Rate
Bitcoin	2022	899.624 Th/s	14
Ethereum	2014	168.59 Th/s	42.2
Hyperledger Fabric	2019	NA	14.5
Litecoin	2020	1.307 Th/s	45.3
Ripple	2021	NA	75.2
Dogecoin	2013	1.4 Th/s	4.55

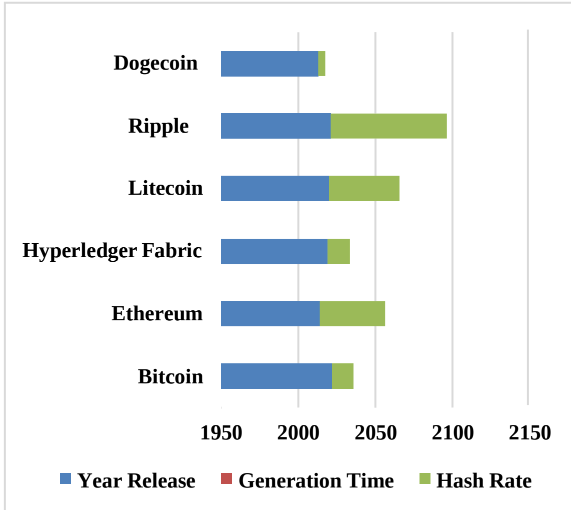


Fig. 7 Flexibility examination of well-known blockchain stages

Table 3 Assessment of picked electronic popularity based plans considering blockchain.

Have voted before	Experience challenges with the material	Thanks, IFC knowledge	Thanks, a south Africa should adopt in e-voting	Find a e-voting appearing
15	44	52	71	25
33	14	25	36	41

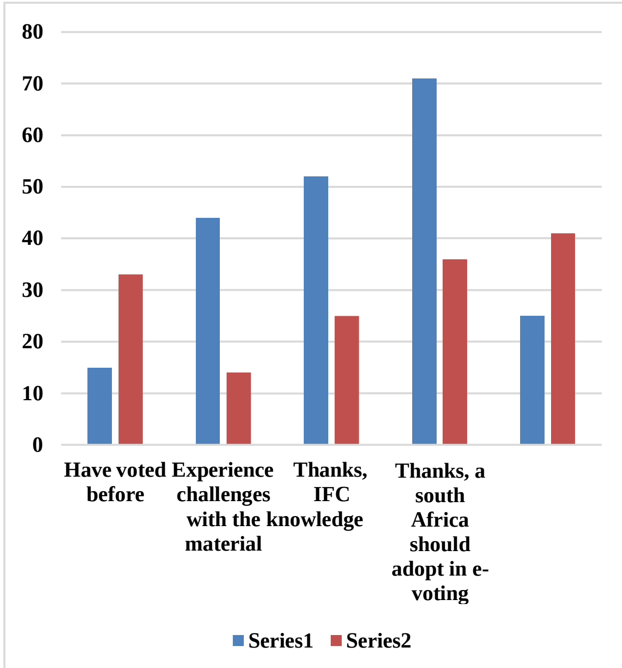


Fig. 8 Assessment of picked electronic popularity based plans considering blockchain

The popular region is being cultivated by the ongoing relationships and affiliations that have been essentially developed over the last few years. They're all good starting points for the blockchain association to try transparency [13].

We suggest a blockchain-based method for e-projecting a voting form that satisfies all of the essential requirements for collaborative e-projecting of a polling form. Block by block, every vote on the blockchain is cryptographically linked. When two blocks have the same timestamp, the block with the higher imprint value is chosen **[14]**.

V. CONCLUSION

Since the 1970s, electronic vote-based systems have been utilized in moving designs. These systems offer several advantages over paper-based ones, including more capability and fewer mistakes. As blockchain technology has advanced dramatically, there has been a push to investigate if blockchain technology may be used to support a workable method for e-projecting a vote form. One such attempt, which leverages blockchain's advantages—such as its cryptographic underpinnings and simplicity—to successfully e-project a vote form is described in this work. The suggested method was implemented using Multichain, and an overall evaluation of the method includes its adequacy in meeting the main requirements for an electronic voting form scheme **[15]**.

VI. FUTURE WORKS

- The blockchain-based e-projecting polling form system can be used for a variety of applications and majority rule scenarios. Even while blockchain technology is secure, it employs ECC public key encryption, which is resistant to quantum computing assaults. Blockchain technology that has defences against quantum computing threats is therefore a future evaluation point.
- This is being worked on as an extra provenance layer to support the present blockchain-based system.

VII. REFERENCES

[1]. Yaga, D.; Mell, P.; Roby, N.; Scarfone, K. Blockchain technology overview. arXiv 2019, arXiv:1906.11078.

[2]. Gao, S.; Zheng, D.; Guo, R.; Jing, C.; Hu, C. An Anti-Quantum E-Voting Protocol in Blockchain with Audit Function. IEEE Access 2019, 7, 115304–115316.

[3]. Pawlak, M.; Poniszewska-Marańda, A. Implementation of Auditable Blockchain Voting System with Hyperledger Fabric. In International Conference on Computational Science; Springer: Berlin/Heidelberg, Germany, 2021.

[4]. Jaffal, R.; Mohd, B.J.; Al-Shayeji, M. An analysis and evaluation of lightweight hash functions for blockchain-based IoT devices. Clust. Comput. 2021.

[5]. Zhang, L.; Peng, M.; Wang, W.; Jin, Z.; Su, Y.; Chen, H. Secure and efficient data storage and sharing scheme for blockchain—Based mobile—Edge computing. Trans. Emerg. Telecommun. Technol. 2021.

[6]. Kim, T.; Ochoa, J.; Faika, T.; Mantooth, A.; Di, J.; Li, Q.; Lee, Y. An overview of cyber-physical security of battery management systems and adoption of blockchain technology. IEEE J. Emerg. Sel. Top. Power Electron. 2020.

[7]. Ometov, A.; Bardinova, Y.; Afanasyeva, A.; Masek, P.; Zhidanov, K.; Vanurin, S.; Sayfullin, M.; Shubina, V.; Komarov, M.; Bezzateev, S. An Overview on Blockchain for

Smartphones: State-of-the-Art, Consensus, Implementation, Challenges and Future Trends. IEEE Access 2020, 8, 103994–104015.

[8]. Hakak, S.; Khan, W.Z.; Gilkar, G.A.; Imran, M.; Guizani, N. Securing smart cities through blockchain technology: Architecture, requirements, and challenges. IEEE Netw. 2020, 34, 8–14.

[9]. Çabuk, U.C.; Adiguzel, E.; Karaarslan, E. A survey on feasibility and suitability of blockchain techniques for the e-voting systems. arXiv 2020, arXiv:2002.07175

[10]. J. Cao, Y. Ding, L. Jiang, et al., A new proxy electronic voting scheme achieved by six-particle entangled states[J]. Int. J. Theor. Phys. 57(3), 674-681 (2018)

[11]. L. Zhang, Z. Zhang, C. Xie, A Choreographed distributed electronic voting scheme[J]. Int. J. Theor. Phys. 57(9), 1-11 (2018)

[12]. E.Ahene,C.Jin, F. Li, Certificateless deniably authenticated encryption and its application to e-voting system [J]. Telecommun. Syst. 70, 1-18 (2018)

[13]. Pop, T. Cioara, M. Antal, et al., Blockchain based decentralized management of demandresponse programs in smart energy grids:[J]. Sensors. 18(1), 162 (2018)

[14]. enry Rossi Andrian, Novianto Budi Kurniawan, Suhardi,” Blockchain Technology and Implementation: A Systematic Literature Review”.2018 International Conference on Information Technology Systems and Innovation (ICITSI) October 22-25,2018.

Blockchain-based Secure Real Estate Information Management

Vivek Sharma
Research Scholar

***Abstract*—** Lately, the amount of blockchain innovation distributions in the real estate industry has increased. Furthermore, according to a recent analysis, blockchains have the potential to significantly improve the housing market even in current half-breed environments, where they are frequently added as a supplement to pre-existing frameworks. This research provides an extensive, up-to-date account of the benefits, limitations, and possible applications of blockchain innovation in the real estate industry. The land company might benefit from blockchain in four areas, according to hypothetical blockchain writing: land organization, land exchanges, tokenization, and land the executives. Furthermore, using subject matter inquiry, 26 genuine applications were found, all but one of which had a role in landing the board. This essay explores the possible applications of blockchain technology in the real estate industry, outlining both its advantages and disadvantages. The article ends with recommendations for additional study.

Keywords*— *blockchain, blockchain technology, real estate, real estate management.

I. INTRODUCTION

The field of land writing has given increasing attention to blockchain innovation, especially in the last five years. Recent research reveals that, despite the current constrained expansion, blockchain applications may provide some additional benefits to the land region. In these scenarios, blockchain essentially functions as an alternative to conventional structures. Despite the hype surrounding blockchains, this scientific analysis shows that grasping the full possibilities and complexities of blockchain technology for the land region is a challenging task. The geological region represents over 60% of the world's abundance, 37% of global energy-related CO_2 outflows, and 36% of total last energy usage, making it of enormous social, environmental, and economic significance. In this article, "land" refers to a parcel of land or a body of water that includes easements, advantages, and registered designs. Land organization and land business are included in this definition of land area, together with their subsectors, such as land improvement, land hypothesis, property commerce, leasing, land organization, and land support [1].

Numerous articles about blockchain innovation in the land domain have been published. Specialized concepts are presented in a significant amount of the disseminated writing. Furthermore, the great majority of the analyses have only focused on one of the land subsectors—tokenization or land organization, for example. Others have examined blockchain's opportunities and challenges in the real estate sector from a singular perspective, such as a legal one. Generally speaking, there aren't many effective assessments that provide a thorough and current overview of blockchain possibilities for the whole geographical region.

A. Blockchain

A blockchain is a distributed ledger that records all transactions in an easy-to-use and durable manner. Any type of financial information, an electronic event, or a crucial informative index trip can all be considered exchanges. The main building block of a blockchain is a block. Many trades may be recorded in a block, which are stored on the blockchain by the majority of interested parties. Furthermore, the idea of blockchain permanence suggests that any information recorded on the network is unchangeable and secure; it is unquestionable. A refined picture of a common blockchain is presented in Figure 1, where each block is made up of two segments: a header segment that stores the hash reference of the previous block, and a data region that contains trades. A hash pointer serves two purposes: first, it stores the previous block's area, ensuring that all blocks are related and identifiable; second, it verifies that the pointers are hash pointers, meaning that it does more than just store the previous block's area; it also verifies that the hash values depend on the items in the block [2]. adjustments made to the things by a vengeful client will necessitate adjustments in every subsequent block, making it difficult to change the items in a blockchain whenever it is recorded. As a result, changelessness is achieved.

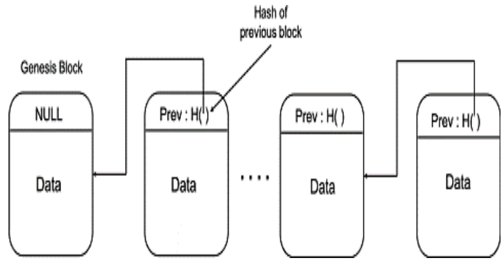


Fig. 1 A worked visible of a blockchain

B. Real Home Administration Framework

This Territory The software application designed for land organizations is called the board Framework. Nowadays, the majority of land companies continue to manage information using outdated methods like books or document frameworks (succeed sheets). This project aims to support the development of an expert data base for the land business, executive clientele, and property postings [3].

This assignment is the result of my designing endeavor; we have completed numerous parts such as section/search, match, property director, and reports. Age in light of the needs of the client.

Fig.2 Real Home Administration Framework

C. Architecture of the Proposed Framework

Our framework, which has four layers as shown in Fig. 3, is designed using a layered method. Each layer's utility is determined by the following:

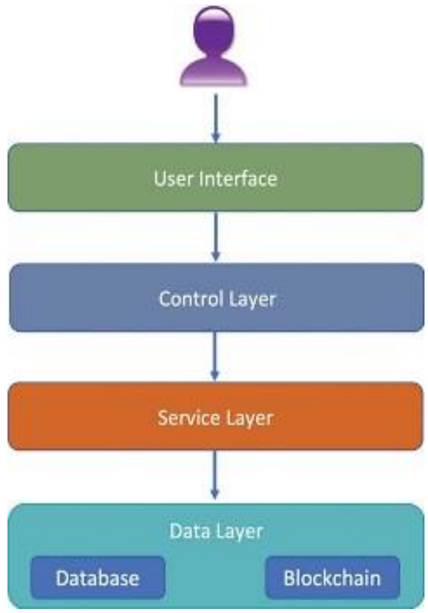


Fig. 3 Architecture of the Proposed Framework

D. User Connection point Layer

The front-end framework, or UI layer, is accessible to all clients and allows them to operate together with other sub-frameworks. Depending on the client's honor level, different functionalities are apparent to the client. The main purpose of the UI layer is to provide a channel of communication between the various sub-frameworks and the end user.

E. Control Layer

Access control is the responsibility of the control layer, which ensures that users can carry out tasks according to their degree of integrity. For example, a user who isn't the owner cannot be allowed to choose how much a resource costs.

F. Service Layer

Between the control layer and the information layer, the help layer functions as a delegate. It covers business rationale rules and functions similarly to the business rationale compartment [4]. Additionally, the administration and control layers guard against customers receiving information immediately.

G. Data Layer

The information layer functions as the framework's overall data set and includes a blockchain. The data set is used to hold client credentials, such as passwords and usernames **[5]**.

II. RELATED WORKS

(H. Wang, S. Ma 2020) Open banking (OB) is the pattern that monetary establishments use for administrative development and reconciliation. In order to discover the optimal arrangements and enhance customer interactions, OB gives outside specialist cooperatives (TSPs) access to client financial data. The expectation of external trust for the accomplishment of the OB biological system raises concerns regarding advanced character reconciliation, information exchange, and conservation of protected areas **[6]**.

(Y. Yan, Y. Dai 2020) The Internet of Things (IoT) is a rapidly evolving field, but as of yet, no established framework exists to protect customer security, secure data transfer between different IoT domains, or enhance edge handling. As a result, we coordinate blockchain innovation to advance secure IoT stages. That being said, the challenges posed by cumbersome registration procedures prevent blockchain from being used in the Internet of Things. **[7]**

(S. L. Cichosz, M. 2019) Patients with diabetes suffer from abnormal blood glucose levels, which can lead to various health problems affecting their kidneys, heart, and eyes. Due to theseconditions, people with diabetes have typically used Self-Checking of Blood Glucose (SMBG) techniques, such as regularly puncturing their fingertips, to check their blood glucose levels. Thesedifficulties may be overcome with the use of a device known as the Endless Glucose Screen (CGM), which measures blood glucose levels continuously throughout the day without requiring the patient to prick themselves **[8]**.

(A. Dorri, S. S. 2019) Blockchain has garnered enormous attention recently because to its noteworthy features, which include auditability, permanence, security, and secrecy. Blockchain has been used in several non-financial applications, such as the Web of Things (IoT), as a result of these noteworthy features. Regardless, blockchain incurs significant processing costs, limits its adaptability, and results in significant data transmission overheads and delays, making it unsuitable for the majority of Internet of Things applications **[9]**.

(P. P. Ray, D 2020) In light of the Web and traditional broadcast communications organizations, the Internet of Things (IoT) functions as a data transporter. One important component of the assistance design that can truly resolve IoT security concerns is access restriction. As a result, various preparations are made for the entrance control to address certain security concerns. However, access control in the IoT can only be effectively communicated with great effort using these current setups. As a distributed ledger technology that is decentralized, immutable, and incredibly reliable, blockchain aligns with the most recent advancements in organizational development **[10]**.

III. METHODOLOGY

the primary assessment. The basis for this assessment was a detailed summary that compiles, examines, and summarizes the known and unknown information on a "specific practice-related topic." The following are the means of exact reviews.

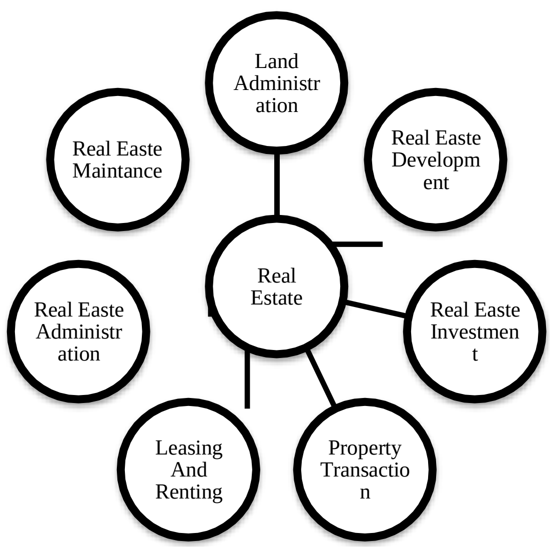


Fig. 4 The calculated construction of the land subsectors utilized in arranging the blockchain writing with land merits

A. Blockchain technology

When introducing Bitcoin in 2008, Satoshi Nakamoto made the initial proposal for the blockchain breakthrough. Since then, it has consistently attracted the attention of states and other large-scale projects. As a result, blockchain innovation has rapidly developed and being used in several industries. The blockchain is essentially a shared data set that documents all transactions or digital events that have been carried out and distributed among participating parties.

B. Hyperledger **Texture**

The Linux establishment launched a project called Hyperledger Texture, which can be used as the foundational blockchain platform. It employs a specific engineering with pluggable agreement conventions, main areas of strength for upholds and personality highlights, and is open-source and compliant with standards. It also uses ring-characterized shrewd contracts.

C. Real home exchange stage in light of blockchain innovation

This section will elaborate on the most effective way to build a land exchange platform considering blockchain technology [11]. Considering the Hyperledger Texture stage, this step is designed. This stage's main skill is to comprehend the delivery and exchange of land exchange data. It may also call into question sincere conversations.

D. System design

Figure 1 depicts the progression of each member's land exchange and data delivery during the land exchange period.

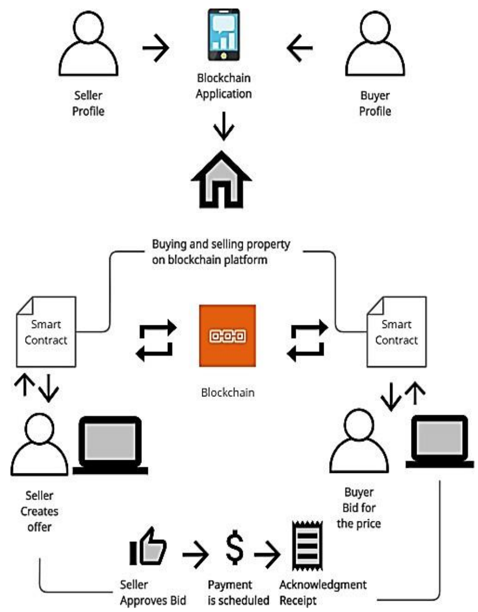


Fig. 5 Land exchange stage in light of blockchain innovation

Although several blockchain gathering proposals may have been classified under many land subsectors, the plan was selected based on the subsector with the most notable emphasis. The layout of the land commercial language implications upgrades the plan. The ruling also allows us to precisely address the question of whether the language guarantees advantages related to blockchain-based land. The distributions that looked into blockchain prospects for a variety of land subsectors without a specific concentration were combined under the eighth order, land business **[12]**.

IV. RESULTS AND DISCUSSION

The section that goes with it provides an analysis of the writing, broken down into the year of distribution, the areas of the primary writers, the type of distribution, and the crucial distribution field. The section also arranges the documents according to paper type and land subsectors.

A. Bibliometric and Record Examination

Over the past five years, the subject has received more attention, as evidenced by the writing presented by dispersing quite a bit in Figure 2. Even with all other factors being equal, observational publications continue to be in the minority. Figure 6 illustrates how the problem has gained international attention. The locations of the top makers address more than 50 countries. The United States, India, the Assembled Realm, Italy, Australia, and Russia are the five most unusual countries. Remarkably, despite India's circulations exhibiting an upward trend, the US distributed the greatest papers in 2018 and 2019 **[13]**.

Table 1. Writing by distributing year and chief writers' areas

Literature by publishing year	Countries	percentage
2013	India	44
2014	UK	42
2015	Russian	52
2016	China	22
2017	Sweden	43
2018	USA	56
2019	Australia	42
2020	Italy	55
2021	Netherlands	43
2022	Malesia	12

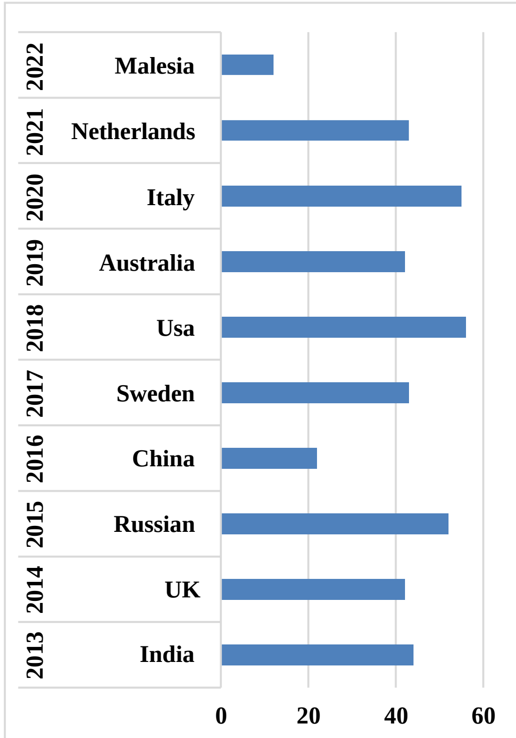


Fig. 6. Writing by distributing year and chief writers' areas

With an 8% offer, land was the second most obvious catchphrase topic. Land included references to several types of land (lodging, business, rental), as well as the land market and business. Land organization was the third-largest group of catchphrase topics, and it also included watchwords like "land vault," "enrollment," and "cadaster." **[14]**.

Table 2 Writing's ten most-frequently utilized writer watchwords

Keywords	Number of occurrences
Blockchain	4
Smart contract	12
Real easte	5
Ethereum	8

Land registry	12
Distributed ledger	4
Land registration	8

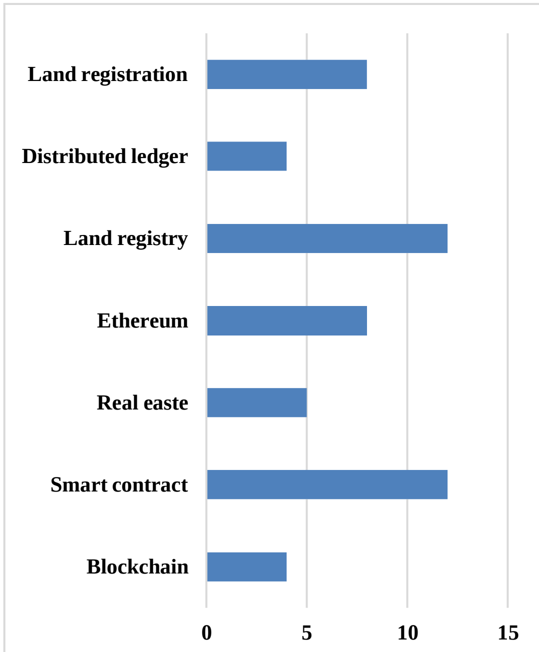


Fig .7 Writing's ten most-frequently utilized writer watchwords

A. Discussion

This section looks at the results and provides answers to the test questions. At that time, the legitimacy of the review and its obstacles are discussed in this section. The distribution pattern suggests rising awareness of blockchain innovation in the geographical region on a global scale. The ongoing analysis found that by referencing current issues that blockchain is supposed to address or its straightforwardness, the text regularly justifies blockchain adoption in the land field **[15]**.

V. CONCLUSIONS

The goal of this analysis was to present a contemporary understanding of the benefits and drawbacks of implementing blockchain technology in the real world, as well as the reasons behind doing so. This analysis addressed these questions by assessing 296 noteworthy reports and conducting a topical ingredient research. Using normalized land lingo, the distinguished blockchain reception proposition and papers were described. Strangely, blockchain within land organizations was the main topic of discussion for the great majority of the writing.

By removing the main flaws in the system, the blockchain-based framework is able to handle the numerous cases of land proprietorship transfers at the value of the participating parties of each and every transaction inside the company. This can potentially promote accountability, expertise, and productivity within the sizable real estate sector.

Academic writing has struggled to keep up with the rate of growth in blockchain innovation. Furthermore, not every legitimate application ends up being disseminated and examined in the text. Furthermore, given that it was impractical to include smart city and IoT blockchain writing, which would likely involve ties to the land the board classification, the land the executives classification may have been underripe-detested in this examination.

VI. FUTURE WORKS

There are several options for upcoming exams. It is still very difficult to evaluate the financial viability of blockchain applications in the land field because the observational experiences to now have provided very little insight. Blockchain architectures have numerous partners and are very confusing.

- Future observational blockchain research should aim to quantify, measure, and effectively communicate the benefits of real-world blockchain implementations. In addition, the effects, including any potential power reallocation effects, should be examined at the level of the biological system.
- More multidisciplinary research on the possible effects of blockchain technology should focus on the land domain, encompassing the expert viewpoint and extending it with, for example,

political, legislative, monetary, ecological, authoritative, legitimate, educational, and behavior viewpoints.

VII. REFERENCES

[1]. J. Wu, M. Dong, K. Ota, J. Li and W. Yang, “Application-aware consensus management for software-defined intelligent blockchain in IoT,” IEEE Network, vol. 34, no. 1, pp. 69–75, 2020.

[2]. M. Du, K. Wang, Y. Liu, K. Qian, Y. Sun et al., “A three-dimensional blockchain architecture for IoT security,” IEEE Wireless Communications, vol. 27, no. 3, pp. 38–45, 2020.

[3]. Ankit Mittal, Bhavyansh Sharma, Pinku Ranjan, IP(2020). Real Estate Management System based on Blockchain

[4]. R.C. Suganthe;N. Shanthi, R.S. Latha, K. Gowtham, S. Deepakkumar, R. Elango, IP (2021). Blockchain enabled Digitization of Land Registration

[5]. Meghali Nandi, Rajat Kanti Bhattacharjee, Amrit Jha, Ferdous A. Barbhuiya, IP(2020). A secured land registration framework on Blockchain

[6]. H. Wang, S. Ma, H. Dai, M. Imran and T. Wang, “Blockchain-based data privacy management with Nudge theory in open banking,” Future Generation Computer Systems, vol. 110, pp. 812–823, 2020.

[7]. Y. Yan, Y. Dai, Z. Zhou, W. Jiang and S. Guo, “Edge computing-based tasks offloading and block caching for mobile blockchain,” Computers, Materials & Continua, vol. 62, no. 2, pp. 905–915, 2020.

[8]. S. L. Cichosz, M. N. Stausholm, T. Kronborg, P. Vestergaard, O. Hejlesen et al., “How to use blockchain for diabetes health care data and access management: An operational concept,” Journal of Diabetes Science and Technology, vol. 13, no. 2, pp. 248–253, 2019.

[9]. Dorri, S. S. Kanhere, R. Jurdak and P. Gauravaram, “LSB: A lightweight scalable blockchain for IoT security and anonymity,” Journal of Parallel and Distributed Computing, vol. 134, pp. 180–197, 2019.

[10]. P. P. Ray, D. Dash, K. Salah and N. Kumar, “Blockchain for IoT-based healthcare: Background, consensus, platforms, and use cases,” IEEE Systems Journal, pp. 1–10, 2020.

[11]. Salman Humdullah, Siti Hajar Othman, Muhammad Najib Razali, Hazinah Kutty Mammi,I (2020). Secured Data Storage Framework or Land Registration using Blockchain Technology.

[12]. Ishita Mishra, Supriya, Astha Sahoo, M. Vivek Anand, I(2020). Digitalization of Land Records using Blockchain Technology

[13]. Archana Sahai, Rajiv Pandey, I (2020). Smart Contract Definition for Land Registry in Blockchain

[14]. Sai Apurva Gollapalli, Gayatri Krishnamoorthy, Neha Shivaji Jagtap, Rizwana Shaikh, I(2020). Land Registration System Using Block-chain

[15]. Fahim Ullah, Fadi Al-Turjman. \"A conceptual framework for blockchain smart contract adoption to manage real estate deals in smart cities\" , Neural Computing and Applications, 2021.

Blockchain-enabled Secure Mobile Financial Transactions

Suresh Kumar

Research Scholar

***Abstract*—** Blockchain is one of the most popular technologies for securing data transfer via decentralized peer-to-peer networks, and it has garnered a lot of attention recently. Blockchain technology provides safe, decentralized transaction processing because it is an unchangeable record. This intricate yet safe technology has grown clientele and a good reputation. Notwithstanding much enthusiasm, there are still obstacles facing the blockchain system. This article describes a system that offers a standardized framework for mobile payments using blockchain technology. The efficiency and security of payment transactions could be raised by implementing blockchain technology in a variety of healthcare settings. Consequently, the only people who can view or alter patient medical information are those who have access to it. When standard data storage and security technologies are employed, there don't seem to be any workable solutions for handling the massive amounts of data produced by the fog computing architecture. Lastly, the difficulties and potential avenues for future study for more secure and scalable blockchain systems for widespread use are emphasized.

Keywords*— *Blockchain-enabled, blockchain technology, Mobile Financial Transactions, normal data storage and security technologies.

I. INTRODUCTION

Blockchain technology provides trusted transactions through external parties within the company as a cryptography-based transmitted record. The Bitcoin blockchain was introduced in 2008, and several other blockchain frameworks, such as Ethereum, are available both publicly and privately. Globally, blockchain awareness has grown and has a significant impact on society. It has had an impact on global monetary systems, affected the growth of financially motivated cyberattacks, such as the disavowal of administrations and the emancipation of products (DoS), and been economically dependent. The main feature of blockchain innovation is its attractive flexibility for most corporate domains, including finance, tactics, smart investments, the pharmaceutical industry, and network security. Every component receives an updated copy of the encoded record, giving them the authority to accept any fresh transaction. At its core, blockchain technology is a distributed database of public or recorded information about all transactions, online events, and partner-shared activities [1]. This incredibly successful framework has a precise, accurate record of each and every previous transaction. "Conveyed record innovation" and "blockchain" are two terms that work well together. As shown in Fig. 1, every block consists of a number of exchanges that are carefully approved by the owner and verified by the reset clients before being added to the block.

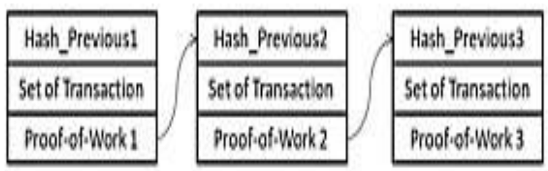


Fig. 1 Designs of Blockchain

A. Trends and Related Exploration

Clarifies the challenges associated with integrating security into unified applications and gives us a thorough analysis of blockchain methods for a security administrations application in a number of areas, including privacy, respectability confirmation in the circulated network, access control, and verification. Numerous research domains can utilize blockchain technology to avoid unified substances, such as cloud, Web of Things (IoT), and Bigdata. Although scientists acknowledge that blockchains have properties that will be used in finance, there are currently no practical uses for blockchains in popular culture.

A. Proposed Solution Framework

This section clearly adds more complexities to the suggested structure.

Given that payments are a remarkable source of revenue for financial institutions, they expand the use of digital money, or computerized monetary standards, to meet the needs of the emerging online consumer and commerce era. From this point forward, monetary organizations must benefit from these new developments and reap the gains, since various oversight authorities and control sheets have recognized the significant status that cryptographic forms of money have gained. Fig. 2, which is shown below, illustrates how the suggested configuration functions **[2]**.

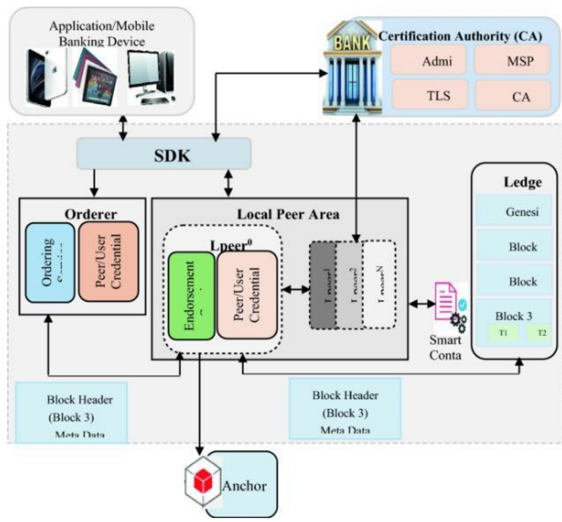


Fig. 2 New portable instalment utilizing blockchain structure

B. Process Security

One of the most robust blockchains, Ethereum, will be handled by the blockchain as a private organization for our financial arrangement. The mobile application will use the LMK to perform an AES-256 calculation in order to jumble the exchange message. Password security is crucial since they are a necessary method of authentication. As a result, there is no method to obtain the initial secret word for the circumstance when hashing is given to complete such an errand. Our computation may be able to detect and prevent such attempts **[3]**.

II. RELATED WORKS

(M. Gupta, V. P. 2022) This research presents a fruitful picture encryption technique based on a hybrid of cryptography and watermarking techniques. The reliable and error-free transfer of images between IoT-enabled devices depends on two-level security. At the primary level, a watermarking scheme based on discrete wavelet changes (DWT) is employed. At a secondary level, a successful image encryption technique based on a combination of hybrid and computed turbulent guidance is implemented **[4]**.

(E. M. Onyema 2022) Exam interest is growing in the application of AI computations for gaze recognition and patient observation. In this review, we introduce a convolutional brain organization (Community) approach for look acknowledgment in light of profound learning computation. Data were taken from the FER2013 dataset, which includes examinations of seven widely used looks for getting ready **[5]**.

(S. Stalin, V 2021) The electroencephalogram (EEG) signals are a significant source of information that are frequently contaminated by remnants of movement. Since human brain illnesses require a strong neurological indicator for conclusion and inquiry. The eradication of the EEG remnants is therefore an important step. Support vector machine (SVM) is used in this research work to identify the primary movement curio from a single channel EEG data, followed by the concealing of further relics. The reflection of the sign highlights are completed by group exact mode disintegration (EEMD) calculation **[6]**.

(D. Jain, P. K. Shukla 2021) Our world has changed as a result of the ubiquitous recognition, growth, and development of the Web and mobile technologies. On the other hand, the globe is witnessing and experiencing wrongdoing tactics that are creatively promoted. These risks—which include, but are not limited to, disruptions and hacking—are security experts' main concerns. Furthermore, there is a clear correlation between the benefits of science and the challenges faced by effective interruption detection techniques. The main goal of this study is to present a host-based interruption recognition framework that makes use of a C4.5-assembler in relation to the renowned Combined Tree Development (CTC) computation, which performs well in the presence of class-imbalanced data **[7]**.

III. METHOD

In light of the Web of Things, Figure 2 illustrates a possible healthcare information security setup that may operate at the hazy layer of a distributed computing viewpoint. It makes use of blockchain innovation for public approval and an electronic computerized signature as security measures. This is accomplished by creating the testament hash using the SHA-256 secure hashing technique, which is connected to the circular bend cryptography advanced mark. A vital component of bitcoin security is frequently employed to protect information-gathering apps and various types of web-based transactions, such as e-commerce. Flaws in standard WSN-IoT security countermeasures necessitated the introduction of blockchain innovation, resulting in increased computational complexity, memory issues, processor power consumption, inactivity, and vulnerability to social attacks, among other issues [8]. Among other benefits, the plan called for the acceptance of a hash calculation in the blockchain to guarantee security, permanence, and transparency among chain participants. This elliptic bent cryptography hashing is several times faster, more efficient, and more secure than RSA, which uses a vital size of 256 pieces (which is similar to the 2048 and 3072 pieces used in RSA), according to the experts. In particular, the fall impact surpasses arbitrary number age in terms of hashing proficiency.

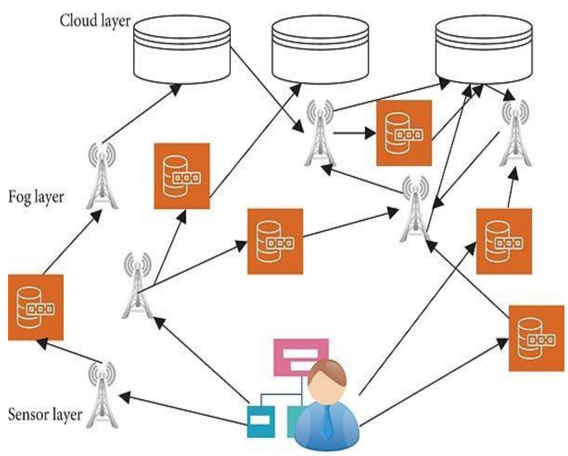


Fig. 3 Generally speaking layer structure

Alternatively, experts may be able to access jumbled sent record information by using a crypto hash-based blockchain and a secret key that has been distributed among the fog levels. This communication is instantly and safely stored in the cloud, where it will be available for long-term maintenance and recovery. The general stream process, the inward cycle exchange for the suggested worldview, and the general stream process are also depicted in Fig 3 **[9]**.

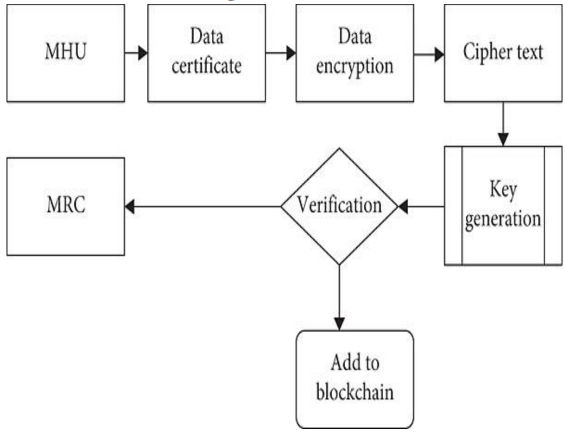


Fig. 4 Input information in the layer stockpiling

IV. RESULTS AND DISCUSSION

The meeting's plan was applied in this review. The meeting was the most appropriate examination technique since it is possible to obtain a vast amount of data to determine the current state of Blockchain-related business reception, promotion, and improvement [9].

In order to gather the necessary information from the professionals and experts associated with blockchain, 100 objective interview subjects were greeted; sixteen of them acknowledged the greeting and were willing to share their knowledge and insights.

Table 1 Engaging data of interview members

Participant	Age	Gender	Sector	Time worked in the R&D or finance team	The country where the team locates
R1	24	Female	Research	2	UK
R2	25	Male	IT	4	India
R3	23	Male	Education	2	Australia

R4	12	Female	Finance	3	China
R5	26	Male	Marketing	4	USA

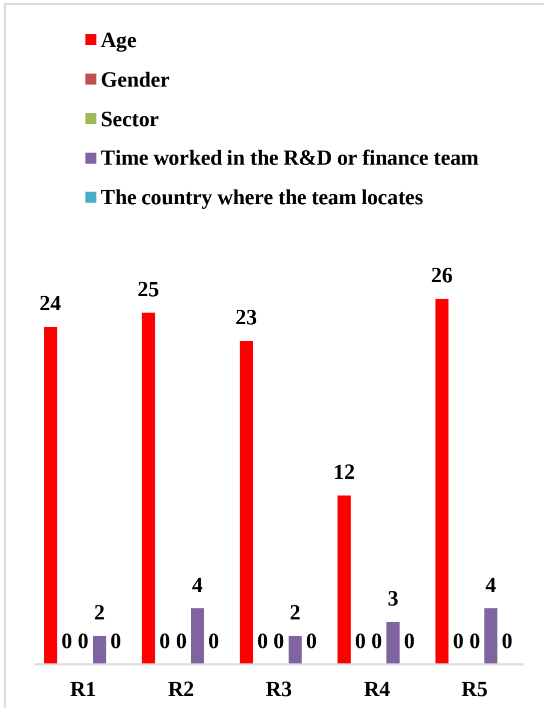


Fig .5 Engaging data of interview members

Table 2 Blockchain can be better utilized, got to and embraced by monetary administrations and different associations.

Year	Number of projects
2013	11
2014	25
2015	10
2017	13
2018	41

2019	20
2020	54

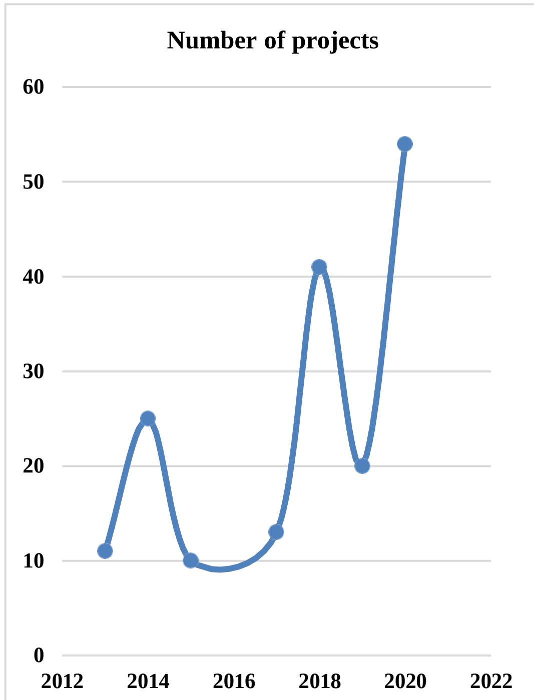


Fig. 6 Blockchain can be better utilized, got to and embraced by monetary administrations and different associations

- **Enough capitals and great monetary administration -** All of the interviewees suggested that businesses wishing to implement blockchain technology should have enough money because doing so is expensive and not all associations can afford the long-term costs.
- **Adjust the association's exercises to Blockchain drives -** The decision to use Blockchain's solutions (either for real estate or testing, or both) should be in line with the core operations of the financial sector. To be sure, the organizations shouldn't have to deal with the issue of entering this new market if they don't engage in financial speculation.
- **Adequate energy and electrical supplies -** Blockchain will need a significant amount of power, and abundant energy sources may make it necessary to have certain resources before adopting or using Blockchain.

- **Dependable high computational power -** In a similar vein, Blockchain will need PCs that are nearly as powerful as supercomputers in order to process hundreds or millions of estimates every second **[10]**.

Table 2 Blockchain for Monetary Consideration and Portable Monetary Administrations

Year	Number of publications
2013	22
2014	42
2015	25
2017	42
2018	33
2019	10
2020	23

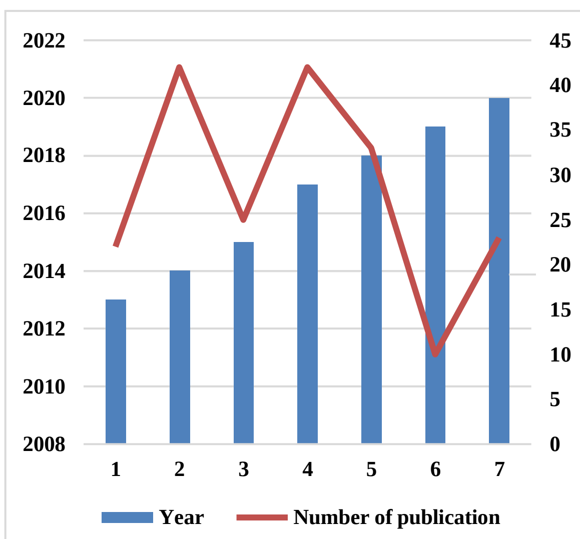


Fig .7 Blockchain for Monetary Consideration and Portable Monetary Administrations

A. *Discussion*

They have been eager to support new services and products, and they anticipate being early adopters of the industry with the usual goal of securing a sizable portion of the market. Blockchain can bring about special improvements for the organization since it can attract more attention and investment opportunities from investors, and financial services are more willing to raise the standard of services and product development. New teams (including marketing, research, and IT) have been formed with more customers and incredible commercial opportunities available **[11]**.

V. CONCLUSION

This article discusses how, thanks to Blockchain, the financial industry is about to enter a new financial era with yet another terrifying framework. The previous products and services that the financial sector offered were thought to be extravagant and inefficient. Consequently, a huge adjustment was needed. mentioned that credit recreation—a cross-time agreement system that enabled people to trust one another without the need for social connections or credit gathering—could be handled by blockchain [12]. Though more work remained to be done to address the hidden difficulties, blockchain innovation has the potential to improve the efficacy and security of the financial business sectors.

However, the respondents provided valuable insights into the current state of Blockchain use. One growing problem was information stowing away. One of the promises of this investigation is that, expanding upon the previous analysis of information hiding, it has discovered a few remarkable justifications for data hiding in the financial industry through the use of Blockchain. acknowledged that storing aside information could make one feel more prominent, aid in receiving praise from others, or lead to better career advancement [13].

Understanding how to handle issues that arise in hazy figuring frameworks, such as centralization and flexibility, as well as idleness and security, is one of the study's goals. From some angles, it is possible to incorporate an ECC digital signature into the hashing process on a blockchain that has received unrestricted public assistance [14]. Our security framework provides an impenetrable arrangement that is unbreakable by human interference because it is based on clinical data. An encoded hash of the document linked to the sharing connection will be provided to everyone who can request the records in the meantime, along with a proof that the original soliciter must view **[15]**.

VI. FUTURE WORKS

The improvement of such a Blockchain reception structure will be the focus of our future work, with a particular emphasis on the mechanical, hierarchical, and individual (TOP) factors that manage profitable Blockchain reception and enable Blockchain to function effectively for a variety of business activities and services.

- We sent an optional exploration plan and, in doing so, split down existing data in a similar domain to help make sense of it all the more cautiously.
- Using direct information would be more enticing than breaking down repeated information, which has predispositions.

VII. REFERENCES

[1]. Revathi, R. Kaladevi, K. Ramana, R. H. Jhaveri, M. R. Kumar, and M. S. P. Kumar, "Early detection of cognitive decline using machine learning algorithm and cognitive ability tesT," Security and Communication Networks, vol. 2022, Article ID 4190023, 13 pages, 2022.

[2]. R. Bhatt, P. Maheshwary, and P. Shukla, "Application of fruit fly optimization algorithm for single-path routing in wireless sensor network for node capture attack," Computing and Network Sustainability, Springer, Singapore, vol. 75, 2019.

[3]. M.S. Ali, M. Vecchio, M. Pincheira, et al. Applications of blockchains in the internet of things: a comprehensive survey IEEE Communications Surveys and Tutorials, 21 (2) (2019), pp. 1676-1717.

[4]. M. Gupta, V. P. Singh, K. K. Gupta, and P. K. Shukla, "An efficient image encryption technique based on two-level security for internet of things," *Multimedia Tools and Applications*, 2022.

[5]. E. M. Onyema, P. K. Shukla, S. Dalal, M. N. Mathur, M. Zakariah, and B. Tiwari, "Enhancement of patient facial recognition through deep learning algorithm: Conv Net," Journal of Healthcare Engineering, vol. 2021, Article ID 5196000, 8 pages, 2021

[6]. S. Stalin, V. Roy, P. K. Shukla et al., "A machine learning-based big EEG data artifact detection and wavelet-based removal: an empirical approach," Mathematical Problems in Engineering, vol. 2021, Article ID 2942808, 11 pages, 2021.

[7]. D. Jain, P. K. Shukla, and S. Varma, "Energy efficient architecture for mitigating the hot-spot problem in wireless sensor networks," Journal of Ambient Intelligence and Humanized Computing, 2022.

[8]. Z. Zheng, S. Xie, H.-N. Dai, X. Chen, H. Wang Blockchain challenges and opportunities: a survey Int. J. Web Grid Serv., 14 (4) (2018), pp. 352-375

[9]. Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. Technical report, Manubot, 2019.

[10]. R. R. Althar, D. Samanta, M. Kaur, A. A. Alnuaim, N. Aljaffan, and M. A. Ullah, "Software systems security vulnerabilities management by exploring the capabilities of language models using NLP," Computational Intelligence and Neuroscience, vol. 2021, Article ID 8522839, 19 pages, 2021.

[11]. S. Pandit, P. K. Shukla, A. Tiwari, P. K. Shukla, M. Maheshwari, and R. Dubey, "Review of video compression techniques based on fractal transform function and swarm intelligence,"International Journal of Modern Physics B, vol. 34, no. 8, 2020.

[12]. X. Wang, A. Zhang, X. Xie, and X. Ye, "Secure-aware and privacy-preserving electronic health record searching in cloud environment," International Journal of Communication Systems,vol. 32, no. 8, article e3925, 2019.

[13]. H. Kaushik, D. Singh, M. Kaur, H. Alshazly, A. Zaguia, and H. Hamam, "Diabetic retinopathy diagnosis from fundus images using stacked generalization of deep models," IEEE Access, vol. 9, pp. 108276–108292, 2021.

[14]. Osho, O., Mohammed, U.L., Nimzing, N.N., Uduimoh, A.A., Misra, S.: Forensic analysis of mobile banking apps. In: International Conference on Computational Science and its Applications, pp. 613–626. Springer, Cham, July 2019.

[15]. Abayomi-Zannu, T.P., Odun-Ayo, I., Tatama, B.F., Misra, S.: Implementing a mobile voting system utilizing blockchain technology and two-factor authentication in Nigeria. In: Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019), pp. 857–872 Springer, Singapore (2020).

Blockchain-Based Dynamic Music Download Payments

Sushmita Reddy

Research Scholar

***Abstract*—** This study examines how blockchain technology is affecting the music industry, with a focus on potential effects on musicians. We argue, based on an analysis of the industry's supply chain, that although on-demand streaming services (like Spotify and Apple Music) have made it easier for customers to obtain music products, they have also created a middleman between musicians and fans, which has led to inefficiencies in the systems that pay royalties. This essay examines the many legal challenges surrounding the application of blockchain technology in the copyright sphere, with a focus on the condition of the online music market at the moment. The copyright information of any collection or song composed by artists is safeguarded by the Ethereum blockchain invention. In light of the Ethereum smart contract, the framework concept and information stockpiling are created and carried out. The framework provides experts with efficient and optimal sovereignty installments and makes use of clever agreements to save music information on the blockchain organization.

Keywords*— *Blockchain, blockchain technology, Music industry.

I. INTRODUCTION

Today's listeners access and experience music in a completely different way thanks to digitalization and the internet. The way musical works are created, produced, distributed, and used online has changed dramatically due to rapid technology advancements. As a result, a strong legal framework has been established to protect copyright holders and artists. Artists have long battled to protect their digital works, as user-upload content platforms frequently result in their being duplicated and reprinted online without their permission. This kind of exploitation is frequently offered for free or at a very low price. With the record industry contributing roughly US$19.1 billion in sales, the music industry is currently estimated to be worth roughly US$45 billion worldwide. 2019 saw global growth for the fourth consecutive year at its fastest rate since 1997.

3. The main factor behind this growth was a 32.9% rise in paid streaming, which now makes up 37.0% of overall revenue. 4 Thus, it is evident how significant the amounts involved are and how crucial new digital methods of monetizing music are in the current online music industry (such as distribution, copying, paid downloads, free or paid streaming, etc.). Therefore, it is necessary to have a strong legal framework that covers all of these new uses and exploitation strategies [1].

Despite the fact that this has been the case, financial research has also been conducted to determine the level of copyright protection that is necessary to provide the creator and distributors of original work with a financial incentive. Online music appropriation (OMD) has a direct impact on the activities of the Recorded Music Industry (RMI). The opportunity for OMD was made possible by innovations in the development, distribution, and storage of digital content, while RMI continued to operate in accordance with a generally accepted plan of action. For a long time, the innovative idea underlying blockchain technology has led to a massive demand for blockchain, mostly to remove outsiders from the financial industry and to reduce the displaying in light of trust and investigation. Blockchain has emerged as one of the most promising innovations since the innovative market's inception.

This is how the remainder of the paper is structured. In Part 2, we provide a well-organized summary of the cycle, blockchain advancement, and its internal responsibilities. The music industry's innovative association is examined in Locale 3, which also discusses how the web has recently impacted the industry and how blockchain might have an impact. We examined the methodology used in this evaluation in Segment 4. The fifth section describes how blockchain might help with some of the issues encountered in the association's creative flow and how it can increase the enormous value that entertainers capture. In addition, we examine the organization's current stage of blockchain development as well as a few operational apps and connections. In the final section of the article, we argue the effectiveness of the current models and how future responsibilities should take into account the display of blockchain associations within the organization.

Given these concepts, the purpose of this investigation is to determine the potential value that BCT could bring to the music industry through its implementation in BC-based advanced music distribution systems (henceforth referred to as BBDMDSs). In this sense, respect suggests the potential to resolve current issues in the music industry while also developing new methods. Furthermore, we should identify and investigate the factors that influence consumers' expectations of using these associations as well as their evaluations of them. As a result, we should also fill in notable research gaps, enhance the body of knowledge in the field by using case studies of BC-based applications as well as customer perspectives, and provide BBDMDSs with suggestions for enhancing their actual capabilities **[2]**.

A. Delimitations

A few cutoff points were imposed in order to restrict the scope of this assessment and to stimulate the development of further exploratory questions. In the section that follows, they are investigated. This analysis focused on BBDMDSs in order to evaluate the potential value that BCT could provide to the music industry. The significance of cutting-edge music dispersion services (DMDSs) in the modern music industry and the scale of their providers led to the selection of these offerings. Therefore, a comprehensive analysis of the value of BCT creation in the music industry is outside the purview of this review. Moreover, the evaluation of these services is limited to the perspective of customers as important partners. As a result, certain music industry partners—such as experts, brands, distributors, and social order gatherings—are avoided from fundamental investigation **[3]**.

B. Technical Highlights Of Blockchain

Due to coincidental turns of events, innovation and regulation have always been closely linked. Swift mechanical adjustments necessitate modifying the legal environment since these creative modifications anticipate being "legitimized" in some way to become standards in our daily lives. These days, blockchain innovation is seeing a similar shift that Wear and Alex Prescott have dubbed "the second time of the web". 6 In 2008, someone discussed blockchain for the first time. In his well-known White Paper7, the anonymous genius who created the bitcoin digital currency, Satoshi Nakamoto, explored the intriguing possibility of creating a different, specialized system for cash disbursements. Because all transactions were saved on separate PCs and verified in a reliable and secure method by all parties involved in the chain, this strategy took into account a safer, more identifiable, and simpler installation. Contrary to popular belief, conveyed records innovation—and its more well-known version, such as blockchain technology8—is not a particularly noteworthy innovation. Rather, it belongs to a class of advancements that generally alter their administrative and specialized setups [4].

To put it simply, blockchain innovation functions as a common, synchronized regular resource library that allows users to track and save information in a simple, safe, and adaptable manner. The name of the invention comes from the way multiple exchanges are requested simultaneously in alleged blocks, which are actually called "records." The characteristic that blockchain innovation is contingent upon consensus among a group of PCs (referred to as "hubs"), participating in a common presumption to verify each exchange's legitimacy before it is stored on the blockchain, is of primary importance. Actually, before creating the block, each hub is supposed to compete with the others to solve a cryptographic puzzle that verifies the trade. The hub organization receives the "evidence of work" once a solution has been identified.

C. The Benefits Guaranteed By Craftsmen For Blockchain Innovation

Many experts in the music industry have recently attested that the system of the industry generally needs to be changed due to the many challenges brought about by the emergence of the internet and the industry's inflexible nature, making it unable to adapt to the barrage of new technologies.12 In an explanation, Senior Speaker O'Dair reaffirmed: The complexity of conflicting copyrights over comparable musical works is the root cause of the need to mend the music industry.14 For instance, under the Berne Show, the owner of the copyright enjoys a series of progressive financial and moral liberties as soon as the work is created, without having to register or be seen by any authority. In particular, the right holder may be granted 177 different public copyrights and associated benefits, which are compiled into a global "fracture" of public copyrights.

II. RELATED WORKS

(Guo, J., Zhou, H., Yang 2020) One of the most significant innovations and creative leaps forwards that is assuming a crucial role in the professional present is blockchain innovation. The advancement of blockchain innovation tends toward revolution and steady transformation. A series of blocks encloses information and maintains consistency between individuals while paying little attention to their separation. The growth of blockchain technology in recent years has forced scholars and professionals to research better ways to apply blockchain innovation in several contexts. Local blockchain innovation development has opened up several new application avenues, including medical applications **[5]**.

(Han, P., Sui, A., 2020) The development of Industry 4.0 in Indonesia is strongly supported by the Service of Industry. From the moment of its inception, Industry 4.0 innovation has advanced very quickly. Almost every aspect of life has embraced cutting-edge innovation. The contemporary insurgency modifies a nation's educational and financial requirements directly or indirectly. The seventh (seventh) leader of the Republic of Indonesia launched the "Making Indonesia 4.0" initiative as part of the administration's commitment to usher in the era of innovation 4.0. There is a lot of conjecture in Indonesian reality around the need to advance and prepare for this disruptive wave of 4.0, including implementing various enhancements and learning techniques **[6]**.

(Peng, W., Yi, L., 2019) Because computerized copyright insurance has such a big impact on society's progress, more thought should be given to it. Although a lot of automated copyright assurance techniques have been developed in the past, there are still a lot of security framework escape clauses that need to be addressed. The results are not tremendous, the security concerns remain quite helpless, pragmatism is lacking, intrusion occurs frequently, and a proper assurance moves slowly. With an emphasis on the aforementioned problems, this research suggests a security technique that makes use of texture's clever agreement innovation to comprehend the programmed management of the entire life cycle of computerized privileges on the blockchain. The suggested structure is reliant on blockchain innovation, which impacts the communicated, well-crafted, and observable characteristics of blockchain **[7]**.

(Zeng, Y. (2020) This will have a negative impact on experts' motivation and vitality. This paper provides artists with an online data set platform that stores music compositions and protects the copyright information of each collection or piece of music produced by industry experts. It does this by leveraging the distributed and thoughtfully crafted innovation of the Ethereum blockchain. Information stockpiling and plan and execution of the framework model are suggested, and information capacity processes in light of the Ethereum astute agreement are fully demonstrated. The system uses the blockchain network to store music data and makes use of this clever arrangement to give artisans a quick and efficient way to get recognition. We use Node.js to do the tests of our framework. We also use Block Adventurer to track music data on the blockchain and test Remote Procedure Calls (RPC) with accessible records and private keys for contract progression. Our architecture makes it possible for music designers to receive credit for copyright earnings, which will help stop the illegal distribution of music on various websites **[8]**.

III. METHODS

A. Research Plan

This section introduces and discusses the methods utilized to address the now resolved fundamental examination topic and its auxiliary hypotheses. Research examining practical applications of blockchain technology in the music industry is lacking, and none of these studies takes into account the opinions of buyers at the time of writing [9]. The evaluation process was primarily based on an experiential approach, which placed a strong emphasis on firsthand experiences and evaluated uniqueness from a different angle (Robson, 2002). In order to facilitate our coverage of the selected investigation opening and data dissemination in the field, this framework is designed to make use of a local appraisal process.

Perceiving clients' social mentalities regarding the use of BBDMDSs and the factors that determine the purpose is one of the proposition's goals. This section will look at the speculative system that is used to analyze the client's viewpoint. It will begin by discussing the rationale behind the selection and application of conjectures in the proceeding setting, before introducing the TPB and Cap and their application in previous evaluations. Finally, the extra model that completes the tried-and-true conjectures is promoted.

Additionally, they believe that a person's behavior is still up in the air because of their expectation that they will behave a certain way. The desire to act out a particular behavior is seen as the goal in relation to the TBP.

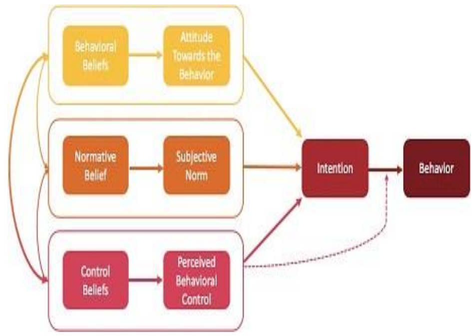


Fig.1 Plan of the Hypothesis of Arranged Conduct

A. Vision sense

The Behavior Disposition (A), defined as the "person's positive or pessimistic assessment of playing out the way of behaving," is the primary determinant of objective (Ajzen, 1985, p. 12). Therefore, not fixed by beliefs that link behavior to a certain desired outcome. We refer to those beliefs as social convictions. In the current context, a model may be a person's belief that using a blockchain application for music streaming (conduct), leads to a more transparent framework, pays craftsmen directly, and democratizes the industry (result) **[10]**.

B. Abstract Norm

"A singular's impression of the overall weights put on him to perform or not play out the direct being alluded to" (p. 12) is the close to home norm (SN), which is the resulting determinant. Furthermore, much like with behaviors, the SN is dependent on beliefs. Ajzen (1985) looked into normalizing convictions for this situation, which are sentiments about how other people or social situations judge whether or not a person should conduct in a certain way (in complete agreement). This actually anticipates that if a person admits that the vast majority of those in his immediate vicinity, with whom he wishes to obtain consent, believe that he or she should or shouldn't act in a certain way, the person will probably feel pressure to act in a certain way or not (common weight) **[11]**.

C. *Social Insights*

In charge An individual's perception of how easy or difficult it is to act in a particular way served as the basis for one of the main forms of social objective setting; this concept is now known as perceived behavior control (PBC). Beginning with the introduction of most behavioral patterns, which rely in part on non-motivational and non-social factors like easy availability to essential supplies and entry points, this section is susceptible to these conditions.

IV. RESULTS AND DISCUSSION

The supplementary section will provide and analyze the observational findings derived from the collection of crucial data in conjunction with the freshly presented theoretical system of the TPB and the Hat. Interviewees' declarations were distributed using the categories created by the factors in the conjectures. In addition, each class was divided based on emerging characteristics, such as impact and assessment (for SN) **[12]**.

Table 1 Socioeconomics of Interview Members

Porband	Age	Sex	Origin	Occupation
Bastian	25	Male	Geramany	Student
Peter temur	26	Male	Geramany	AI
Michal	24	Male	Urbekistaan	Student – e business
Karolis	26	Male	Creez republic	Student marketing

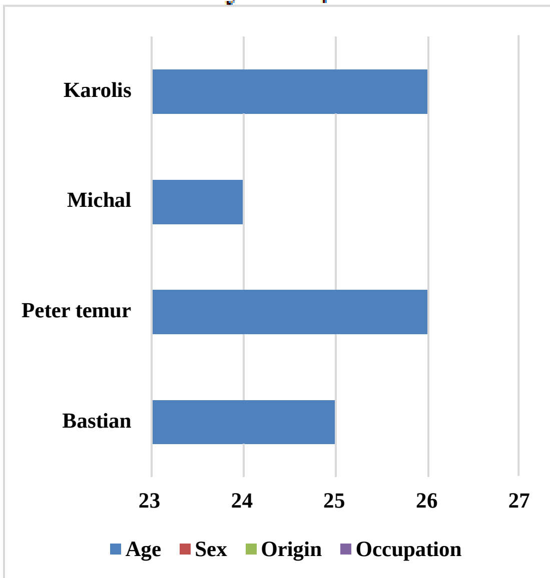


Fig.1 Socioeconomics of Interview Members

Table 2 Incomes from real time features

Streaming platform	Unsinged artist	Signed artist
Musicoin	125	114
Apple music	469	561
Google play music	134	475
Deezer	115	461
Spotify	889	461

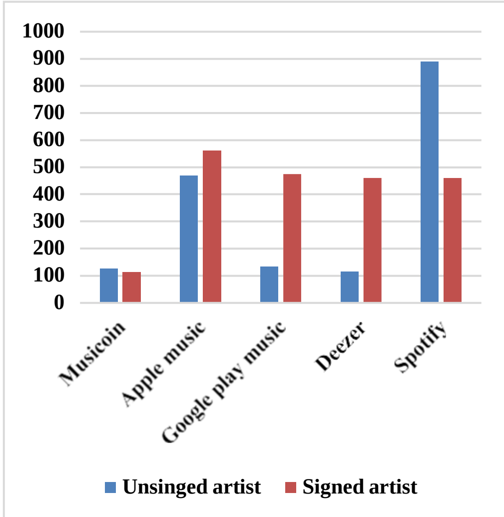


Fig.2 Incomes from real time features

Table 3 Record of shopper interview

Name	Bastian	Origin
Age	25	City
Occupation	Student- mechanical side jobs	Self- employed artificial intelligence

A. Blockchain Use Cases and Made Worth

BBDMDSs hope to overcome a few obstacles in the music industry, which could eventually outperform more established forms of automated music movement. In this section, we'll examine the potential advantages of blockchain development for the music industry in comparison to the current convoluted music allocation system **13]**.

A. Facilitating Processes

By all accounts, the BCT application in computerized music circulation administrations is suitable for expediting, obtaining, and streamlining present jobs. By enabling the organization to save data across PCs, BCT implementation makes the framework decentralized and appropriated. All things considered, the information and trade records that are dispersed throughout the entire organization make the creators' works far more secure than they were previously **[14]**.

V. CONCLUSION

This study suggests an IPFS-based blockchain-based music-sharing platform. Our proposal uses blockchain technology, which also functions with the web's checks for duplicate metadata, to eliminate and limit content creators' illegal music sharing. This report sharing association was established with the Ethereum blockchain in mind, which employs a more effective understanding technique than achieving the stated goals of a sharp agreement in a quick and efficient manner [15]. The different developments anticipated to envision the movement of the suggested structure are presented by our generation. In order to ensure the protection of music documents, our reconstruction included the enrollment and access control feature that was added to the IPFS convention. The astute arrangement ensures that the requirements that have to be fulfilled before beginning the music recording are carried out. The organization's closed state was demonstrated with the intention of ensuring that each interest hub ensured that records were maintained on downloaded music documents in relation to the average revenue. Responsible income management is achieved with our suggested approach.

Collected findings indicate that consumers are quite picky about the content that is available on their preferred music delivery service.

VI. FUTURE WORK

Obtaining sufficient information about the current stages' presentation proved to be difficult due to the increasing phases that underlie them. Ideally, comparing those blockchain-powered stages' aftermath to the underlying effects of their competitors using the membership model would show how disconnected they are from mainstream use. Similarly, not much research was done on blockchain in the music industry.

Not only is writing confined, but the innovation's suitability for this particular business is also limited. With this analysis, we hope to evaluate blockchain's advancement in the music industry further by analysing it from the perspectives of customers and experts regarding its presentation and reception.

VII. REFERENCES

[1]. Aki, J. (2018). Report Links 74% of Bitcoin Mining to China, Sees Threat to Network. Retrieved March 18, 2019,

[2]. Akolkar, B. (2018). China Officially Bans All Crypto-Related Commercial Activities. Retrieved March 18, 2019, from

[3]. Bartmann, J. (2019). List of blockchain music streaming platforms to watch in 2019. Retrieved March 21, 2019, from

[4]. Bazinet, J., Singlehurst, T., May, M., Suva, J., Ezawa, K., Yap, A. (2018). Putting the brand back together. Remastering the World of Music. Citi GSP. Retrieved March 13, 2019, from

[5]. Guo, J., Zhou, H., Yang, L., & Chen, X. (2020). Research on digital copyright blockchain technology. Proceedings - 2020 3rd International Conference on Smart BlockChain, SmartBlock 2020, 114–118.

[6]. Han, P., Sui, A., Jiang, T., & Gu, C. (2020). Copyright Certificate Storage and Trading System Based on Blockchain. Proceedings of 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications, AEECA 2020, 611–615.

[7]. Peng, W., Yi, L., Fang, L., Xinhua, D., & Ping, C. (2019). Secure and Traceable Copyright Management System Based on Blockchain. 2019 IEEE 5th International Conference on Computer and Communications, ICCC 2019, 1243–1247.

[8]. Zeng, Y. (2020). Digital Music Resource Copyright Management Mechanism Based on Blockchain. Proceedings - 2020 3rd International Conference on Smart BlockChain, SmartBlock 2020, 158–162.

[9]. Beaumont-Thomas, B., & Rushe, D. (2018). Spotify sued for $1.6bn in unpaid royalties as it reportedly files for IPO. Retrieved March 15, 2019,

[10]. Capgemini Consulting. (2016). Smart Contracts in Financial Services: Getting from Hypeto Reality.

[11]. Chen, Y. (2018). Blockchain tokens and the potential democratization of entrepreneurship and innovation. Business Horizons, 61(4), 567-575.

[12]. Choon. (2018). Whitepaper. Retrieved April 06, 2019, from.

[13]. Choon. (2019). About. Retrieved April 06, 2019.

[14]. Daniel, R. (2019). Digital disruption in the music industry: The case of the compact disc. Creative Industries Journal, 1-8.

[15]. Naumoff, A. (2017). Why Blockchain Needs 'Proof of Authority' Instead of 'Proof of Stake.' Retrieved May 11, 2019,

Blockchain Smart contracts for Smart Supply Chains

Jitendra Yadav

Research Scholar

Abstract— An electronic exchange convention called a beautiful agreement is used to meticulously handle, verify, and preserve the specifics of a legal agreement. Because of this, astute agreements aim to reduce trade expenses such as requirement and discretion charges by enabling identifiable and irreversible exchanges via the use of blockchain technology for distributed data sets. On the other hand, clever agreements may have a much greater impact than just setting aside money for reserves because they foster the kind of cross-hierarchical business process collaboration that smart supply chains are known for. The purpose of this post is to provide users with advice about the work being done in this area in relation to blockchain innovation. Based on the review's findings, companies should invest in blockchain innovation to improve the transparency, flexibility, and security of their supply chains. Unquestionably, blockchain innovation greatly contributes to the construction of trust between store network participants. The discussion of the benefits and potential effects of blockchain innovation in coordination and participation brings the review to a close.

Keywords— Blockchain, Industry 4.0, Smart Contracts, Smart Supply Chains.

I. INTRODUCTION

In today's global marketplace, supply chain management, or SCM, is essential. It has an effect on the global economy. SCM is sometimes described as the evolution of goods from manufacturer to consumer. It comprises the manufacturer, merchant, and retailer and is divided into multiple phases, starting with the inventory of raw materials and concluding with the customer. It's a global cycle in which components are taken from a single location, packaged, and distributed globally. While conventional store network executives accomplish a broad goal, they fall short in terms of complete consistency. There are a few cutoff lines for allowing the final customer to change the exchange and ensuring the quality of the goods supplied. Generally speaking, it is comparable to advance streams, or the flow of goods from the source to the recipient [1].

Blockchain is a P2P distributed record technology. Blockchain is therefore distinguished by decentralization, incredibly robust material, and noticeable data. A typical example of blockchain 2.0 is splendid understanding, which is essentially a piece of code that can function automatically and self-check. The brilliant plan expands the use cases for the blockchain, and the blockchain ensures the brilliant understanding's code security. Finally, a description of blockchain-smart arrangement advancement has been provided. In this analysis, we provided strong points for a blockchain-based plan for the rice store network as well as excellent actions that may fulfil the specific management of the rice store association. The security and brightness of the entire retail organization are increased by the highly automated and skilful planning of these creative courses of action. Our model design may provide associations involved in rice production with internal association and information availability capabilities. It may provide controllers with an efficient and effective way to gradually manage the entire store association, encourage the real controllers, and provide buyers with services like rice discovery.

A. Contribution

We first oversaw a broad analysis of the network of rice stores and established a potent management framework in light of blockchain technology and astute agreements. We gave the model the ability to automatically collect data and manage the rice production network by modifying clever contracts. A model framework for the special administration of the network of rice stores was established based on that idea. This investigation can often be divided into three parts:

- To provide a legitimate environment for the management of the rice inventory network, we employed blockchain as the innovation stage. The entire rice life cycle is covered under the scope of management. This allows information to be shared and interconnected among all links within the network of rice stores and its many divisions.
- We completed the unique ongoing oversight of store network information, the administration of bosses, and the programmed assortment of information for the entire rice production network by reconstructing and assembling dynamic dazzling agreements.
- The suggested approach may deftly extract and coordinate the knowledge obtained during repeated project cycles, providing a reliable information source for the ensuing oversight.

B. Organization

Here is how the rest of the paper is organized. The fictitious foundation is discussed in the section that follows [2]. While "Savvy contract configuration" displays the agreement's pseudo-code execution, "Administrative model development" briefly presents the model. "Results and examination" presents the findings, uses a model to investigate the model's reasonableness, and clarifies the model's common sense. "End and future work" concludes.

B. Blockchain And Shrewd Agreements

As used currently, the term "blockchain" refers to two distinct components: a shared information index and a data structure (e.g., a list of linked blocks with trades, each of which is cryptographically secured to the previous block by incorporating its hash value and a cryptographic signature, making it difficult to alter a previous block without recreating the entire chain since that block). The concept of "smart contracts," which are scripts that run whenever a particular kind of transaction takes place and have the ability to read and write from the blockchain, is linked to blockchain innovation [3]. Excellent plans enable groups to make sure that further transactions take place after a particular exchange.

Think about the scenario of transportation within a stock organization, where all information is distributed via the creation organization's informative indexes on providers, recipients, items, and business conditions. Selling work and goods as part of an exchange is feasible, provided that there is a clever agreement for conditional exchanges and that both the vendor and the buyer cryptographically approve of the transaction. Excellent agreements can enable the proper implementation of beneficial, innovative cycles by authorizing the exchange of associated resources and liberties when the transaction takes place and any other agreed-upon requirements, such as charge portions, paperwork, or quality checks, are satisfied. Petri-Nets that describe the center's work cycles and tokens that manage the majority of approvals (also known as "authorization tokens") or information are widely used to mathematically express the main business tasks and associated legal actions **[4]**.

C. Case: Conveyance Robots

examined supplier chains and organized cross-organization value in the context of a measured assembly organization. The contextual analysis illustrates every aspect of structured business structures in relation to fractals, self-association, self-enhancement, and distributed frameworks working together as a planned, measured project; that is, the structured venture idea is based on current models for carrying out pre-Industry 4.0 designs. As a result, the use of clever contracts and blockchain innovation is generally classified as belonging to the category of measurable organizations. The concept of Industry 4.0, in general, embraces communications between CPS that coordinate online across organizational boundaries inside astute assembly and planned operations organizations. Business procedures demand arrangements in a conveyed IT environment, and the concealed cross-organization esteem creation activities involve automatized items in various fractals along with M2M correspondence and the associated work processes. Here, clever agreements and blockchain innovation provide a fair arrangement that provides a recognizable and simple approach that may be used for the lifetime of an Industry 4.0 item.

D. Regulatory model development

- Examination of management data in the rice production network

Researchers are generally concerned on the production network examination of different yields in the food and agricultural industries. Previous studies designed a second closed-loop production network for the pecan industry, incorporating homestead advancement, acquisition targets, partition targets, and so forth. Focuses on recycling and assortment, as well as related industries, go in opposite directions. provides a closed-loop supply chain network for the sugarcane sector, encompassing producers, processing facilities (or sugarcane industry), distribution centers, merchants, other businesses, composting facilities, and manure markets **[5]**.

II. RELATED WORKS

One of the most important ways to ensure sterilization is to control rice quality and welfare better. In order to support equitable rice management, standard administration methods have been modified, primarily from the perspectives of morality, guidelines, development, and declaration. Morality-wise, the expertise of creating and managing the workforce has been improved to reduce the likelihood of problems inside the rice store chain. Upgrading rice inspecting criteria and guidelines, strengthening the board, and organizing legal training have all been part of the legitimate viewpoint. Regarding, it has been decided that the monitoring of heavy metals should be strengthened through the development of a biogeochemical cycle model, focusing on the sources of heavy metal pollution in rice fields, and identifying the movement of heavy metals between different rice paddies. In terms of rice checking, it might be strengthened by emphasizing the dangerous effects of contaminated rice and poorly thought out development and maintenance techniques.

(Yanjun, R. E 2020) Rice is the primary cereal grain that provides the necessary amount of cadmium in a human diet. In this review, we examine how dietary differences in cadmium admission are exacerbated by geological, genetic, and handling (taking care of) factors that affect rice grain cadmium and rice utilization rates. Based on a review of 12 countries across four landmasses, Bangladesh and Sri Lanka had the highest levels of cadmium in rice grain, and these two countries also had the highest per capita rice consumption **[5]**.

(Wei, L. et al 2020) The use of maker rice may be the cause of the greater health risk to adults than to children. The primary sources were identified as regular sources (heavily stacked with Cr and Ni, such as soil parent materials and lactogenic parts), rural exercises (heavily stacked with Cd, Cu, and Zn), especially over the top use of pesticides and manures, and modern exercises (heavily stacked with Hg, Pb, and As), such as vehicle outflows, coal ignition, and those of the material and synthetic ventures. Effective policies should be maintained in order to guarantee the safety of ranch produce and protect the environment and public health in the area under consideration **[6]**.

(Ouyang, L., Yuan 2020) Recently, nitric oxide (NO), a chemical particle associated with redox, has been considered an essential regulator of plant growth and development as well as a response to abiotic stressors. One of the most dangerous factors influencing crop development and formation is heavy metal (HM) stress. Plant oxidative pressure is triggered by HM stress, which is linked to the production of reactive oxygen species (ROS) **[7]**.

III. METHODS

With the current shortcomings in rice supply chain control, we developed a potent administration strategy based on blockchain technology and clever agreements. Organizations, purchasers, and leaders working with the rice creation organization are supported by this model. There are two types of observational data in this successful administration perspective. The most basic type includes information about heavy metals, my zootoxins, pesticide stores, bugs, fumigants, herbicides, and other harmful, unsafe pollutants in the store organization; seed sources, object gatherings, packaging materials, accruing time, and transportation data; and trade records for seeds, compost, purchases, labor costs, and the dressing and neatness requirements of each component in the institution. The second type of data involves closely observing the behavior of participating firms, customers, and authorities. This layout might provide the rice store organization with a specific administrative process and a trustworthy information hub **[8]**.

Fig.1 Order table of key data of the entire rice production network.

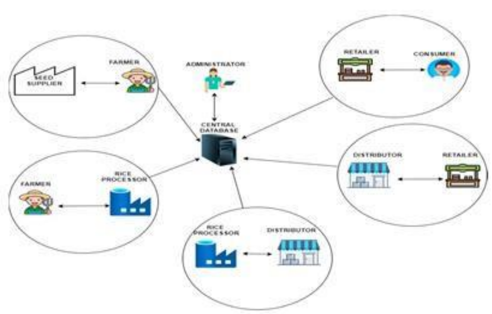


Fig.2 Development of rice production network oversight model driven by blockchain shrewd agreement

A. Model

The rice stock organization's unique supervisory methodology is depicted in Figure 1. It is separated into three modules: storing, management, and presentation. Store network users are divided into three kinds by the establishment module: customers, authoritative associations, and attempts. Six categories are used to categorize attempts: creation, moving and caring for, caring for, limit, transportation, and arrangements **[9]**.

Table 1 The second kind of information is connected with powerfully checking the way of behaving.

Variety	12
Sun hours	42
Tadiff	46
Avn	10
Tmean	49
Elevation	75
Aspect	12
TPI	26

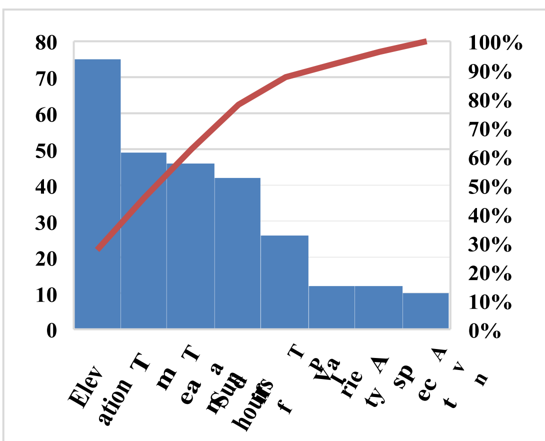


Fig.3 The second kind of information is connected with powerfully checking the way of behaving

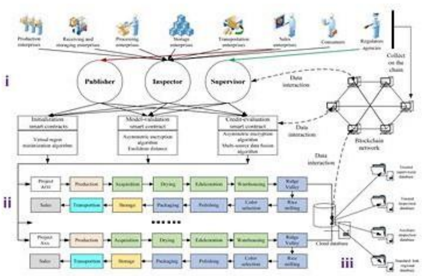


Fig .4 Dynamic oversight model for the rice inventory network

IV. RESULTS AND DISCUSSION

A. Examination of model activity process

In light of blockchain technology and brilliant agreements, the activity rationale of the rice inventory network's powerful oversight model can be divided into three phases that correspond to the three shrewd agreements: the introduction stage is analogous to the savvy contract [10], the model-confirmation stage is analogous to the savvy agreement on the model-check, and the credit assessment stage is analogous to the savvy contract on the credit assessment. The project is divided into three stages with a single focus, and the model uses many threads to address the special management of multi-business rice. Figure 6 displays the grouping graph in operation.

Table 2 Lament esteem union bend.

Regtet value	Number of iterations
0.00	10
0.10	25
0.002	31
0.005	46
0.016	23
0.02	10

0.13	28

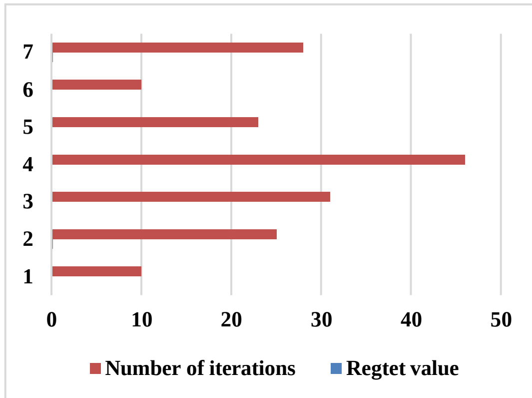


Fig. 5 Lament esteem union bend

Table 3 equal planning relationship with the blockchain network.

Trusted supervisors	Trusted inspectors	Untrusted supervisors	Untrusted inspectors
A0005	A0023	A0023	A0023
A0002	A0100	A1340	A0100
A0025	A1005	A0005	A1005
A0126	A1005	A0002	A1005
A0123	A0023	A0025	A0023
A1002	A0100	A0126	A1230

This model and the blockchain network have an equal planning relationship. Upon reaching hub agreement, the model records each participant on the chain. Based on the information obtained during recruitment, the model is divided into three groups: interest organizations, administrative offices, and shoppers. After that, the model interaction is divided into three stages in order to achieve the special continuous monitoring of the rice plant's whole life cycle. The specific activity process is shown below **[11]**.

B. Model framework confirmation

- **Framework engineering plan**

This investigation promoted a model system taking into account the outstanding oversight model of the rice creation organization. Figure 7 shows the outline of the structure. The application layer, the magnificent arrangement layer, the organization understanding layer, and the data resource layer make up the four layers of the design **[12]**.

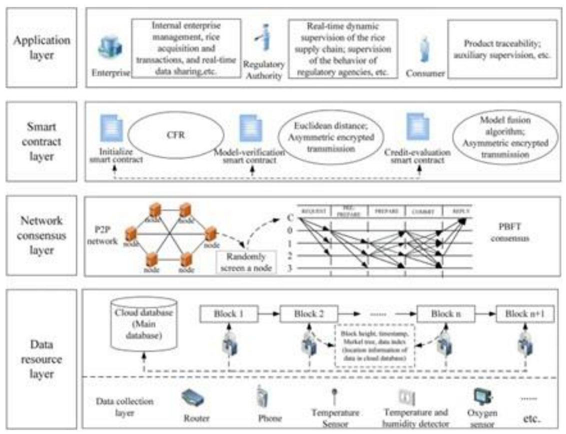


Fig.6 Framework engineering plan

Applications, purchasing groups, and administrative agencies are all served by the application layer throughout the inventory network. This layer provides initiatives such as internal endeavor boards, rice exchanges and purchase, and ongoing information sharing for chain ventures. The model-check shrewd agreement, the credit-assessment shrewd agreement, and the introduction savvy contract together form the brilliant agreement layer and serve to mutually characterize the model's operating theory. Through the design of relevant trigger events, the agreement subsequently implements the existing capabilities **[13]**.

V. CONCLUSION

This study created a model to work with the strong oversight of the existence pattern of rice supply chains in light of blockchain technology and astute agreements. Initially, we deconstructed the overall path of the rice store network and severed its core links [14]. Next, we identified the participating jobs across the entire production network, created an approved information assortment table based on each participating job, and created a key information grouping table for the inventory network based on its major connects. Next, we created a special management model just for the network of rice stores. We managed the boss's behavior and the information in the store network by using astute contracts. To achieve model usefulness, we applied the multisource information combination innovation, Euclidean distance, lopsided encryption calculation, and virtual lament minimization calculation. Finally, we used a model to deconstruct the model framework's accuracy **[15]**.

VI. FUTURE WORK

This investigation of model action speed and limit time is hindered by the limitations of resource stockpiling and blockchain management. In order to solve such issues, we will later on make use of cross-chain development and an organization with a higher setup. Subsequent analysis could also aim to improve the CFR and encryption computations in order to reduce blockchain computing resources. The rice production network places a greater emphasis on risk than other supply chains for grains and oilseeds. It is mostly shared through a variety of organizations, dangerous material information, other important data, and participating associations.

VII. REFERENCES

[1]. Giraldo, F. D., Milton, B. C. & Gamboa, C. E. Electronic voting using blockchain and smart contracts: Proof of concept. IEEE Latin Am. Trans. 18(10), 1743–1751. (2020).

[2]. Peng, X. et al. A refned supervision model of rice supply chain based on multi-blockchain. Foods 11(18), 2785. (2022).

[3]. Casino, F., Dasaklis, T. K. & Patsakis, C. A systematic literature review of blockchain-based applications: Current status, classifcation and open issues. Telemat. Inform. 36, 55–81 (2019).

[4]. Perboli, G., Musso, S. & Rosano, M. Blockchain in logistics and supply chain: A lean approach for designing real-world use cases. IEEE Access 6, 62018–62028. (2018)

[5]. Yanjun, R. E. N. et al. Contamination assessment, health risk evaluation, and source identifcation of heavy metals in the soil-rice system of typical agricultural regions on the southeast coast of China. Environ. Sci. Pollut. Res. 4, 1–11. (2020).

[6]. Wei, L. et al. Recent progress in the knowledge on the alleviating efect of nitric oxide on heavy metal stress in plants. Plant Physiol. Biochem. 147, 161–171. 2019.12.021 (2020).

[7]. Ouyang, L., Yuan, Y. & Wang, F. Learning markets: An AI collaboration framework based on blockchain and smart contracts. IEEE Internet Tings J. 15(12), 6367–6378. (2020).

[8]. Zhou, Z., Wang, B., Dong, M. & Ota, K. Secure and efcient vehicle-to-grid energy trading in cyber physical systems: Integration of blockchain and edge computing. IEEE Trans. Syst. Man Cybern. Syst. 50(1), 43–57. 23 (2020).

[9]. Saad, M. et al. e-PoS: Making proof-of-stake decentralized and fair. IEEE Trans. Parallel Distrib. Syst. 32(8), 1961–1973. (2021).

[10]. Shayan, M. et al. Biscotti: A blockchain system for private and secure federated learning. IEEE Trans. Parallel Distrib. Syst. 32(7), 1513–1525. (2021).

[11]. Li, W. et al. A scalable multi-layer PBFT consensus for blockchain. IEEE Trans. Parallel Distrib. Syst. 32(5), 1146–1160. (2021).

[12]. Peng, X. et al. Multi-chain collaboration-based information management and control for the rice supply chain. Agriculture 12(5), 689. (2022).

[13]. Peng, S. et al. An efcient double-layer blockchain method for vaccine production supervision. IEEE Trans. Nanobiosci. 19(3), 579–587.(2020).

[14]. Peng, X. et al. Research on the cross-chain model of rice supply chain supervision based on parallel blockchain and smart contracts. Foods 11(9), 1269. (2022).

[15]. Saini, A. et al. A smart contract based access control framework for cloud smart healthcare system. IEEE Internet Tings J. 8(7), 5914–5925. (2021).

Advance Analyses for Blockchain Legal Issues and Its Various Output

Rashmi Soni
Research Scholar

***Abstract*—** Blockchain is a distributed point-to-point technology utilized in big data, cloud computing, edge computing, finance, and the Internet of Things. Conversely, artificial intelligence is accelerating many firms' intelligent growth. It makes logical to combine blockchain technology with artificial intelligence, as both are emerging fields. Blockchain makes artificial intelligence more independent and credible, and it could even make blockchain smarter in the long run. In this study, we examine the relationship between blockchain technology and artificial intelligence from a wider, three-dimensional perspective. In addition, we draw attention to shortcomings found in the pertinent literature, including the constraints placed on blockchain technology and the ways in which these constraints manifest themselves in various sectors and establishments. We suggest numerous research gaps and potential exploratory directions based on these findings, which will be highly interesting to scholars and practitioners alike.

Keywords*— *Artificial Intelligence Technologies, Blockchain, Blockchain Technology, Blockchain Legal Issues.

I. INTRODUCTION

Often, an exile group oversees and promotes monetary transactions amongst individuals or groups. To complete an electronic transaction or money exchange, a bank or Mastercard backer should act as a mediator. Additionally, a bank or the Mastercard affiliation incurs costs as a result of such exchange. For a few domains, such as gaming, music, programming, etc., a connected cycle essentially works the same way. Not the two head components included in the exchange, but an untouchable association oversees and coordinates all information and data. It has been attempted to address this issue through the advancement of blockchain innovation. Providing a decentralized biological system where everyone is accountable for information exchanges is the goal of Blockchain development [1].

Distributed ledger technology, or blockchain, maintains an ever-expanding list of information records that have been authorized by the hubs of the organization. Subtle details of every completed exchange are included in a publically accessible report that comprises the information. As a decentralized system, blockchain doesn't require external affiliation. Every focus is on trading the data associated with all Blockchain-based exchanges [2]. Compared to aggregated deals that include an untouchable, this trademark simplifies the system. Furthermore, a few hubs can confirm deals more safely because Blockchain hubs are anonymous. The key application that brought Blockchain innovation to the masses was Bit Coin. Bit Coin established a decentralized platform for digital currency, enabling anybody to exchange goods using virtual currency.

Nevertheless, while Blockchain is generally seen as an appropriate option for overseeing transactions utilizing digital currency standards, it has a number of drawbacks and restrictions that need to be considered and addressed. The great security and steadfastness of the exchanges and focuses on Blockchain should thwart threats and attempts to modify exchanges. In essence, calculating assets is requested in order to verify exchanges on the Blockchain.

It's critical to understand which Blockchain-related topics have already been discussed and worked on, as well as which issues and needs still require evaluation. In order to ascertain these requirements, we made the conscious decision to employ a cycle of arrangement to identify pertinent articles related to blockchain. We used a specially designed research show in the systematic arrangement investigation to look for content in reliable informational indexes. The use of blockchain technology in energy research will assist a range of professionals and researchers in identifying potential assessment areas and requests for additional study [3].

Blockchain and artificial intelligence both have advantages, but they also clearly have disadvantages. While computerized speculation has interpretability and sufficiency difficulties, blockchain has problems with energy utilization, flexibility, security, protection, and proficiency. Combining them would allow them to benefit from standard coordination while standing out as two distinct request heads. These two approaches could be beneficial to each other as they have common requests for information security, inquiry, and trust. The blockchain, for example, can handle the development of calculations, enlisting power, and information assets because of its inherent qualities like decentralization, consistency, and secrecy. Human reasoning, on the other hand, depends on three main components: assessments, handling power, and information.

Alternatively, this paper examines the viability of combining blockchain technology with human intelligence from a three-pronged standpoint, and it comprehensively compiles and illustrates the sites of intersection between the two in numerous areas of investigation [4]. Here are the key tasks for this study that follow:

(1) We investigate how blockchain and artificial intelligence are related, as well as the possibility that they will diverge.

(2) Based on the stream nearby and a recent evaluation on the combination of blockchain and artificial intelligence, we create a detailed graph from various orders.

(3) To explore the possibilities of using blockchain technology and artificial intelligence, we select a few application scenarios and practical use cases from various fields.

(4) Here, we highlight the concerns and challenges associated with using blockchain technology in conjunction with artificial intelligence, and we anticipate that the evaluation process will continue.

A. Background

To realize the benefits of this article's main promises, the coordinating sector must grow further and blockchain technology must evolve generally. This section addresses the hashing and mining parameters as well as the blockchain network's notion of commitment (QoE).

A. Blockchain Development: The High-level Thought

Information records that are compiled to be records are part of the progress of blockchain. It makes use of a sent device for record validation. To maintain the dynamic of the connection, innumerable individuals from all over the world regard blocks as essential. This section will walk through the complete cycle of reasoning that is propelling the development of blockchain technology.

B. Agreement System

A convention known as an agreement system is put in place to ensure that all participants in the blockchain network are abiding by the agreed-upon guidelines. By requiring each member to consent to the terms of the communicated record, it ensures that the exchanges originate from a legitimate source. Since the public blockchain is a decentralized innovation, the anticipated demonstration is not managed by a single, centralized authority. As a result, in order for the organization to examine and authorize any transactions that take place on the blockchain network, the people' consents are required. Because the entire cycle is carried out with the consent of the association's members, the blockchain is a reliable, safe, and trustless advancement for electronic transactions. specific understanding components adhere to specific measurements, enabling association members to accept such requirements. A few comprehension frameworks have been suggested in view of the requirements for safe automated transactions. In any event, the three agreement conventions used by the main digital currencies are appointed evidence of stake (DPoS), verification of stake (PoS), and confirmation of work (PoW).

C. Hashing and Computerized Marks

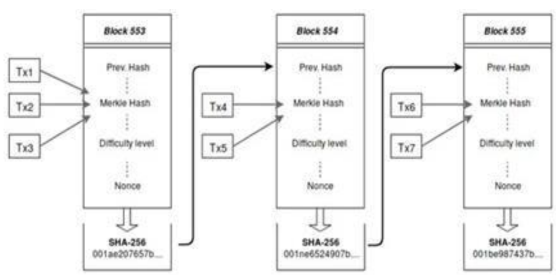


Fig.1 Hashing and Computerized Marks

D. Security Dangers

This section looks at the top seven security risks that blockchain innovation faces. When those attacks are successfully carried out, they result in a shortage of funds or the resignation of certain blockchain hub administrators.

E. 51% Assault

An adversary may launch a 51% assault once they possess 51% of the hashing power. This attack starts by surreptitiously creating a chain of blocks that is noticeably different from the chain's central point. The separated chain is then presented to the organization and circulated as an authentic chain. This works perfectly with the two-pronged spending attack. Aggressors will try to move the longest chain by persuading the affiliation communities to join their chain in the case that they have at least 51% of the hashing power, as the blockchain technique recognizes the longest chain rule [5]. However, obtaining 51% of the hashing power is not necessary; if aggressors fail to obtain at least 50% of the hashing power [12], the double spending attack remains viable albeit with a lower likelihood of success. The attack gets more intense the more notable the hashing power of the entire blockchain network is. As a result, automated financial systems that use hashing are expected to be more resistant to the 51% attack.

F. Long-Reach Assault

A long-range assault is a method that shows itself due of the poor subjectivity idea. This assault strategy offers a correlation technique for handling the 51% attack. Nonetheless, rather of overcoming the 6-block attestation, it will for the most part fork the chain from the main block. The chance of this attack is basically non-existent in piece cash, however it very well may be lethal to the settlement on affirmation of stake (PoS) and allocated proof of stake (DPoS). Taking into account a PoS understanding in which a couple of aggressors start with a proper number of coins not long after the Starting block, they may clandestinely mine their sort of chain to lead the assault. Thinking about their restricted stake, they are permitted to provide limited blocks in the start of the cycle; nevertheless, as the cycle advances, they will be allowed to shape a more widened chain. Since PoS doesn't designate an end for chain extension, the chain might develop eternally.

II. RELATED WORK

(A,Barredo Arrieta 2020) In the last few years, artificial intelligence, or computer-based intelligence, has grown to be a powerful force that, when used appropriately, can express the best practices in many other fields of application. The entire local area must overcome the obstacle of make sense of capacity in AI for this to occur naturally. This is an inherent problem of the most recent procedures introduced by sub-imagery (such as outfits or Profound Brain Organizations), which were lacking in the previous advancement of simulated intelligence (more specifically, master frameworks and rule-based models). The field of purported logical simulated intelligence (XAI), which is widely acknowledged as essential to the functional organization of computer-based intelligence models, is where standards hiding this issue belong. This article's outline examines the writing that is currently being done and the commitments that have already been completed in the field of XAI, as well as potential future directions **[6]**.

(Y. Song, Y, 2020) Despite being widely used in many rush-hour traffic jams, vehicular worldwide positioning systems (GPS) fall well short of path level situating due to their poor accuracy. Current helpful situating (CP) techniques have improved vehicle positioning accuracy to varying degrees, but they struggle to further improve the security and robustness of the system. In this paper, we present a novel architecture for the Web of Vehicles (IoV) combining blockchain technology and CP to improve the accuracy, robustness, and security of vehicle GPS locating. First, a self-situating remediation scheme is suggested for intelligent cars to improve their positioning accuracy. This scheme uses the multitrophic signs as reference points to address the vehicle position (provided by GPS) through deep neural network (DNN) computation **[7]**.

(R. Gupta, S. Tanwar 2019) Telesurgery offers the enormous potential to provide continuous, meticulous medical care over a remote communication station with excellent accuracy to distant or isolated places. With more advanced accuracy and precision in determination strategies, it benefits the general public. Nevertheless, the security, protection, and interoperability problems with the current telesurgery framework limit its suitability for use in medical services jogs over the globe in the future **[8]**.

(J. Wang, C. Jiang 2020) Future distant enterprises should expect to offer a wide range of complex persuasive applications in both military and civilian domains, where clients can benefit from high-quality, low-cost, less intrusive, and reliable data administrations. Due to the complicated and heterogeneous character of the organization designs and remote administrations, achieving this ambitious goal calls for new radio tactics for adaptable learning and astute decision making. Artificial Intelligence (ML) algorithms have made remarkable strides in enabling massive data analysis, efficient boundary evaluation, and user-friendly navigation. As a result, we discuss controlled learning, unaided learning, support learning, and profound learning in this article's assessment of the thirty years of machine learning history. Additionally, we examine their efforts in the persuasive applications of remote organizations, such as machine-to-machine (M2M), Web of Things (IoT), mental radios (CR), heterogeneous organizations (Het Nets), etc **[9]**.

III. METHOD

For this audit, a calculated arrangement study was chosen as the assessment method. A fruitful planning study aims to provide a framework for an evaluation area, disperse in case confirmation of investigation exists, and quantify the amount of evidence. We adhere to the deliberate arrangement technique in this investigation. Furthermore, for a calculated composition overview, we employ rules. Look for large-format documents. Because our objective was to investigate the continuing assessments related to Blockchain development, we chose the precise arrangement procedure as our assessment approach.

A. Meaning of exploration questions

Interpreting the exploration questions is the first step in the precise planning process. This study's goal was to provide a summary of the momentum behind blockchain innovation research. Consequently, we defined four research questions:

- **RQ1:** What research subjects have been tended to in Blockchain force research?

Understanding the evolution of Blockchain research topics is the main investigative question of this planned review. If we were to compile all of the foundational papers from reliable information sources, we could create a basic understanding of Blockchain evaluation and direct the recurrent pattern focus on areas [10]. Streamlining the study on Blockchain development would enable various subject matter specialists and experts to gain a deeper understanding of the energy research areas, thereby advancing the Blockchain assessment significantly.

- **RQ2:** What applications have been created because of Blockchain advancement?

Most people associate blockchain with mausoleum cash from Spot currency. Blockchain development is included in bit coin monetary trades. Coin tomb currency, however, isn't the core strategy that makes advantage of Blockchain innovation. Therefore, it is essential to locate ongoing applications created with Blockchain technology. Understanding different categories and techniques to using Blockchain might be aided by perceiving different applications.

IV. RESULT AND DISCUSSION

In this section, we examine how well five security validation techniques work against the 51% attack. There are research models among them that ought to be used on actual structures. An assessment framework is shown in Figure 6, where we classify each security cycle's gamble sections as low, medium, or high [11]. The security-based framework also illustrates the advantages, difficulties, costs, and practicality of the guiding techniques.

Understanding the back-and-forth movement of exploration openings is made easier with a coordinated investigation method. The identifiable evidence of assessment opportunities will assist various experts and professionals in focusing their investigation on areas that need further investigation. Finding gaps in the literature will aid in understanding and reveal unanswered exploratory questions. Blockchain development.

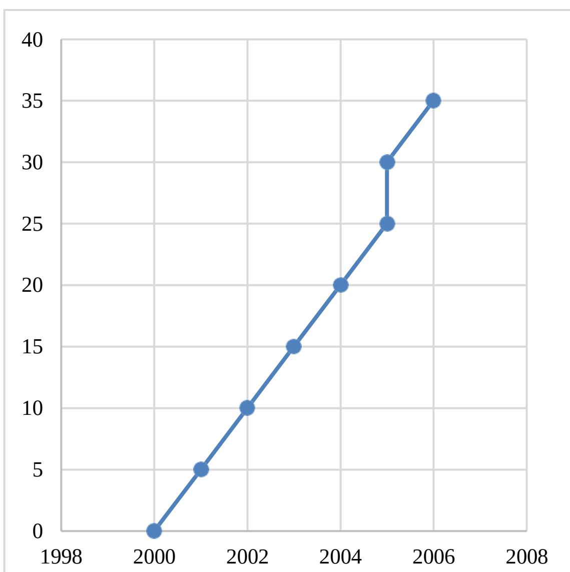


Fig .2 Publication year of the selected primary papers.

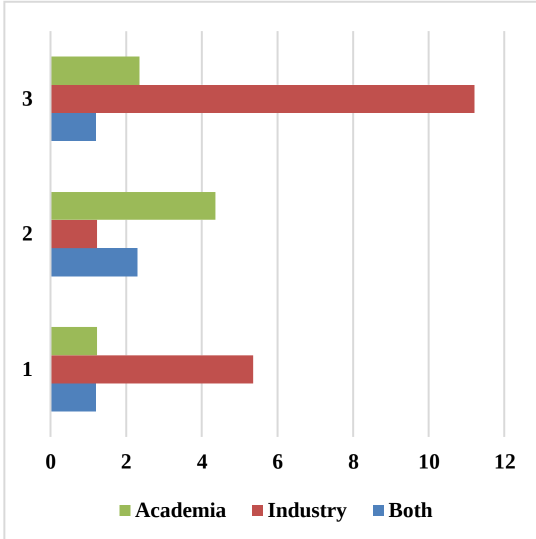


Fig.3 Wellspring of the chose essential papers.

The distribution of the chosen studies' geological features is shown in Figure 3. Institutions or organizations in the United States produced the majority of the papers (13, 31%) [12]. Switzerland had the most normal distribution with 5 papers (12.2%) and Germany with 6 articles (14.6%) after that. Less than four articles were sent to the other nations. The scholarly community's worldwide interest in Blockchain innovation is evident from the geographical distribution of the chosen pertinent publications.

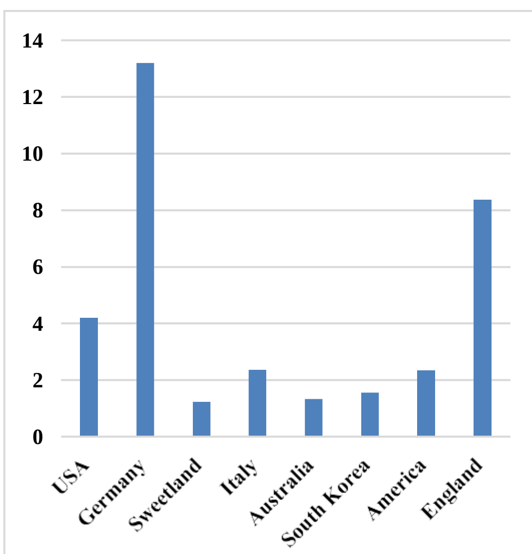


Fig .4 Geographic dissemination of the chose essential papers.

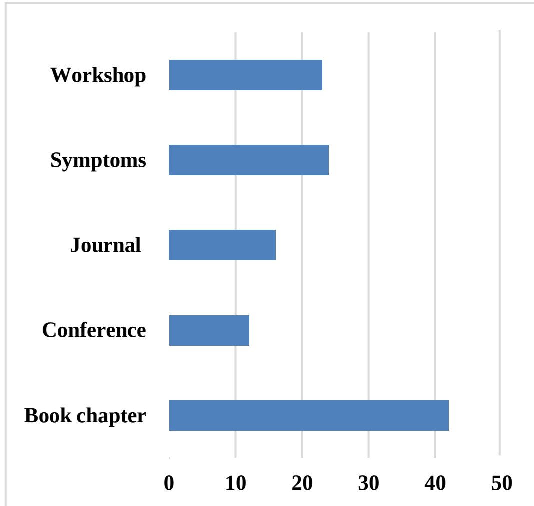


Fig 5 Distribution type and channel

Fig 5 shows the distribution kind of the chose papers. Distribution type implies the channel where the paper has been distributed **[13]**. The distribution types remembered for this planning study were gathering, diary, studio, conference, and book part.

Determination predisposition alludes to the twisting of measurable investigation inferable from the models used to choose the distributions **[14]**. We resolved this issue by planning our hunt convention cautiously. We likewise directed a pilot search with various catchphrases, to guarantee that we included however many papers as could be expected under the circumstances in this planning study. We characterized thorough incorporation and avoidance standards, to guarantee that every one of the chose papers were essential for our examination point, and addressed the exploration questions. Notwithstanding, there is one significant impediment that should be tended to.

V. CONCLUSION

Blockchain technology powers electronic cash in the form of piece coins. It is a decentralized environment for trading in which every transaction is recorded and visible to all users in an open report. Giving its customers privacy, security, certainty, and transparency is the only goal of blockchain technology [15]. All the same, these traits present a great deal of unique challenges and restrictions that need to be addressed.

In this paper, we provide a comprehensive overview of the foundational data on blockchain and artificial intelligence, conduct a top-to-bottom analysis of the feasibility of coordinating blockchain and artificial intelligence, and thoroughly compile the research on the integration of blockchain and AI in both domestic and international markets.

VI. FUTURE WORK

Blockchain and artificial intelligence are two of the most cutting-edge technological developments. Despite their individual advantages, they have the potential to open up important doorways that could revolutionize data innovation in the future.

Finally, we discuss future work and possible application scenarios.

VII. REFERENCES

[1]. W. Sun, N. Xu, L. Wang, H. Zhang, and Y. Zhang, "Dynamic digital twin and federated learning with incentives for air-ground networks," IEEE Transactions on Network Science and Engineering, p. 1, 2020.

[2]. X. Hou, Z. Ren, J. Wang et al., "Reliable computation offloading for edge computing-enabled software-defined IoV," IEEE Internet of Things Journal, vol. 7, pp. 7097–7111, 2020.

[3]. J. Wang, C. Jiang, Z. Han, Y. Ren, and L. Hanzo, "Internet of vehicles: sensing-aided transportation information collection and diffusion," IEEE Transactions on Vehicular Technology, vol. 67, no. 5, pp. 3813–3825, 2018.

[4]. J. Guo, Y. Zhou, P. Zhang, B. Song, and C. Chen, "Trust-aware recommendation based on heterogeneous multi-relational graphs fusion," Information Fusion, vol. 74, pp. 87–95, 2021.

[5]. M. B. Mollah, J. Zhao, D. Niyato et al., "Blockchain for the internet of vehicles towards intelligent transportation systems: a survey," IEEE Internet of Things Journal, vol. 8, pp. 4157–4185, 2020.

[6]. A, Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser et al., "Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI," Information Fusion, vol. 58, pp. 82–115, 2020.

[7]. Y. Song, Y. Fu, F. R. Yu et al., "Blockchain-enabled internet of vehicles with cooperative positioning: a deep neural network approach," IEEE Internet of Things Journal, vol. 7, no. 4, pp. 3485–3498, 2020.

[8]. R. Gupta, S. Tanwar, S. Tyagi, N. Kumar, M. Obaidat, and B. Sadoun, "Habits: blockchain-based telesurgery framework for healthcare 4.0," in Proceedings of the 2019 International Conference on Computer, Information and Telecommunication Systems (CITS), pp. 1–5, IEEE, Beijing China, August 2019.

[9]. J. Wang, C. Jiang, H. Zhang, Y. Ren, and K-C. Cheng, "Thirty years of machine learning: the road to Pareto-optimal wireless networks," IEEE Communications Surveys & Tutorials, vol. 22, pp. 1472–1514, 2020.

[10]. Ahmed, A., 2018. How Blockchain Will Change HR Forever.

[11]. Lafaille, C. What is Blockchain Technology? 2018.

[12]. Baliga, A. Understanding Blockchain Consensus Model. 2018.

[13]. Walker, H. How Digital Signatures and Blockchains Can Work Together. 2018.

[14]. Daoui, S. (2019). GDPR, Blockchain and the French Data Protection Authority: Many Answers but Some Remaining Questions.

[15]. Levi, S. D., and Lipton, A. B. (2018). An Introduction to Smart Contracts and Their Potential and Inherent Limitations. Skadden, Arps, Slate, Meagher & Flom LLP.

IN ASSOCIATION WITH

Published by Airo Journals
ISBN Number : 978-93-95305-72-3
Proceeding Available On
www.airo.co.in

DISCLAIMER

www.airo.co.in