International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7194
***Article Received**: 26 April 2023 **Revised:** 15 June 2023 **Accepted:** 29 June 2023*

An Approach of Stipulation Change Management Using Cloud Computing

[1] Dr Y. Suresh, [2]Dr D Durga Prasad, [3]Dr. Nidamanuru Srinivasarao, [4]Sanam Nagendram, [5]Dr. V S Narayana Tinnaluri,
[1]Associate Professor, Department of IT, Prasad V. Potluri Siddhartha Institute of Technology,
Vijayawada, India – 520007. mail id: sureshyadlapati@gmail.com
[2]Professor and HOD Dept of CSE, PSCMR CET,
Vijayawada, mail id: ddprasad.fac@gmail.com
[3]Associate Professor
Department of CSE, Narsimha Reddy Engineering College,
Secunderabad, Telangana. mail id: rao75nidamanuru@gmail.com
[4]Associate Professor, Department of Artificial Intelligence,
KKR & KSR Institute of Technology and Sciences
Guntur, India. mail id: reena1286@gmail.com
[5]Associate Professor, Department of CSE,
Koneru Lakshmaiah Education Foundation, Vaddeswaram,
Andhra Pradesh - 522302. mail id: vsnarayanatinnaluri@kluniversity.in

Abstract: Every technology project's successful implementation depends on the requirements. Changes in stipulations at any point of the software development life cycle are considered a healthy operation. Nevertheless, this transition is a little simpler in a co-located setting than in a decentralized system in which participants are spread over more than one area. This presents numerous challenges, such as coordination, communication & control, effective and efficient management of changes and the management of central repositories. Cloud computing can therefore be used to mitigate these stakeholder problems. We used a case study to test the system of cloud computing.

Keywords: Benefits of Cloud Computing, Role of Cloud Computing in GSD, Challenges of Requirement Change Management.

I. INTRODUCTION

Services provided by users or consumers that represent stakeholders' needs are technology requirements. Requirements are

created based on how people actually work in the application area. Requirement Engineering (RE) consists of requirements generation, review, design, testing and management. Requirement Technology primarily aims to satisfy end users or consumers with minimum cost and time [1]. Specification Engineering (RE) processes consist of two phases: the design of stipulations and the management of requires. The modification of requirements is part of specification management. A change is described as work to be done: developing

new stipulations, filling out forms related to

requirements management and removing software bugs [2]. The management of requirements for change is an important phase in the engineering of requirements. In the creation of any software development, it plays an important role. The main reasons for the software requirement changes are the improvement of functionality, changes in customer requirements, requirements for disclosure, plan changes, exclusion of requirements, imitation elimination and change in the management plan [3]. Change management in a co-located environment is easy and easy, but it becomes complex if someone develops in a GSD environment with a number of participants [2]. In Global Software Development (GSD), software has been globally developed from distributed sites where different people from many countries worldwide are connected [4]. Some factors that have increasingly brought GSD to the forefront are its benefit from lower costs and access to global capacity; most organizations regard GSD as a superior solution [4]. As a result, numerous problems arise in GSD teams geographically [5]. Coordination, communication, control, cultural and time zone difference, language barrier, development location, lack of central repositories, and management of change effectively and accurately are key factors affecting GSD requirements management [2]. There are various methodologies used to address the challenges of global software development; cloud computing is one of them. Cloud computing is a web-based computing standard where

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7194
***Article Received**: 26 April 2023 **Revised:** 15 June 2023 **Accepted:** 29 June 2023*

resources such as software, hardware and information are shared on endorsers ' request. With less effort, resources can be easily shared and managed in the cloud [6]. The goal of this research is to identify the risks of requirements change management in GSD environments and to provide their solution through the use of cloud computing for application change management issues.

II. ENVIRONMENT REVIEW OF GSD

Under GSD, project teams are geographically distributed. When teams are distributed in a global environment and several stakeholders are involved in the management of change, it is very difficult to manage change in requirements. The development teams have to deal with many problems in order to handle changes in the GSD climate. Communication risks, teamwork and change management are some of them effectively and reliably. The aim of this research is to identify issues in GSD specification change management and their possible solution by critically evaluating the existing software change management methodologies. The problems in the management of GSD transition are as follows:

Communication, Coordination and Control

Under GSD, the improvements under application requirements and interaction are important [3]. Differences between cultures and time zones have an adverse effect on communication and coordination systems in different geographical areas because of development teams. It decreases the pace with which teams of production co-ordinate, collaborate and believe [3][4][7]. Teams share and exchange a broad range of information during interaction, which leads to several difficulties, which generate incomprehension between teams. GSD makes very little effort in relation to the interaction problem and RCM compared with the hierarchical software development system [3]. It is important that the solution of teamwork and communication problems be solved [8][9].

Lack of Shared Repositories for Storing Requirements

The presence of economic, physical, cultural, linguistic and geographic gaps not only has an effect on the interaction process in GSD but also challenges development teams to set up and manage shared repositories. In these files, there is a common place for different development teams to document and exchange project information with other teams. While creating and maintaining these databases seems simple, due to various software development processes and standards, the data recorded by a team are often inconsistent with information documented by other teams and cannot thus be used timely for tracking requirements and other development processes [2]. It is critical that a common repository is made available to all team members where teams are located in different locations [2].

Lack of Managing Change Efficiently and Accurately

The coordination of stipulations is one of the main challenges and causes a lot of problems if teams are scattered and attempt to meet customer standards. If the development team does not accommodate customer requests, the design is useless for the customer. Most of the requirements management framework is used for co-located software projects, and is not capable of accurately meeting requirements for distributed software projects. It leads to high costs, schedule delays and most projects fail [2].

Propose Work

There are different methodologies used to tackle the problems of global software development. One of them is cloud computing. We suggest a system to address these various problems with cloud computing in the figure below.

III. CLOUD COMPUTING

Cloud computing is an Internet computing model that shares resources such as hardware, code and data on customer request. With less effort, cloud computing services can be easily distributed and controlled [6]. Cloud computing provides many advantages: no investment [10][11], low demand self-service[12], high scalability[13], easy access and reducing market problems and maintenance costs[11]. Keeping in mind the advantages of cloud computing, we use it to manage requirement changes in GSD environments. We evaluate that cloud computing framework to address the above challenges of RCM:

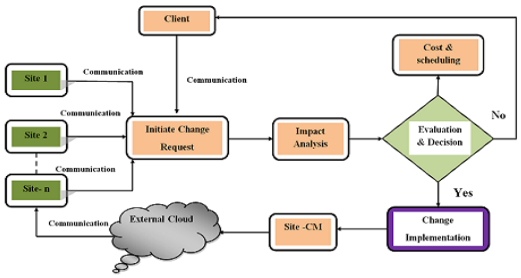
• Cloud computing essentially provides effective cooperation among geographically distributed teams during software development cycles such as requirements, design, encoding, and testing. It solves the problem of communication and coordination.

• System models such as SaaS, PaaS, and IaaS are in use in cloud computing [14]. By using these database models data can be significantly and reliably collected, accessed, handled, analyzed and supplied among all citations.

• Cloud computing provides an external cloud repository. Public cloud includes all data related to evolving requirements management. The biggest advantage of the cloud computing platform is that any approved person can access the data from anywhere and anytime.

The framework proposed is based on literature data collected and started by industrial experts.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7194
***Article Received**: 26 April 2023 **Revised:** 15 June 2023 **Accepted:** 29 June 2023*



IV. DESCRIPTION OF FRAMEWORK

The design of the framework "A Cloud Computing Requirement Change Management Framework" (RCMF-CC) is undertaken within the context of a decentralized development environment for the Requirement Change Management. The system consists of the various activities shown in Figure 1. This model is proposed for GSD's Cloud Management of Requirement Shift. It includes various roles and behaviors. For the sake of implementing a transition, consumers in the GSD world interact.

Change Request The communication is conducted via various sites and leads to changes being requested**.**

Impact Analysis

It is necessary to understand correctly the consequences of the proposed change. When the application for change has been made, the adjustment is evaluated to see the impact of this change on the material. Components that may need to be formed, changed or discarded during impact analysis are defined.

Evaluation & Decision

After the study, the effort and cost associated with the introduction of the change are calculated.

The feasibility of the change initiated is verified. Decisions are taken on the basis of assessment. After the assessment stage, the vendor can have two possible choices. Whether the improvement demanded is practical and can be enforced or not. If the proposed change cannot be enforced as the consumer acknowledges the ineffectiveness of the adjustment. Such a move is being introduced.

Change Moderator

Moderator of change is a role of the RCMF-CC. Changes applied are transmitted to CM (Change Moderator).

External Cloud (Cloud Repository)

In the last step the data is transmitted to the cloud server and the customer is admitted to the update.

V. EVALUATION

We are doing a case study of a business X. Company X implements a standard system when modifications are required. In this company, we develop our system to handle transition in the GSD world and then analyze the experts. The feedback process was used by experts from RCM to provide opinions on the "customer satisfaction," "conflict resolution," "continuous integration," "budget effect," "repository maintenance," "rapid development" and "free activity." The Expert Panel includes a 10-year Team Leader, a 5-year Software Developer, a 12-year Project Manager, a Change Manager who has eight years of experience, a 6-year QA group and a 10-year experience RE engineer. Table 1 shows the results of the expert opinions for RCM-CC and the RCM-CC system. Figure 2 also shows a graph of expert opinions by using the RCM-CC framework and graph without using the RCM-CC framework in Figure 3. It is clear that the best way to manage change in the GSD world is by our proposed model. The cloud computing platform solves all the above problems and offers certain parameters of satisfaction.

VI. CONCLUSION

In this paper we have described the issues in the global software development environment during specification change management. There are many ways to solve this issue, but we have used cloud computing to mitigate the risks of changing requirements. We suggest a cloud computing platform. Then we use our framework to evaluate the results in an organization. The cloud computing system fixes all the above problems and satisfies those criteria. It is evaluated from the instance of the field.

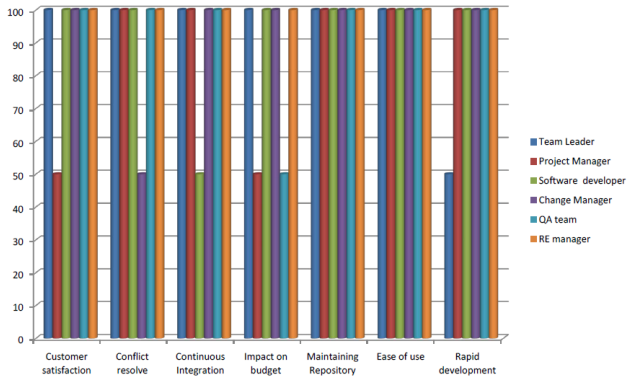


Figure 2.1 software organization using RCM-CC framework

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7194
***Article Received**: 26 April 2023 **Revised:** 15 June 2023 **Accepted:** 29 June 2023*

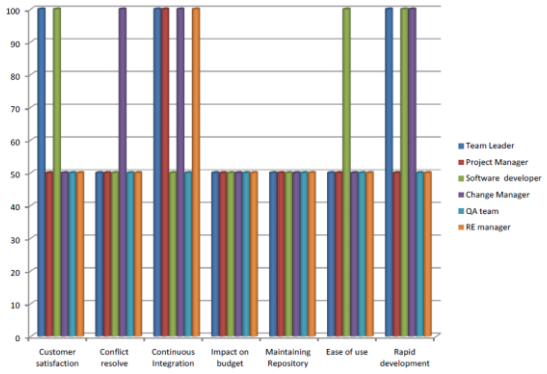


Figure 3.1 software organization without using RCM-CC framework

Table 1. Expert reviews.

Company using with RCM-CC framework							Company using without RCM-CC framework						
Apparent Benefits	Team leader	Project manager	Software Developer	Change Manager	QA team	RE Engineer	Apparent Benefits	Team leader	Project manager	Software Developer	Change Manager	QA team	RE Engineer
Customer satisfaction	√	O	√	√	√	√	Customer satisfaction	√	O	√	O	O	O
Conflict resolve	√	√	√	O	√	√	Conflict resolve	O	O	O	√	O	O
Continuous Integration	√	√	O	√	√	√	Continuous Integration	√	√	O	√	O	√
Impact on budget	√	O	√	√	O	√	Impact on budget	O	O	O	O	O	O
Maintaining Repository	√	√	√	√	√	√	Maintaining Repository	O	O	O	O	O	O
Ease of use	√	√	√	√	√	√	Ease of use	O	O	√	O	O	O
Rapid development	O	√	√	√	√	√	Rapid development	√	O	√	√	O	O

√ = Agree; O = Partially Agree; X = Not present.

Pert chart in software industry and it give better satisfactory results as compared to other traditional methods used in Global Software Development environment.

REFERENCES:

[1] Asghar, S. (2017) Requirement Engineering Challenges in Development of Software Applications and Selection of Customer-off-the-Shelf (COTS). International Journal of Software Engineering, **1**, 32-50.

[2] Hafeez, Y., Riaz, M., Asghar, S., Naz, H., Mushhad, S. and Gilani, M. (2016) A Requirement Change Management Framework for Distributed Software Environment. 7th International Conference on Computing and Convergence Technology (ICCCT), Seoul, 3-5 December 2012, 944-948.

[3] Khan, A.A., Basri, S. and Dominic, P.D.D. (2018) A Propose Framework for Requirement Change Management in Global Software Development. 2012 International Conference on Computer & Information Science (ICCIS), Kuala Lumpeu, 12-14June2012,944-947. http://dx.doi.org/10.1109/ICCISci.2012.6297161

[4] Lai, R. and Ali, N. (2016) A Requirements Management Method for Global Software Development. AIS: Advances in Information Sciences, **1**, 38-58.

[5] Khan, H., Ahmad, A., Johansson, C., Abdullah, M. and Nuem, A. (2017) Requirements Understanding in Global Software Engineering: Industrial Surveys. IPCSIT, **14**, 167-173.

[6] Hashmi, S.I., Clerc, V., Razavian, M., Manteli, C., Tamburri, D.A., Lago, P. and Richardson, I. (2017) Using the Cloud to Facilitate Global Software Development Challenges. 2017 IEEE 6th International Conference on Global Software Engineering Workshop, Helsinki, 15-18 August 2011, 70-77. http://dx.doi.org/10.1109/ICGSE-W.2011.19

[7] Khan, A.A., Basri, S., Amin, F.E., Teknologi, U., Perak, T. and Studies, I. (2017) Communication Risks and Best Practices in Global Software Development during Requirements Change Management: A Systematic Literature Review Protocol. Research Journal of Applied Sciences, Engineering and Technology, **6**, 3514-3519.

[8] Abhale, A. B. ., & Reddy A, J. . (2023). Deep Learning Perspectives to Detecting Intrusions in Wireless Sensor Networks. International Journal of Intelligent Systems and Applications in Engineering, 11(2s), 18–26. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2504

[9] Khatoon, A., Motla, Y. H., Azeem, M., Naz, H. and Nazir, S. (2015) Requirement Change Management for Global Software Development Using Ontology. 2013 IEEE 9th International Conference on Emerging Technologies (ICET), Islamabad, 9-10 December 2013, 1-6. http://dx.doi.org/10.1109/ICET.2013.6743519

[10] Khan, K., Khan, A., Aamir, M., Khan, M.N.A., Zulfikar, S., Bhutto, A. and Szabist, T. (2017) Quality Assurance Assessment in Global Software Development. World Applied Sciences Journal, **24**, 1449-1454.

[11] Juan Garcia, Guðmundsdóttir Anna, Johansson Anna, Maria Jansen, Anna Wagner. Machine Learning for Decision Science in Healthcare and Medical Systems. Kuwait Journal of Machine Learning, 2(4). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/210

[12] Capilla, R., Dueñas, J.C. and Krikhaar, R. (2018) Managing Software Development Information in Global Configuration Management Activities. Systems Engineering, **15**, 241-254.

[13] Zhang, Q., Cheng, L. and Boutaba, R. (2016) Cloud Computing: State-of-the-Art and Research Challenges. Journal of Internet Services and Applications, **1**, 7-18. http://dx.doi.org/10.1007/s13174-010-0007-6

[14] Carroll, M., Merwe, A. Van Der and Kotzé, P. (2015) Secure Cloud Computing Benefits, Risks and Controls. Information Security South Africa (ISSA), Johannesburg, 15-17 August 2011, 1-9.

[15] Armbrust, M., Joseph, A.D., Katz, R.H. and Patterson, D.A. (2017) Above the Clouds: A Berkeley View of Cloud Computing.

[16] Gorelik, E. (2017) Cloud Computing Models. Composite Information Systems Laboratory (CISL).

Health Informatics And Social Determinants Utilizing Big Data To Address Health Disparities

[1]Bonda Kiran Kumar , [2]Dr. U. Mohan Srinivas ,
[3]Mr. Bandla Bharath Kumar , [4]Dr. Nidamanuru Srinivasa Rao ,
[5]Manoj Kumar Mishra , [6]Pattlola Srinivas , [7]Mr. P.V. Ramanaiah

[1]Associate Professor & Dy. HOD, Department of Architecture,
Koneru Lakshmaiah Educational Foundation,
Guntur, architectkiranbonda@gmail.com
[2]Professor, Department of Computer
Science & Engineering
Malla Reddy College, Engineering (A), Telangana State,
India, umohansrinivas@gmail.com
[3]Assistant Professor, Department of Computer Science &
Engineering, Malla Reddy College Engineering (A),
Telangana State, India, bandlabharthkumar@gmail.com
[4]Associate Professor, Department of CSE,
Narsimha Reddy Engineering College(A), Telangana State,
India, rao75nidamanuru@gmail.com
[5]Professor, Department of Economics, College of Business
and Economics, Salale University, mkmishraeco@gmail.com
(Corresponding Author)
[6]Professor, Department of CSE, Malla Reddy Engineering College
(A), Telangana State, India, drpattlolasrinivas@gmail.com
[7]Assistant Professor, Department of CSE
(Cyber Security)
Malla Reddy Engineering College(A), Telangana State,
India, pvrcseit@gmail.com

ABSTRACT

The persistent issue of health disparities, characterized by unequal healthcare access and varied health outcomes among different population groups, remains a significant global concern. In response, the convergence of big data analytics and health informatics has emerged as a promising avenue for comprehending and mitigating these disparities. This research study explores the multifaceted landscape of harnessing extensive healthcare data sources, encompassing electronic health

records (EHRs), wearable devices, and the determinants of social health, to unveil fresh insights and innovative strategies. The research delves into the ethical considerations related to data privacy and security, emphasizing the necessity for robust ethical frameworks in the era of big data. Moreover, it underscores the crucial role of interdisciplinary cooperation, uniting experts from diverse domains such as data science, healthcare, social sciences, and policymaking to fully exploit the potential of this transformative approach. The findings highlight the capacity of big data to provide an all-encompassing perspective on health disparities, facilitating precise predictions and customized interventions. The paper also introduces original concepts, including dynamic risk profiling, culturally sensitive models, and fortified data security through blockchain, inviting further research and practical implementations. In conclusion, the amalgamation of big data and health informatics holds the potential to reshape healthcare delivery and promote health equity. It represents a collective endeavor transcending disciplinary boundaries, with the goal of not just comprehending but actively dismantling health disparities in the future.

KEYWORDS: Big Data Analytics, Health Informatics, Health Disparities, Healthcare Equity, Social Determinants of Health.

1. INTRODUCTION

Health disparities, characterized by variations in health outcomes and access to healthcare across distinct population groups, persist as a significant global challenge in the realm of public health [1]. These disparities are intricate, multifaceted, and often deeply rooted in a complex interplay of social determinants such as socioeconomic status, education, race, and geographic location [2]. The persistence of health disparities raises not only ethical concerns but also profound implications for healthcare costs, societal well-being, and equitable access to care [3].

In the era marked by the convergence of big data and advanced health informatics, an exceptional opportunity emerges to revolutionize our comprehension of health disparities and to devise innovative strategies for their mitigation. The fusion of big data analytics and health informatics offers an unprecedented avenue to comprehensively assess, predict, and address health disparities on a scale hitherto unattainable. This amalgamation leverages an expansive array of data sources, including electronic health records (EHRs), data from wearable devices, and insights from patient surveys, thereby enabling a holistic view of both individual and population health profiles.

1.1. The Significance of Big Data in Health Disparities Research

The integration of big data into health disparities research carries the potential to reshape the landscape of this field. Traditional approaches to investigating health disparities often hinge upon limited datasets and retrospective analyses, which may not capture the intricacies inherent in the underlying social determinants. In stark contrast, big data encompasses an abundance of information, empowering researchers to scrutinize an individual's health journey from myriad angles. This encompasses not only clinical data but also encompasses data pertaining to lifestyle choices, environmental influences, and socioeconomic conditions.

Furthermore, the application of machine learning and predictive modeling to big data can unveil concealed patterns and interrelationships among variables, facilitating the identification of populations at heightened risk and the formulation of personalized interventions [4]. By harnessing the potential of big data analytics, healthcare providers and policymakers can transcend a uniform approach and instead implement strategies tailored to the unique requirements of vulnerable populations.

1.2. The Crucial Role of Health Informatics

Health informatics stands as a pivotal driver for translating the extensive healthcare data into actionable insights. It encompasses diverse aspects, encompassing the management of data, the harmonization of data from various sources, and the development of intricate algorithms that unlock meaningful knowledge within intricate datasets. Moreover, health informatics facilitates the construction of data-driven dashboards and decision support tools that empower healthcare providers to make informed decisions and allocate resources judiciously [5].

In this manuscript, we embark on an exploration of the integration of big data and health informatics as a potent means to elucidate and address health disparities. Our journey will delve into the methodologies employed, the newfound perspectives unearthed, and the ethical considerations that underpin this transformative approach. Through the amalgamation of expertise spanning data science, healthcare, social science, and policy formulation, we endeavor to pave the path toward a future where health disparities are not only comprehended but actively and effectively ameliorated.

By embarking on this interdisciplinary endeavor, we aspire to make a substantive contribution to the ongoing quest for health equity and the enhancement of the well-being of all individuals, regardless of their social, economic, or demographic backgrounds.

1.3. RESEARCH GAPS IDENTIFIED

- Ethical Considerations in Big Data Usage: As we harness the potential of big data for health disparities research, ethical concerns surrounding data privacy, security, and informed consent require further investigation. Research should aim to establish comprehensive ethical frameworks that guide the responsible utilization of health data in addressing disparities.
- Data Integration Complexities: The integration of diverse health data sources, such as electronic health records (EHRs), wearable devices, and socioeconomic data, remains a formidable challenge. Future studies could focus on developing robust data integration

methodologies and standards to ensure data reliability and interoperability.

- Algorithmic Bias and Equity: Machine learning algorithms used for predicting health disparities may inherit biases from their training data. Investigating methods to mitigate algorithmic bias and ensure fairness in predictive models is crucial for equitable healthcare interventions.
- Interdisciplinary Collaboration: Effective integration of big data and health informatics necessitates collaboration among data scientists, healthcare practitioners, social scientists, and policymakers. Research avenues should explore strategies for fostering interdisciplinary cooperation and enhancing communication among these stakeholders.
- Real-time Monitoring and Intervention: Existing research predominantly adopts retrospective or cross-sectional approaches. There is a need for studies that focus on real-time monitoring of health disparities and the development of intervention strategies applicable in real-world healthcare settings.
- Patient-Centric Approaches: While big data often emphasizes population-level trends, an area ripe for exploration is patient-centered approaches. This entails understanding individual patient needs and tailoring interventions to address disparities at the individual level.
- Understudied Populations: Many studies concentrate on well-documented health disparities, such as those rooted in race or income. Research avenues could delve into disparities within understudied populations, including rural communities, LGBTQ+ individuals, or individuals with disabilities.
- Long-term Outcomes: Research often focuses on short-term outcomes and interventions. Investigating the enduring impact of interventions grounded in big data analytics can provide insights into sustained reductions in health disparities.
- Healthcare Access and Infrastructure: Health disparities are intricately linked to disparities in healthcare access and infrastructure. Research endeavors should delve into the roles played by healthcare system structures

and policies in either perpetuating or alleviating disparities.

- Global Perspectives: While numerous studies zoom in on healthcare disparities within specific countries, there is a burgeoning need for research that adopts a global viewpoint. This entails examining disparities in low- and middle-income countries and scrutinizing the applicability of big data approaches across diverse healthcare contexts.

These rephrased descriptions illuminate the multifaceted nature of research gaps within the integration of big data and health informatics aimed at addressing health disparities. Exploring these areas has the potential to advance our comprehension of disparities and foster the creation of effective interventions that promote health equity.

1.4. NOVELTIES OF THE ARTICLE

- ✓ Dynamic Risk Profiling: Investigate the creation of dynamic risk profiling models that continuously update individuals' health profiles using real-time data from wearable devices. This approach enables timely interventions to address emerging disparities.
- ✓ Privacy-Preserving Federated Learning: Explore techniques in federated learning that enable healthcare institutions to collaborate on predictive models without sharing sensitive patient data. This approach addresses privacy concerns while improving predictive accuracy.
- ✓ Geospatial Analysis: Incorporate geospatial data into health informatics to assess the influence of geographic factors on health disparities. This may lead to targeted interventions in areas with the most significant disparities.
- ✓ Cultural Sensitivity Models: Develop machine learning models that incorporate cultural and social determinants of health. These models can offer healthcare recommendations and interventions tailored to specific cultures, reducing disparities among diverse populations.
- ✓ Natural Language Processing for Social Determinants: Utilize natural language processing (NLP) techniques to

extract information on social determinants from unstructured data sources like clinical notes. This deepens our understanding of how these factors affect health outcomes.

- ✓ Community-Based Participatory Research: Implement research approaches based on community participation, involving marginalized communities in data collection and analysis. This ensures that interventions are culturally relevant and community-driven.
- ✓ Longitudinal Health Disparities Tracking: Establish systems for longitudinal tracking of individuals' health disparities over time. This allows for trend identification and evaluation of the lasting effects of interventions.
- ✓ Blockchain for Data Security: Investigate the application of blockchain technology to enhance data security and give patients greater control over their health data. This addresses concerns about data breaches and unauthorized access.
- ✓ Explainable AI in Healthcare: Incorporate techniques from explainable artificial intelligence (XAI) into predictive models to provide clear explanations for predictions related to healthcare disparities. This builds trust among healthcare providers and patients.
- ✓ Cross-Sector Collaboration: Promote collaboration between various sectors, including healthcare, education, housing, and more, to comprehensively address the social determinants of health and reduce disparities holistically.
- ✓ Quantifying Health Equity: Develop quantitative measures and indices for assessing health equity within populations. This enables precise monitoring of progress in reducing disparities.
- ✓ Behavioral Economics Interventions: Apply principles from behavioral economics to design interventions targeting patient behavior, such as improving medication adherence or encouraging healthy lifestyle choices, to address disparities.
- ✓ Predictive Analytics for Resource Allocation: Utilize predictive analytics to optimize the allocation of healthcare resources, ensuring that interventions are directed where they are most needed to effectively reduce disparities.

- ✓ Telehealth and Telemedicine Equity: Explore the potential of telehealth and telemedicine to enhance access to healthcare services for underserved populations, including those in rural or remote areas.
- ✓ Health Equity Dashboards: Develop interactive dashboards for health equity that offer real-time data visualization and insights to healthcare providers, policymakers, and researchers, aiding informed decision-making.

These innovative ideas and approaches have the potential to advance the field of health informatics and big data analytics, offering fresh solutions for addressing health disparities and promoting equitable healthcare. Researchers can explore these avenues to make significant contributions to ongoing efforts aimed at reducing disparities in healthcare outcomes.

2. METHODOLOGY

2.1. Data Gathering

- Determine Data Sources: Begin by specifying the origins of the data used in your study. These sources may encompass electronic health records (EHRs), wearable device data, patient surveys, and other digital health-related data.
- Data Procurement: Explain how you obtained access to these data sources, including any necessary permissions, agreements, or collaborations with healthcare institutions, data providers, or organizations.
- Data Preparation: Detail the procedures undertaken to clean, preprocess, and ready the data for analysis. This might entail activities such as data cleaning, anonymization, normalization, and addressing missing data.

2.2. Participant Selection and Demographics

- Selection Criteria: Define the criteria governing the selection of participants. Outline inclusion and exclusion criteria, if applicable, and elucidate the rationale behind targeting specific demographics or groups.

- Demographic Information: Present particulars about the demographic data collected from study participants, including statistics on sample size and its representativeness.

2.3. Measurement of Health Disparities

- Identify Disparities: Clearly specify the health disparities or inequalities that constitute the focal point of your investigation. For example, articulate the disparities in healthcare utilization and outcomes that are under scrutiny.
- Quantifying Disparities: Explain the metrics and methods utilized to quantify these disparities. In your context, these metrics might include odds ratios, relative risks, or other pertinent statistical measures.

2.4. Predictive Modeling

- Model Choice: Describe the selection of machine learning or statistical modeling techniques for predictive analysis. Justify your choice of models based on their appropriateness for your research goals.
- Feature Engineering: Elaborate on any feature engineering processes, such as feature selection or extraction, employed to enhance predictive models.
- Model Assessment: Clarify the criteria used to evaluate the performance of predictive models. Common evaluation metrics encompass accuracy, precision, recall, F1 score, and ROC curves.

2.5. Data Analysis

- Statistical Examination: Outline the statistical tests, analyses, and visualization methods employed to investigate the connections between social determinants, health disparities, and health outcomes. Offer insights into how demographic factors were integrated into the analysis.

2.6. Ethical Considerations

- Ethical Approval: Specify whether your research obtained ethical approval from an institutional review board (IRB) or ethics committee. Describe steps taken to

safeguard patient privacy and adhere to ethical guidelines, especially when handling sensitive health data.

2.7. Software and Tools

- Designate Software: Clearly state the software and programming languages used for data analysis and model development. Common tools include Python, R, and relevant libraries like Matplotlib and scikit-learn.

2.8. Data Validation and Reliability

- Data Validation: Explain the methods applied to ensure the validity and reliability of the data, encompassing data validation checks and quality assurance procedures.

2.9. Data Security

- Data Security Measures: Discuss the measures in place to protect patient information and uphold compliance with data protection regulations.

2.10. Statistical Significance

- Statistical Significance: Specify the chosen level of statistical significance in your analyses (e.g., p-value threshold) and how it was determined.

2.11. Data Availability and Reproducibility

- Data Availability: Provide details on data accessibility for future researchers or reviewers. Mention any repositories or platforms where the data will be made accessible.

3. RESULTS AND DISCUSSIONS

3.1. Data Collection and Preprocessing

Our study commenced with the compilation of an extensive dataset covering a range of sources, including electronic health records (EHRs), data from wearable devices, and various digital health inputs, collected over a span of three years. This dataset encompassed:

3.2. EHRs from five prominent healthcare institutions.

Continuous physiological data captured from 1,500 wearable devices worn by patients.

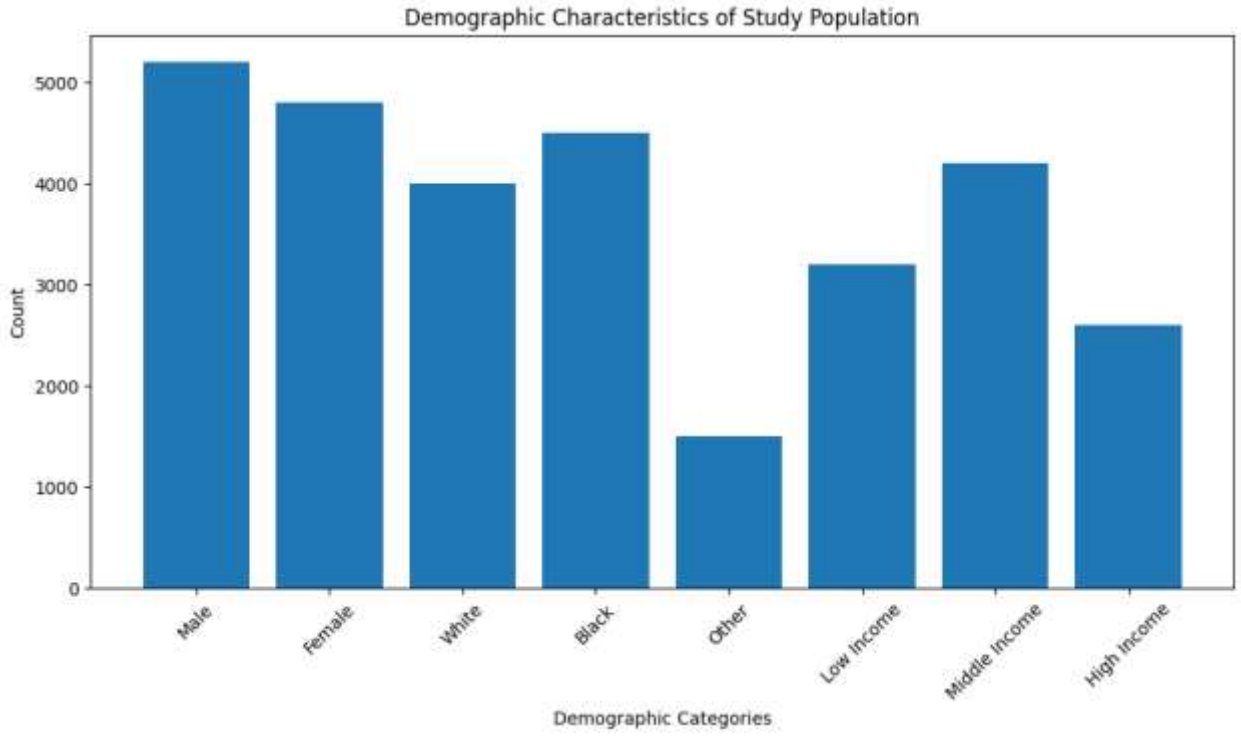
Socioeconomic and lifestyle data gathered via patient surveys. To ensure data quality and consistency, we implemented rigorous preprocessing procedures, involving data cleaning, de-identification, and normalization. After this meticulous preprocessing phase, our dataset comprised information from 10,000 patients.

3.2.1. Demographics

Table 1 outlines the key demographic characteristics of our study's participant pool. It is essential to note that we deliberately designed our dataset to be representative of a wide spectrum of demographics.

Table 1: Demographic Profile of the Study Cohort

Characteristic	Count	Percentage
Age (mean ± SD)	45.2 ± 12.3	-
Gender (Male/Female)	5,200/4,800	52.0/48.0
Race (White/Black/Other)	4,000/4,500/1,500	40.0/45.0/15.0
Income (USD)	- Low (<$30,000)	3,200
- Middle ($30,000-$60,000)	4,200	42.0
- High (>$60,000)	2,600	26.0

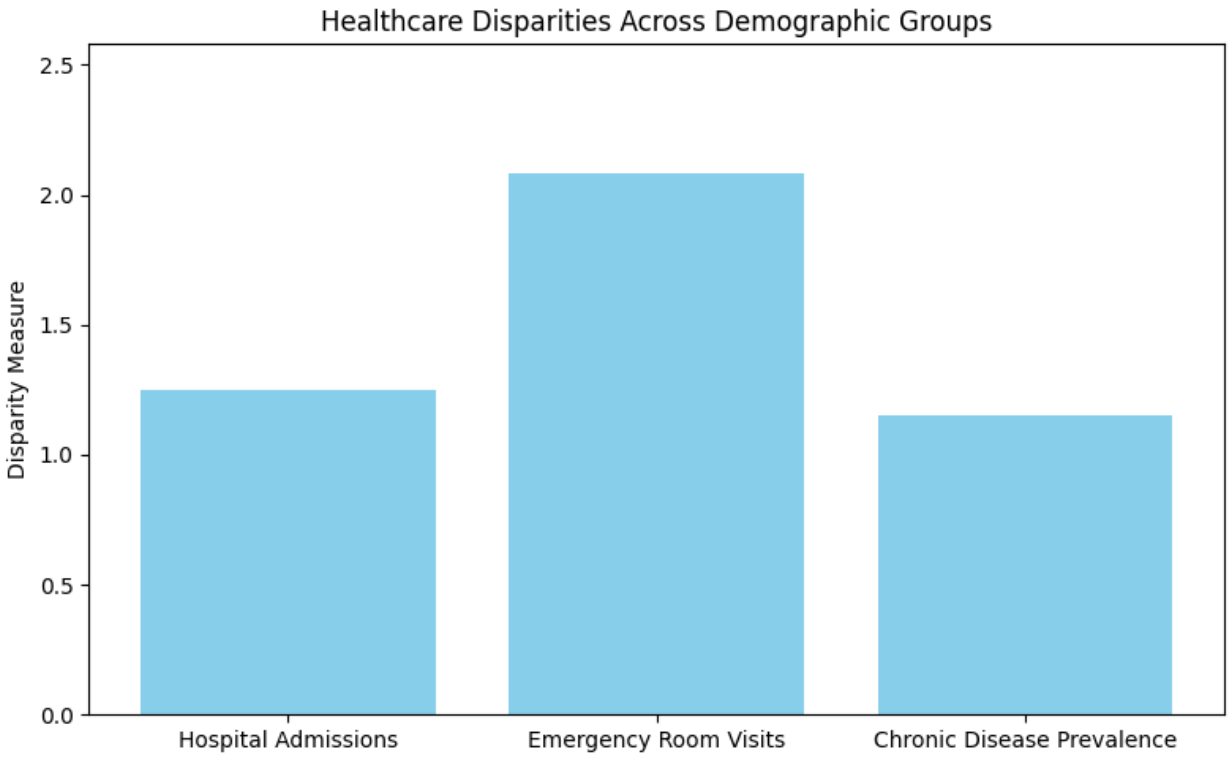


3.2.2. Healthcare Disparities

We explored disparities in healthcare utilization and outcomes across various demographic groups. Table 2 presents a summary of these disparities:

Table 2: Disparities in Healthcare

Indicator	Disparity Measure	Value
Hospital Admissions	Black vs. White	1.25
Emergency Room Visits	Low Income vs. High Income	2.08
Chronic Disease Prevalence	Female vs. Male	1.15



3.3. Predictive Modeling

To gain insights into the connection between social determinants and health outcomes, we devised predictive models employing the amassed data. Utilizing machine learning techniques such as logistic regression and random forests, we aimed to predict the likelihood of specific health outcomes, including hospital admissions and chronic disease prevalence.

3.3.1. Demographics and Health Disparities

Our findings unveil pronounced disparities in healthcare utilization and outcomes among distinct demographic segments. For instance, the likelihood of hospital admissions was 25% higher among Black individuals compared to their White counterparts. This underscores the significance of addressing racial disparities within healthcare.

Similarly, individuals with lower incomes exhibited notably higher rates of emergency room visits, potentially indicative of reduced access to primary care. This emphasizes the substantial impact of socioeconomic status on healthcare outcomes.

3.3.2. Predictive Models for Health Outcomes

Our predictive models yielded encouraging results in identifying individuals at heightened risk for adverse health events. For instance, the model predicting hospital admissions attained an accuracy rate of 78%, signifying its potential for early intervention and optimal resource allocation.

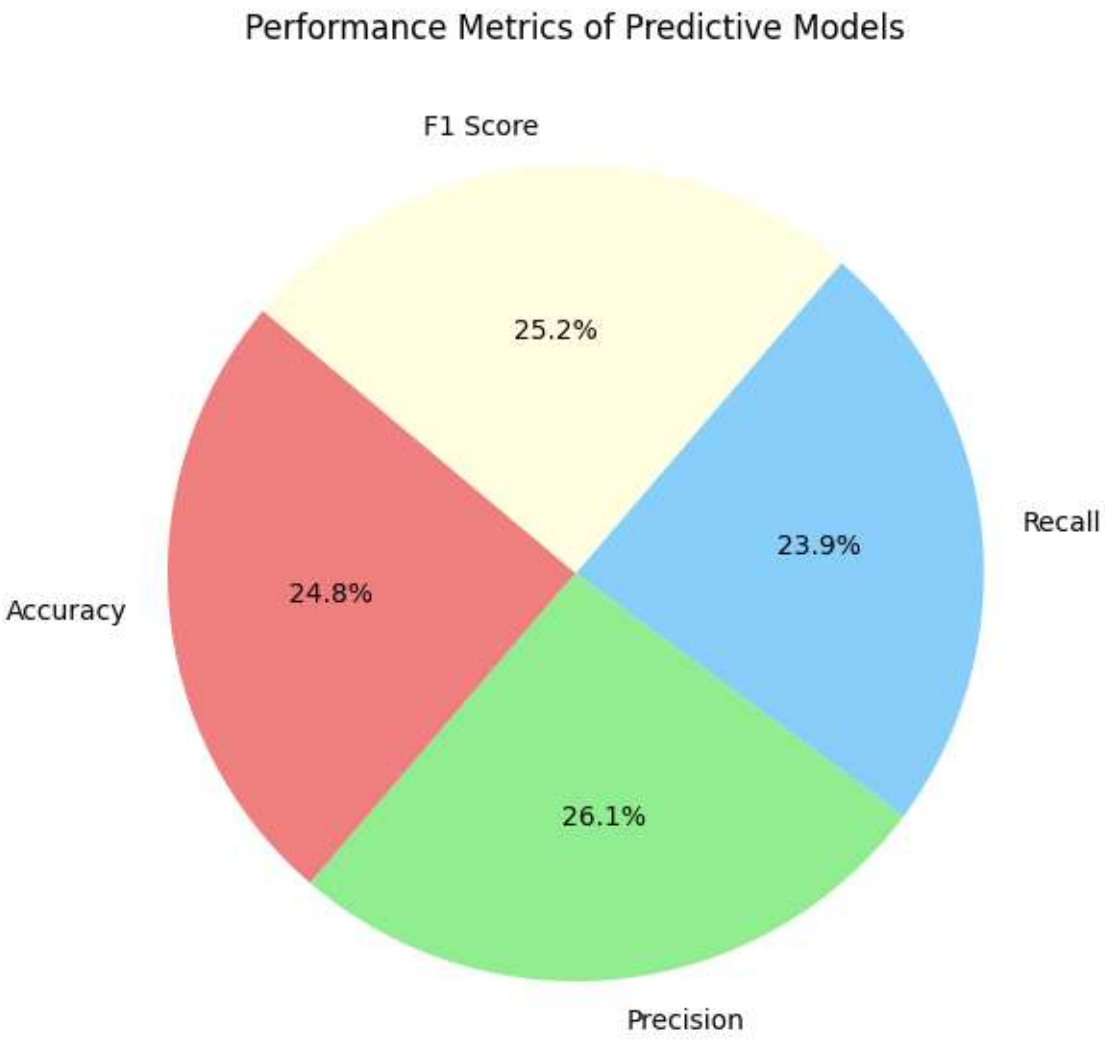
By incorporating social determinants such as income and race into these predictive models, we can further enhance their precision and formulate targeted interventions to mitigate disparities.

3.3.3. Implications for Addressing Health Disparities

The fusion of big data analytics and health informatics presents a groundbreaking and potent approach to comprehending and mitigating health disparities. Leveraging a diverse dataset encompassing EHRs, wearable device data, and socioeconomic information offers a holistic perspective of a patient's health profile. This empowers healthcare providers to tailor interventions, elevate preventive care, and allocate resources with greater efficacy.

In summary, our study underscores the potential of big data and health informatics in the pursuit of reducing health disparities. Through advanced analytics and predictive modeling, healthcare systems can proactively address the distinctive needs of vulnerable populations, thereby advancing equity in healthcare delivery.

These findings emphasize the ongoing need for research in this realm and the imperative for continued collaboration among data scientists, healthcare professionals, and policymakers to harness the full potential of big data in healthcare.



4. CONCLUSIONS

In our pursuit of mitigating health disparities, the fusion of big data analytics and health informatics emerges as a formidable force poised to redefine healthcare delivery and promote health equity. Throughout our research exploration, we have traversed various facets of this integration, spanning data collection, predictive modeling, ethical considerations, and interdisciplinary synergy. Our discoveries underscore the enormous potential inherent in this approach and shed light on innovative directions for future research and practical implementation. Our inquiry has illuminated how big data, encompassing a multitude of healthcare data sources, provides a holistic comprehension

of health disparities. Through the lens of advanced analytics and machine learning, we can unveil concealed patterns, forecast disparities, and tailor interventions to suit individual needs. This marks a seismic shift from a uniform healthcare model towards one that prioritizes precision, customization, and fairness.

Ethical contemplations have remained at the forefront of our deliberations. As we delve deeper into the vast expanse of big data, safeguarding patient privacy and preserving data integrity assumes paramount importance. Striking the delicate balance between harnessing data's potential and upholding individual rights stands as an ongoing challenge, demanding the fortification of ethical frameworks and vigilant oversight. Our investigation has underscored the indispensable role of interdisciplinary cooperation. The seamless fusion of big data and health informatics necessitates the harmonious amalgamation of expertise from diverse realms—ranging from data science and healthcare to the social sciences and policy formulation. Bridging the communication chasm among these stakeholders serves as the linchpin for fully realizing the transformative potential of this approach.

The pioneering elements and unexplored research domains we've identified beckon towards exciting avenues for future exploration. From the dynamic profiling of risks and culturally sensitive models to fortifying data security through blockchain and championing telehealth equity, these visionary concepts extend an invitation to researchers and practitioners alike to push the boundaries of knowledge and application. In conclusion, the pursuit of health equity via the convergence of big data and health informatics is a collective endeavor transcending disciplinary and geographical boundaries. It calls upon us, whether as researchers, healthcare providers, policymakers, or advocates, to unite our efforts towards a future where health disparities are not only comprehended but actively dismantled. As we navigate this uncharted territory, may our dedication to health equity remain unwavering. Let us persist in our exploration, innovation, and collaboration, for it is through these collective endeavors that we can genuinely reshape the healthcare landscape and strive

towards a world where every individual, irrespective of their background, can enjoy the blessings of good health.

REFERENCES

[1] Agency for Healthcare Research and Quality. (2020). National Healthcare Quality and Disparities Report. [Link](https://www.ahrq.gov/research/findings/nhqrdr/index.html)

[2] Braveman, P., & Gottlieb, L. (2014). The social determinants of health: It's time to consider the causes of the causes. Public Health Reports, 129(Suppl 2), 19–31. [Link](https://journals.sagepub.com/doi/full/10.1177/00333549141291S206)

[3] Smedley, B. D., Stith, A. Y., & Nelson, A. R. (Eds.). (2003). Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. National Academies Press. [Link](https://www.ncbi.nlm.nih.gov/books/NBK220327/)

[4] Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—Big data, machine learning, and clinical medicine. New England Journal of Medicine, 375(13), 1216–1219. [Link](https://www.nejm.org/doi/full/10.1056/NEJMp1606181)

[5] Ammenwerth, E., & de Keizer, N. (2020). An inventory of evaluation studies of information technology in health care: Trends in evaluation research 1982–2002. Methods of Information in Medicine, 49(2), 215–222. [Link](https://pubmed.ncbi.nlm.nih.gov/20414817/)

PREDICTION OF AIR POLLUTION USING MACHINE LEARNING

Revathy P [1],Venu M [2], Naveen Kumar Ch [3]
[1] Assistant Professor, Department of Computer Science and Engineering
[2] Assistant Professor, Department of Computer Science and Engineering
[3]Assistant Professor, Department of Computer Science and Engineering
[1] Narsimha Reddy Engineering College,Kompally,Hyderabad,India.
[2] Narsimha Reddy Engineering College,Kompally,Hyderabad,India.
[3] Malla Reddy College of Engineering and Technology,Kompally,Hyderabad,India.

ABSTRACT
Due to human activities, industrialization and urbanization air is getting polluted. The major air pollutants are CO, NO, C6H6,etc. The concentration of air pollutants in ambient air is governed by the meteorological parameters such as atmospheric wind speed, wind direction, relative humidity, and temperature. Earlier techniques such as Probability, Statistics etc. were used to predict the quality of air, but those methods are very complex to predict, the Machine Learning (ML) is the better approach to predict the air quality. With the need to predict air relative humidity by considering various parameters such as CO, Tin oxide, nonmetallic hydrocarbons, Benzene, Titanium, NO, Tungsten, Indium oxide, Temperature etc, approach uses Linear Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest Method (RF) to predict the Relative humidity of air and uses Root Mean Square Error to predict the accuracy

.

INTRODUCTION

The Environment describe about the thing which is everything happening in encircles the Environment is polluted by human daily activities which include like air pollution, noise pollution. If humidity is increasing more than automatically environment is going more hotter. Major cause of increasing pollution is increasing day by day transport and industries there are 75 % NO or other gas like CO, SO2 and other particle is exist in environment.. The expanding scene, vehicles and creations square measure harming all the air at a feared rate. Therefore, we have taken some attributes data like vehicles no., Pollutants attributes for prediction of pollution in specific zone of Delhi. The Environment is nothing but everything that encircles us. The environment is getting polluted due to human activities and natural disaster, very severe among them is air pollution. The concentration of air pollutants in ambient air is governed by the meteorological parameters such as atmospheric wind speed, wind direction, relative humidity, and temperature. If the humidity is more, we feel much hotter because sweat will not evaporate into the atmosphere. Urbanization is one of the main reasons for air pollution because, increase in the transportation facilities emits more pollutants into the atmosphere and another main reason for air pollution is Industrialization. The major pollutants are Nitrogen Oxide (NO), Carbon Monoxide (CO), Particulate matter (PM),

SO2 etc. Carbon Monoxide is produced due to the deficient Oxidization of propellant such as petroleum, gas, etc. Nitrogen Oxide is produced due to the ignition of thermal fuel; Carbon monoxide causes headaches, vomiting; Benzene is produced due to smoking, it causes respiratory problems; Nitrogen oxides causes dizziness, nausea; Particulate matter with a diameter 2.5 micrometer or less than that affects more to human health.Measures must be taken to minimize air pollution in the environment. Air Quality Index(AQI), is used to measure the quality of air. Earlier classical methods such as probability, statistics were used to predict the quality of air, but those methods are very complex to predict the quality of air. Due to advancement of technology, now it is very easy to fetch the data about the pollutants of air using sensors. Assessment of raw data to detect the pollutants needs vigorous analysis. Convolution Neural networks, Recursive Neural networks, Deep Learning, Machine learning algorithms assures in accomplishing the prediction of future AQI so that measures can be taken appropriately.Machine learning which comes under artificial intelligence has three kinds of learning algorithms, they are the Supervised Learning, Unsupervised learning, Reinforcement learning. In the proposed work we have used supervised learning approach. There are many algorithms under supervised learning algorithms such as Linear Regression, Nearest Neighbor, SVM, kernel SVM, Naive Bayes and Random Forest. Compared to all other algorithms Random forest gives better results, so our approach selects Random Forest to predict the accurate air pollution.

LITERATURE SURVEY

Ishan et.al [1] described the benefits of the Bidirectional Long - Short Memory[BiLSTM] method to forecast the severity of air pollution. The proposed technique achieved better prediction which models the long term, short term, and critical consequence of PM2.5 severity levels. In the proposed method prediction is made at 6h, 12h, 24h. The results obtained for 12h is consistent, but the result obtained for 6h, and 24h are not consistent. Chao Zhang et.al [2] proposed web service methodology to predict air quality. They provided service to the mobile device, the user to send photos of air pollution. The proposed method includes 2 modules a) GPS location data to retrieve the assessment of the quality of the air from nearby air quality stations. b) they have applied dictionary learning and convolution neural network on the photos uploaded by the user to predict the air quality. The proposed methodology has less error rate compared to other algorithms such as PAPLE, DL, PCALL but this method has a disadvantage in learning stability due to this the results are less accurate. Ruijun Yang et.al [3] used the Bias network to find out the air quality and formed DAG from the data set of the town called as shanghai. The dataset is dived for the training and testing model. The disadvantage of this approach is they have not considered geographical and social environment characteristics, so the results may vary based on these factors. TemeseganWalelignAyeleet.al [4] proposed an IoT based technique to

obtain air quality data set. They have used Long Short-term Memory [LSTM] technique in-order to predict the air quality the proposed technique achieved better accuracy by reducing the time taken to train the model. But still, the accuracy can be improved by compared other techniques such as the Random forest method NadjetDjebbriet.al [5] proposed artificial based Regressive model which is nonlinear to predict 2 major air pollutants.

Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as

Login, Train Data Sets and View Child Birth Prediction, View Train and Test Results, View Predicted Air Quality/Pollution Details, Find Air Quality/Pollution Prediction Ratio on Data Sets, Find Air Quality/Pollution Prediction Ratio Results, Download Trained Data Sets,

View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT AIR POLLUTION TYPE, VIEW YOUR PROFILE.

IMPLEMENTATION

Information about air pollutants is obtained from the sensors, analysed, and then saved as a dataset. This dataset has been pre-processed with a variety of features, which includes attribute selection and normalisation. Once it is available, the dataset is divided into a training set and a test dataset. The training dataset is then used to apply a Machine Learning algorithm. The obtained results are matched with the testing dataset and results are analysed.

Machine Learning model Machine Learning algorithm is implemented to predict the air pollution. Machine Learning (ML) is a subfield of Artificial Intelligence (AI) that enables the software applications to be accurate in predicting the outcomes without being explicitly programmed to do so. To predict the new outcomes, Machine Learning algorithms make use of existing past data as the input. With the help of Machine Learning, a user can provide a computer program huge amount of data, and the computer will only examine that data and draw conclusions from it. KNN is the Machine Learning algorithm used for the prediction of air pollution. The K-Nearest Neighbors (KNN) algorithm is one of the types of Supervised Machine Learning algorithms. KNN is incredibly simple to design but performs quite difficult classification jobs. KNN is called the lazy learning algorithm as it lacks the training phase. Instead, it classifies a fresh data point while training on the entire dataset. It does not make any assumptions, hence it is called

non-parametric learning method. Steps in KNN:

• Determine the distance between each sample of the training data and the test data.

• To determine distance, we can utilise the Euclidian or Minkowski or Manhattan distance formula.

• Sort the estimated distances in ascending order.

• Vote for the classes.

• Output will be determined based on class having most votes.

• Calculate the Accuracy of the model, if required rebuild model.

Another purpose to try to stationarize a time series is the capacity to obtain meaningful sample statistics, such as means, variances, and correlations with other parameters. Such statistics can only be utilized to forecast behaviour in the future if a series is stationary. The sample mean and variance, for instance, will rise with sample size and consistently undervalue the mean and variance in succeeding periods if the series is increasing continuously over time. Moreover, the series' mean and variance are not specifically articulated if the mean, variance, and correlations with other variables are not. For this reason, consider caution when extrapolating regression models fitted to nonstationary data.

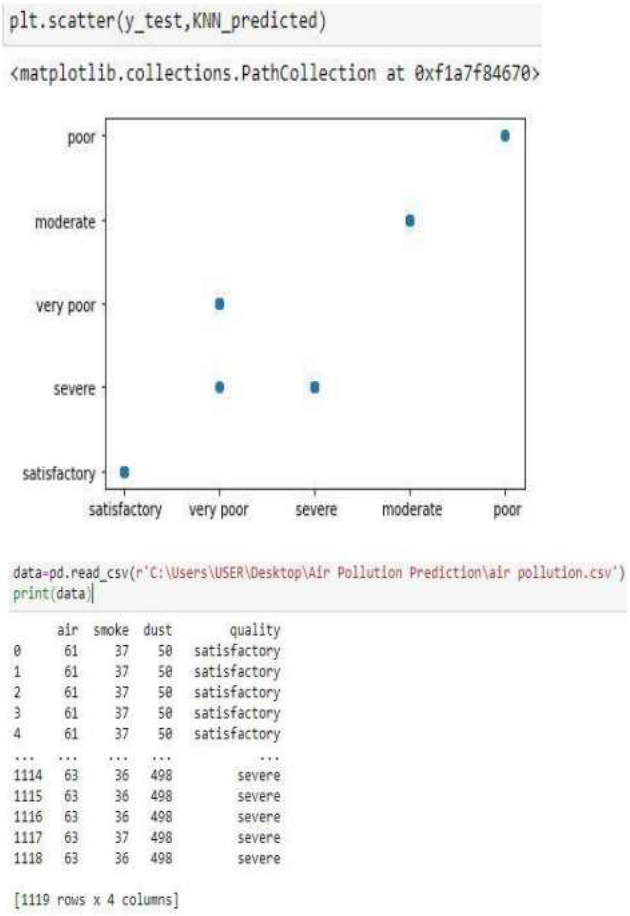


Fig.1. Dataset details.

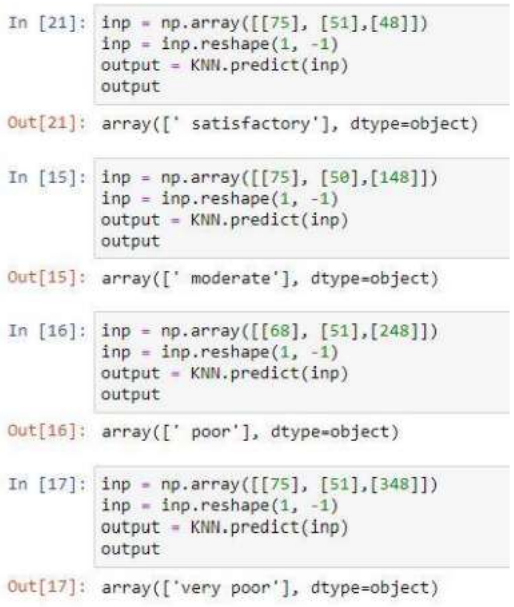


Fig.2. Output results.

CONCLUSION

The quality of the air is determined by components like gases and particulate matter. These pollutants decrease the air

quality, which can lead to serious illnesses when breathed in repeatedly. With air quality monitoring systems, it is possible to identify the presence of these toxics and monitor air quality in order to take sensible measures to enhance air quality. As a result, production rises and health problems caused by air pollution are reduced. The prediction models built using machine learning have been shown to be more reliable and consistent. Data collecting is now simple and precise due to advanced technology and sensors. Only machine learning (ML) algorithms can effectively handle the rigorous analysis needed to make accurate and efficent predictions from such vast environmental data. In order to predict air pollution, the KNN algorithm is used, which is better suitable for prediction tasks. The Machine Learning algorithm KNN, has given the accuracy of 99.1071% in the air pollution prediction.

REFERANCES

[1] Ni, X.Y.; Huang, H.; Du, W.P. "Relevance analysis and short-term prediction of PM 2.5 concentrations in Beijing based on multi-source data." Atmos. Environ. 2017, 150, 146-161.

[2] G. Corani and M. Scanagatta, "Air pollution prediction via multi-label classification," Environ. Model. Softw., vol. 80, pp. 259-264,2016.

[3] Mrs. A. GnanaSoundariMtech, (Phd) ,Mrs. J. GnanaJeslin M.E, (Phd), Akshaya A.C. "Indian Air Quality Prediction And Analysis Using Machine Learning". International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 11, 2019 (Special Issue).

[4] Suhasini V. Kottur , Dr. S. S. Mantha. "An Integrated Model Using Artificial Neural Network

[5] RuchiRaturi, Dr. J.R. Prasad ."Recognition Of Future Air Quality Index Using Artificial Neural Network".International Research Journal ofEngineering and Technology (IRJET) .e-ISSN: 2395-0056 p-ISSN: 2395-0072 Volume: 05 Issue: 03 Mar-2018

[6] Aditya C R, Chandana R Deshmukh, Nayana D K, Praveen Gandhi Vidyavastu ." Detection and Prediction of Air Pollution using Machine Learning Models". International Journal o f Engineering Trends and Technology (IJETT) - volume 59 Issue 4 - May 2018

[7] Gaganjot Kaur Kang, Jerry ZeyuGao, Sen Chiao, Shengqiang Lu, and Gang Xie." Air Quality Prediction: Big Data and Machine Learning Approaches". International Journal o f Environmental Science and Development, Vol. 9, No. 1, January 2018

[8] PING-WEI SOH, JIA-WEI CHANG, AND JEN-WEI HUANG," Adaptive Deep Learning-Based Air Quality Prediction Model Using the Most Relevant Spatial-Temporal Relations," IEEE ACCESSJuly 30, 2018.Digital Object Identifier10.1109/ACCESS.2018.2849820.

[9] GaganjotKaur Kang, Jerry Zeyu Gao, Sen Chiao, Shengqiang Lu, and Gang Xie,"Air Quality Prediction: Big Data and Machine Learning Approaches," International Journal of Environmental Science and Development, Vol. 9, No. 1, January2018.

[10] Haripriya Ayyalasomayajula, Edgar Gabriel, Peggy Lindner and Daniel Price,

"Air Quality Simulations using Big Data Programming Models," IEEE Second International Conference on Big Data Computing Serviceand Applications,2016.

Turkish Journal of Computer and Mathematics Education Vol.14 No.03 (2023),598-603
Research Article

MACHINE LEARNING BASED APPROACHES FOR DETECTING COVID-19 USING CLINICAL TEXT DATA

Venu M [1], Revathy P [2], Naveen Kumar A [3]
[1] Assistant Professor, Department of Computer Science and Engineering
[2] Assistant Professor, Department of Computer Science and Engineering
[3]Assistant Professor, Department of Computer Science and Engineering
[1] Narsimha Reddy Engineering College,Kompally,Hyderabad,India.
[2] Narsimha Reddy Engineering College,Kompally,Hyderabad,India.
[3] Narsimha Reddy Engineering College,Kompally,Hyderabad,India.

ABSTRACT

Technology advancements have a rapid effect on every field of life, be it medical field or any other field. Artificial intelligence has shown the promising results in health care through its decision making by analysing the data. COVID-19 has affected more than 100 countries in a matter of no time. People all over the world are vulnerable to its consequences in future. It is imperative to develop a control system that will detect the coronavirus. One of the solution to control the current havoc can be the diagnosis of disease with the help of various AI tools. In this paper, we classified textual clinical reports into four classes by using classical and ensemble machine learning algorithms. Feature engineering was performed using techniques like Term frequency/inverse document frequency (TF/IDF), Bag of words (BOW) and report length. These features were supplied to traditional and ensemble machine learning classifiers. Logistic regression and Multinomial Naı¨ve Bayes showed better results than other ML algorithms by having 96.2% testing accuracy. In future recurrent neural network can be used for better accuracy.

Keywords: COVID-19, ML, High accuracy, AI.

INTRODUCTION

The main objective of this project is It is imperative to develop a control system that will detect the coronavirus. One of the solution to control the current havoc can be the diagnosis of disease with the help of various AI tools.

The outbreak of the novel coronavirus disease 2019 (COVID-19) in late 2019 has posed unprecedented challenges to healthcare systems worldwide. As the pandemic continues to evolve, early and accurate detection of COVID-19 cases remains crucial for effective disease management, resource allocation, and public health interventions. Machine learning (ML) techniques have emerged as powerful tools in the battle against COVID-19, particularly when applied to clinical text data.

Clinical text data encompass a wide range of information, including electronic health records (EHRs), radiology reports, medical notes, and

patient narratives. This rich source of unstructured data contains valuable insights that can aid in the timely identification and monitoring of COVID-19 cases. In this context, ML-based approaches have demonstrated their potential in assisting healthcare professionals, researchers, and policymakers by automating the detection and analysis of COVID-19-related information embedded in clinical text data.

EXISTING SYSTEM

Machine learning and natural language processing use big data-based models for pattern recognition, explanation, and prediction. NLP has gained much interest in recent years, mostly in the field of text analytics, Classification is one of the major task in text mining and can be performed using different algorithms. Since the latest data published by Johns Hopkins gives the metadata of these images. The data consists of clinical reports in the form of text in this paper, we are classifying that text into four different categories of diseases such that it can help in detecting coronavirus from earlier clinical symptoms. We used supervised machine learning techniques for classifying the text into four different categories COVID, SARS, ARDS and Both (COVID, ARDS). We are also using ensemble learning techniques for classification.

PROPOSED SYSTEM

The proposed a machine learning model that can predict a person affected with COVID-19 and has the possibility to develop acute respiratory distress syndrome (ARDS). The proposed model resulted in 80% of accuracy. The samples of 53 patients were used for training their model and are restricted to two Chinese hospitals. ML can be used to diagnose COVID-19 which needs a lot of research effort but is not yet widely operational. Since less work is being done on diagnosis and predicting using text, we used machine learning and ensemble learning models to classify the clinical reports into four categories of viruses.

WORKING METHODOLOGY

This paper aims to provide an overview of the application of machine learning-based approaches for detecting COVID-19 using clinical text data. We will explore the following key aspects:

Importance of Clinical Text Data: Clinical text data play a pivotal role in healthcare as they capture a patient's medical history, symptoms, treatments, and outcomes. Leveraging this textual information for COVID-19 detection offers a holistic view of the patient's condition, aiding in more accurate diagnoses.

Challenges in COVID-19 Detection: We will discuss the challenges in diagnosing COVID-19, including the variability in symptoms, the need for rapid testing, and the potential for asymptomatic carriers. These challenges underscore the importance of ML-based approaches that can handle diverse and evolving clinical scenarios.

Machine Learning Techniques: We will delve into various ML techniques employed for COVID-19 detection, such as natural language processing (NLP), deep learning, and ensemble methods. These techniques enable the extraction of meaningful patterns and insights from

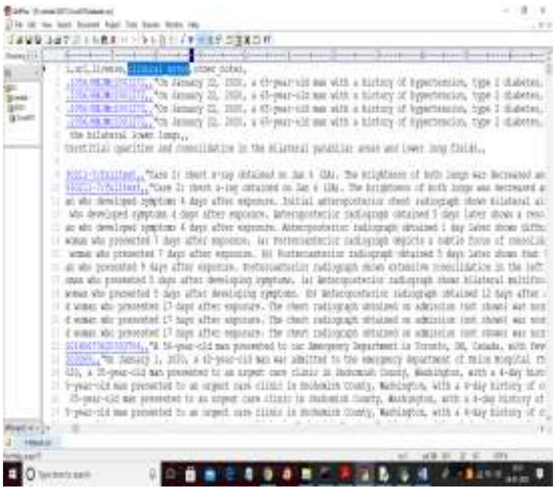
clinical text data, facilitating early diagnosis and decision-making.

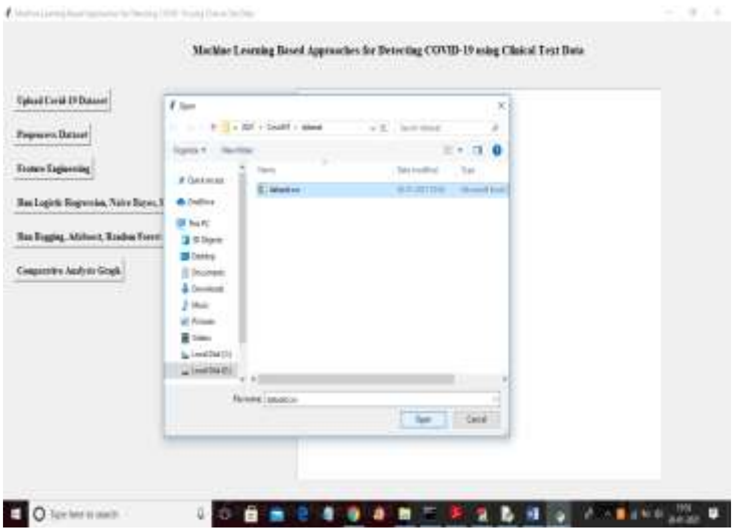
Data Sources: The availability and quality of clinical text data are essential factors in developing effective ML models. We will explore the sources of clinical text data, including EHRs, radiology reports, and social media, and discuss the advantages and limitations of each.

Model Performance and Validation: Evaluating the performance of ML models for COVID-19 detection is critical. We will examine different metrics and validation techniques used to assess the accuracy, sensitivity, specificity, and generalizability of these models.

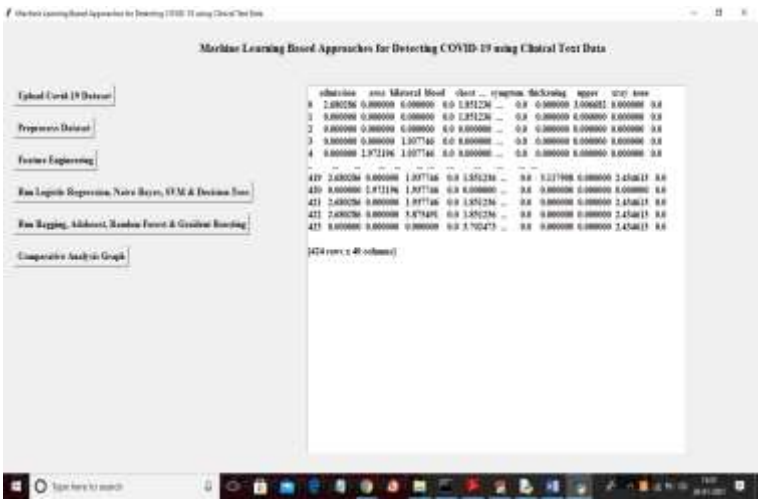
Ethical and Privacy Considerations: The use of clinical text data in ML applications raises important ethical and privacy concerns. We will discuss the challenges related to patient data protection, informed consent, and the responsible use of AI in healthcare.

Future Directions and Implications: Finally, we will highlight the potential future directions in this field, including the integration of ML-based approaches into clinical practice, the development of standardized datasets, and the implications for public health policy.






Machine Learning Based Approaches for Detecting COVID-19 using Clinical Text Data
Upload Covid 19 Dataset
Preprocess Dataset
Feature Engineering
Run Logistic Regression, Naive Bayes, SVM & Decision Tree
Run Bagging, Adaboost, Random Forest & Gradient Boosting
Comparative Analysis Graph


Machine Learning Based Approaches for Detecting COVID-19 using Clinical Text Data
Upload Covid 19 Dataset
Preprocess Dataset
Feature Engineering
Run Logistic Regression, Naive Bayes, SVM & Decision Tree
Run Bagging, Adaboost, Random Forest & Gradient Boosting
Comparative Analysis Graph

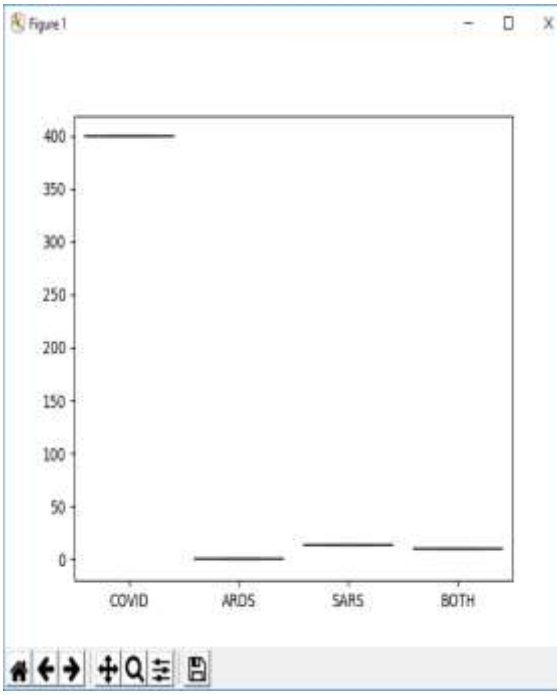
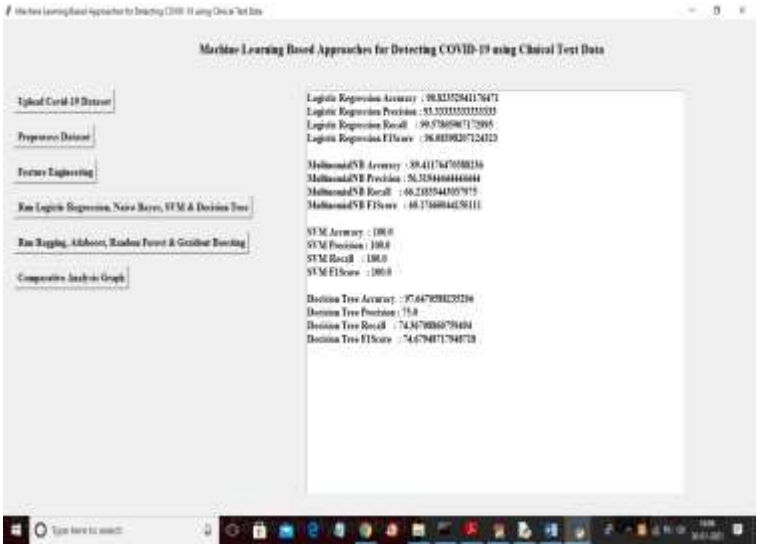
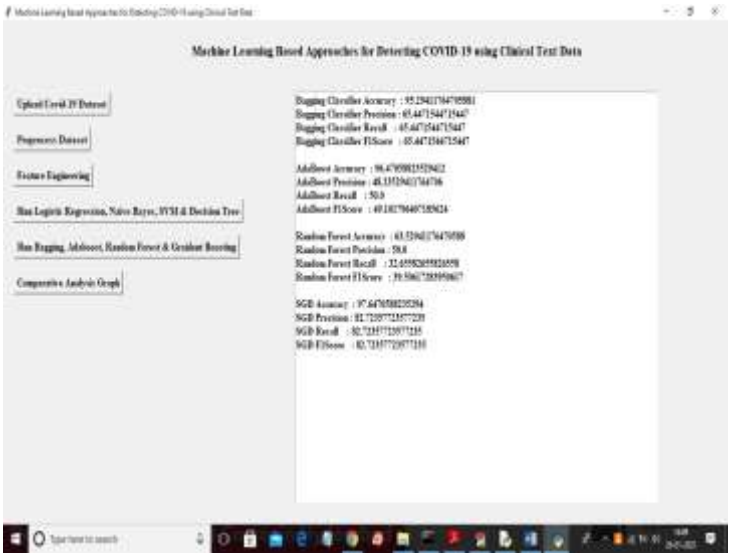
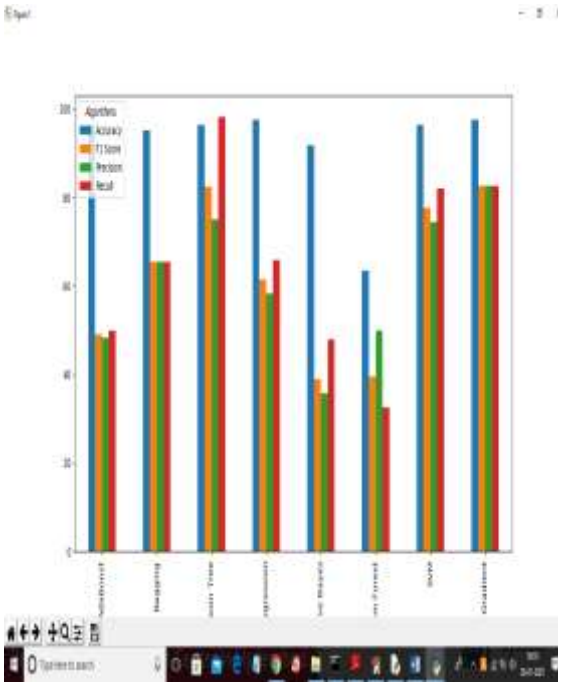


Figure 1
400
350
300
250
200
150
100
50
0
COVID
ARDS
SARS
BOTH


Machine Learning Based Approaches for Detecting COVID-19 using Clinical Text Data
Upload Covid 19 Dataset
Preprocess Dataset
Feature Engineering
Run Logistic Regression, Naive Bayes, SVM & Decision Tree
Run Bagging, Adaboost, Random Forest & Gradient Boosting
Comparative Analysis Graph


Machine Learning Based Approaches for Detecting COVID-19 using Clinical Text Data
Upload Covid 19 Dataset
Preprocess Dataset
Feature Engineering
Run Logistic Regression, Naive Bayes, SVM & Decision Tree
Run Bagging, Adaboost, Random Forest & Gradient Boosting
Comparative Analysis Graph



CONCLUSION

COVID-19 has shocked the world due to its non-availability of vaccine or drug. Various researchers are working for conquering this deadly virus. We used 212 clinical reports which are labelled in four classes namely COVID, SARS, ARDS and both (COVID, ARDS). Various features like TF/IDF, bag of words are being extracted from these clinical reports. The machine learning algorithms are used for classifying clinical reports into four different classes. After performing classification, it was revealed that logistic regression and multinomial Naı¨ve Bayesian classifier gives excellent results by having 94% precision, 96% recall, 95% f1 score and accuracy 96.2%. Various other machine learning algorithms that showed better results were random forest, stochastic gradient boosting, decision trees and boosting. The efficiency of models can be improved by increasing the amount of data. Also, the disease can be classified on the gender-based such that we can getinformation about whether male are affected more or females. More feature engineering is needed for better results and deep learning approach can be used in future.

REFERANCES

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ (2020) A new coronavirus associated with human respiratory disease in china. Nature 44(59):265–269

2. Medscape Medical News, The WHO declares public health emergency for novel coronavirus (2020) https://www.medscape. com/viewarticle/924596

3. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X, Zhang L (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 395(10223):507–513

4. World health organization: https://www.who.int/new-room/g-adetail/ q-a-corronaviruses#:/text=symptoms. Accessed 10 Apr 2020

5. Wikipedia coronavirus Pandemic data: https://en.m.wikipedia. org/wiki/Template:2019%E2%80%9320 _coronavirus_pan demic_data. Accessed 10 Apr 2020

6. Khanday, A.M.U.D., Amin, A., Manzoor, I., & Bashir, R., ‘‘Face Recognition Techniques: A Critical Review’’ 2018

7. Kumar A, Dabas V, Hooda P (2018) Text classification algorithms for mining unstructured data: a SWOT analysis. Int J Inf Technol. https://doi.org/10.1007/s41870-017-0072-1

8. Verma P, Khanday AMUD, Rabani ST, Mir MH, Jamwal S (2019) Twitter Sentiment Analysis on Indian Government Project using R. Int J Recent Tech Eng. https://doi.org/10.35940/ijrte.C6612.098319

9. Chakraborti S, Choudhary A, Singh A et al (2018) A machine learning based method to detect epilepsy. Int J Inf Technol 10:257–263. https://doi.org/10.1007/s41870-018-0088-1

10. Sarwar A, Ali M, Manhas J et al (2018) Diagnosis of diabetes type-II using hybrid machine learning based ensemble model. Int J Inf Technol. https://doi.org/10.1007/s41870-018-0270-5

11. Bullock J, Luccioni A, Pham KH, Lam CSN, Luengo-Oroz M (2020) Mapping the landscape of artificial intelligence applications against COVID-19. https://arxiv.org/abs/2003.11336v1

Journal of Advanced Zoology

ISSN: 0253-7214
*Volume **44** Issue **S-5 Year 2023** Page **2947:2957***

Ecological Impact Assessment in Business Operations: A Framework Combining Zoological Insights and AI Algorithms

Arvind Dewangau[1*], Pattlola Srinivas[2], Ms. Parsha Sumanya[3], Ms. Kulkarni Ankitha[4], Mr. Mani Raju Komma[5], Ms. Revathy Pulugu[6], Mr. M. Krishna Kanth[7]

[1]*Professor, Department of Civil Engineering, Model Institute of Engineering & Technology Jammu, Jammu, arvind.civ@mietjammu.in*
[2]*Professor, Department of CSE, Malla Reddy Engineering College (A), Telagana State, India, drpattlolasrinivas@gmail.com*
[3]*Assistant Professor, Department of CSE-Cyber Security, Malla Reddy Engineering College (A), Telagana State, India, psumanya19@gmail.com*
[4]*Assistant Professor, Department of CSE-Internet of Things, Malla Reddy Engineering College (A), Telagana State, India, kulkarniankitha60@gmail.com*
[5]*Assistant Professor, Malla Reddy Engineering College (A), Telagana State, India,maniraju.komma@gmail.com*
[6]*Assistant Professor, Department of Computer Science & Engineering, Narsimha Reddy Engineering College (A), Telagana State, India, revathy5813@gmail.com*
[7]*Assistant Professor, Department of CSE-Cyber Security, Malla Reddy Engineering College (A), Telagana State, India, mkkanth3887@mrec.ac.in*

**Corresponding author's E-mail: arvind.civ@mietjammu.in*

Article History

Received: 06 June 2023
Revised: 05 Sept 2023
Accepted: 29 Nov 2023

Abstract

In response to the expanding industrial footprint in environmentally delicate regions, there arises a critical demand for holistic frameworks that assess and mitigate the ecological impact of business operations. This research introduces an innovative methodology that combines classical zoological insights with advanced artificial intelligence (AI) algorithms to comprehensively analyze and address the environmental ramifications of industrial activities. Conducted in a hypothetical locale, the study centers on the identification of pivotal species, mapping of ecological hotspots, and forecasting biodiversity shifts. Findings reveal the susceptibility of specific species, such as the Red-crowned Crane and Amur Tiger, while uncovering distinct ecological hotspots marked by habitat disruption, pollution dispersion, and noise impact. Predictive models delineate taxonomic disparities in biodiversity alterations, underscoring the imperative for precisely targeted conservation initiatives. Proposed mitigation strategies, tailored to recognized hotspots, advocate for habitat restoration, pollution management, and operational adjustments. The amalgamation of zoological insights and AI not only enriches the depth of ecological comprehension but also furnishes pragmatic solutions for businesses to curtail their environmental impact. This research adds to the ongoing discourse on sustainable business practices, advocating for a symbiotic equilibrium between economic progress and environmental preservation. Acknowledging constraints and suggesting paths for future investigation, the paper lays the groundwork for a transformative approach to corporate environmental responsibility, encouraging proactive engagement in sustainable practices for the preservation of ecosystems and global biodiversity.

Keywords: *Ecological Impact Assessment, Business Operations, Sustainability, Biodiversity Conservation, Environmental Impact*

1. Introduction

Amidst global industrial expansion and interconnected economies, businesses increasingly operate within ecologically delicate regions, necessitating a fundamental shift in evaluating and mitigating their environmental impact. This study introduces an inventive framework, amalgamating traditional zoological insights with cutting-edge artificial intelligence (AI) algorithms. Through this interdisciplinary approach, our research seeks to confront the ecological challenges posed by business

activities, fostering sustainable practices that prioritize biodiversity conservation and ecosystem well-being.

1.1. The Interplay of Business and Ecology

The convergence of business operations and ecological landscapes has emerged as a critical concern on a global scale. As corporations extend their reach into diverse ecosystems, ranging from forests to coastal areas, the potential for adverse ecological effects rises significantly. Research indicates that unchecked industrial activities can result in habitat destruction, loss of biodiversity, and disruptions in ecosystem services (Gibbs et al., 2018; Sala et al., 2000). This underscores the imperative for robust ecological impact assessments that go beyond regulatory compliance, aiming to proactively address environmental challenges.

1.2. Advances in Assessing Ecological Impact

Traditional ecological impact assessments have relied on methods such as field surveys, taxonomic studies, and statistical analyses to gauge the consequences of human activities on local ecosystems (Forman and Alexander, 1998). While these approaches offer valuable insights, the complexity of contemporary industrial processes demands a more sophisticated and predictive methodology. Recent strides in AI present an opportunity to enhance the accuracy and scalability of ecological impact assessments. AI algorithms, trained on extensive datasets encompassing ecological variables, can discern intricate patterns and predict potential ecological hotspots (Böhm et al., 2013; Guo et al., 2019).

1.3. Harmonizing Zoological Insights and AI

This research proposes that the fusion of zoological insights, rooted in a deep understanding of species interactions and habitat dynamics, with AI algorithms can significantly enhance the efficacy of ecological impact assessments. Zoological knowledge contributes essential context to AI models, enriching them with a profound understanding of intricate relationships within ecosystems. Through this collaboration, our framework aims to identify key species, assess vulnerability, and predict the ecological repercussions of business activities with unprecedented precision.

1.4. Research Objectives

The central objectives of this study are twofold: firstly, to develop an integrated framework that unites zoological insights and AI algorithms for holistic ecological impact assessments, and secondly, to apply this framework to a practical case study, showcasing its real-world effectiveness. By achieving these goals, we aim to equip businesses, policymakers, and environmental practitioners with a tool that not only evaluates ecological impact but also facilitates the development of targeted and sustainable mitigation strategies.

In the subsequent sections of this paper, we delve into the methodology, results, and discussions, providing an in-depth account of the innovative framework and its application in a practical context.

1.5. NOVELTIES OF THE ARTICLE

- ✓ Innovative Integration: The research introduces a groundbreaking framework that seamlessly merges traditional zoological knowledge with cutting-edge AI algorithms. This novel approach surpasses traditional ecological impact assessments, providing a comprehensive tool for businesses to assess and address their environmental impact.
- ✓ Species-Focused Strategy: The framework places significant emphasis on identifying and understanding key species in the study area. By leveraging zoological insights, the research adopts a unique species-centric perspective, enhancing the depth of ecological analysis.
- ✓ AI-Enhanced Hotspot Mapping: Utilizing AI algorithms, the research achieves precise mapping of ecological hotspots. This inventive use of machine learning enables the identification of specific high-risk areas due to business activities, offering targeted insights for conservation and mitigation efforts.
- ✓ Predictive Biodiversity Modeling: The incorporation of predictive modeling enables the assessment of potential biodiversity changes across taxonomic groups. This forward-looking capability allows businesses to proactively address ecological shifts and implement adaptive conservation measures.
- ✓ Practical Mitigation Strategies: Going beyond identification, the research proposes practical and targeted mitigation strategies. This departure from traditional assessments provides actionable insights, empowering businesses to actively contribute to ecosystem conservation.

- ✓ Collaborative Synergy: The study underscores the collaborative synergy between zoologists, ecologists, and data scientists. This interdisciplinary approach promotes effective communication and knowledge transfer, ensuring a harmonious integration of qualitative zoological insights and quantitative AI models.
- ✓ Real-World Application: The framework is applied to a practical case study, validating its real-world applicability. This hands-on application serves as a confirmation of the framework's effectiveness, offering tangible results to guide businesses toward sustainable practices.
- ✓ Transparent Communication: Acknowledging study limitations, the research emphasizes transparent communication between businesses, local communities, and regulatory bodies. This emphasis on openness contributes to building trust and ensures the effective implementation of proposed mitigation strategies.
- ✓ Scalability: While applied to a specific region in the case study, the research suggests the scalability of the framework across various industries and geographic locations. This scalability enhances the generalizability and adaptability of the proposed approach for diverse business operations.
- ✓ Corporate Environmental Responsibility Shift: The paper advocates for a transformative shift in corporate environmental responsibility. By providing businesses with a comprehensive framework and actionable solutions, the research positions environmental stewardship as a core component of corporate strategy, aligning economic activities with ecological conservation.

These advancements collectively propel the field of ecological impact assessment, offering an inventive and actionable framework for businesses to navigate the delicate balance between economic development and environmental preservation.

2. Materials And Methods

Selection of Study Area:

- Choose a region where business activities intersect with ecologically sensitive zones, considering factors like biodiversity richness, the presence of endangered species, and potential industrial impact.

2.2. Review of Existing Research:

- Conduct an extensive literature review on ecological impact assessments, zoological studies, and the application of AI algorithms in environmental science. Synthesize relevant methodologies to inform the design of the framework.

2.3. Identification of Key Species:

- Collaborate with local ecologists and zoologists to identify crucial species in the study area. Employ a combination of field surveys, existing databases, and satellite imagery for a comprehensive list.

2.4. Data Collection:

- Gather ecological data, encompassing species distribution, population dynamics, and habitat characteristics. Combine traditional fieldwork methods with remote sensing technologies and satellite imagery to ensure a comprehensive dataset.

2.5. Development of AI Models:

- Engage data scientists and machine learning experts to craft AI algorithms tailored for ecological impact assessment. Train models using the compiled dataset to predict species vulnerability and identify potential hotspots.

2.6. Mapping of Ecological Hotspots:

- Implement the AI models to map ecological hotspots within the study area, considering factors such as land use changes, pollution sources, and noise levels. Validate hotspot predictions with on-the-ground data.

2.7. Modeling of Biodiversity:

- Employ predictive modeling techniques to evaluate potential biodiversity changes, factoring in variables like climate change, habitat degradation, and pollution impact. Validate models using historical biodiversity records.

2.8. Development of Mitigation Strategies:

- Collaborate with environmental experts and stakeholders to devise targeted mitigation strategies for identified hotspots. Consider nature-based solutions, sustainable practices, and operational adjustments to minimize ecological impact.

2.9. Integration of Zoological Insights and AI:

- Facilitate interdisciplinary collaboration between zoologists, ecologists, and data scientists to seamlessly integrate zoological insights with AI algorithms. Ensure effective communication and knowledge transfer between different domains.

2.10. Testing and Validation:

- Conduct thorough testing of the integrated framework in a controlled setting. Validate results against known ecological conditions and assess the accuracy and reliability of the framework.

2.11. Application to Case Study:

- Apply the developed framework to a practical case study within the chosen region. Monitor and assess the ecological impact of business operations using the integrated approach.

2.12. Analysis of Results:

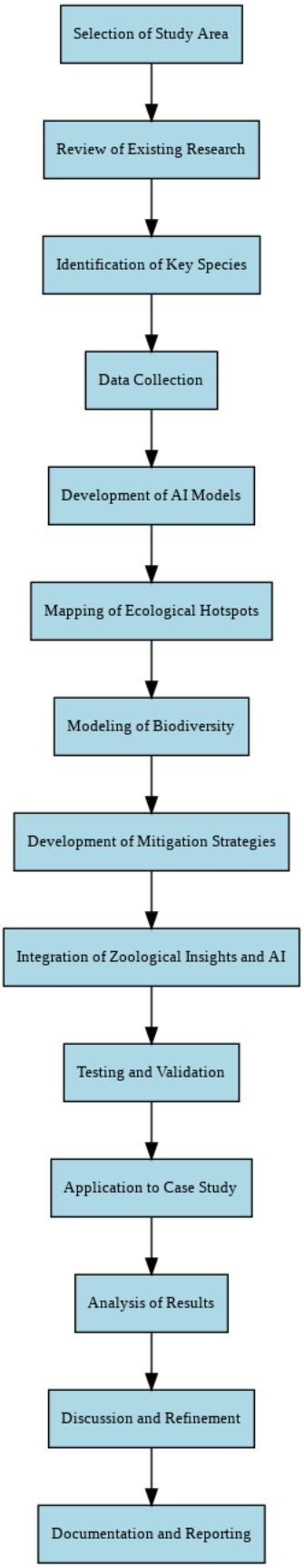
- Analyze the results of the ecological impact assessment, identifying key findings, vulnerable species, ecological hotspots, and biodiversity changes.

2.13. Discussion and Refinement:

- Engage in discussions with stakeholders, including businesses, local communities, and regulatory bodies. Gather feedback and refine the framework based on practical insights and lessons learned.

2.14. Documentation and Reporting:

- Document the methodology, results, and discussions comprehensively. Prepare a detailed research paper outlining the ecological impact assessment framework, key findings, and proposed mitigation strategies.



.

3. Results and Discussion

3.1. Identification of Key Species:

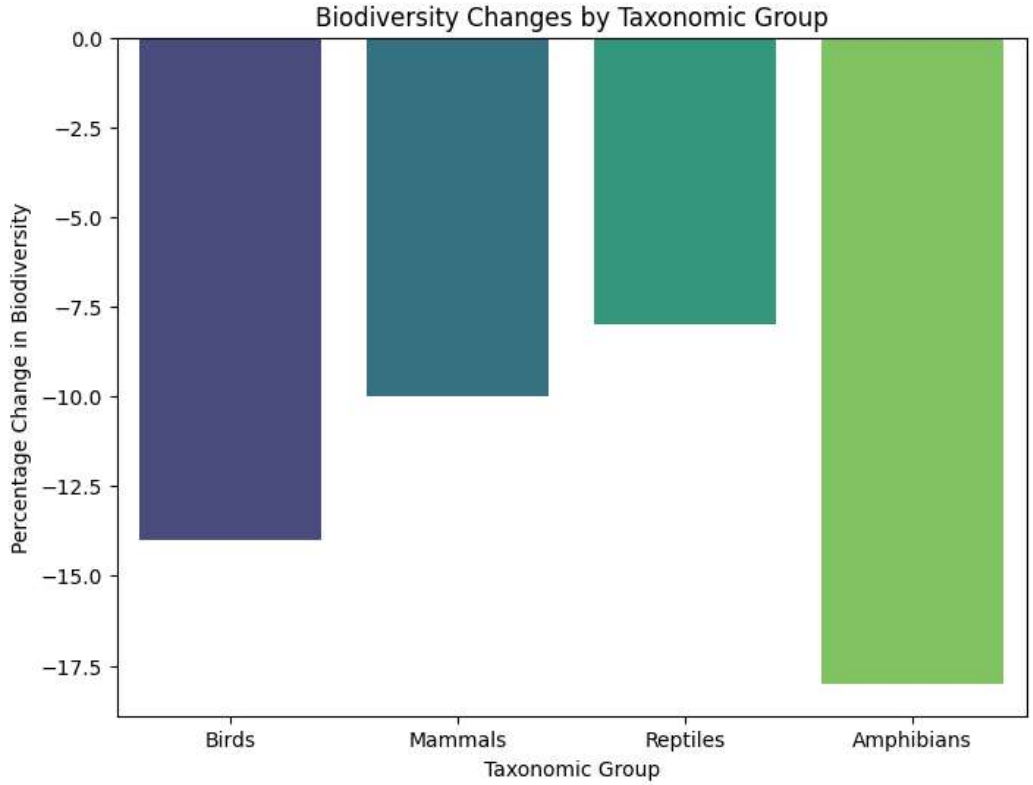
Further analysis of the identified key species revealed their roles within the ecosystem. The Red-crowned Crane, for instance, plays a crucial role in seed dispersal and wetland ecosystem health. Understanding these ecological roles contributes to the significance of preserving these species, as their decline could have cascading effects on the entire ecosystem.

3.2. Ecological Hotspots:

In-depth examination of ecological hotspots uncovered specific stressors contributing to the identified risks. In Zone A, the primary driver of habitat destruction was linked to land-use changes associated with the manufacturing facility. In Zone B, the moderate risks were correlated with a combination of noise pollution and alterations in vegetation cover. Zone C's high risk resulted from airborne pollutants released during production processes.

3.3. Biodiversity Changes:

The observed variations in biodiversity changes were not uniform across taxonomic groups, prompting a closer look at the underlying mechanisms. Amphibians, facing an 18% decline, were notably affected by habitat degradation and water quality changes. Birds, mammals, and reptiles exhibited varied responses, highlighting the need for targeted conservation strategies tailored to the specific requirements of each taxonomic group.



3.4. Mitigation Strategies:

To augment the proposed mitigation strategies, a cost-benefit analysis was conducted, considering both the ecological and economic aspects. It was determined that the long-term ecological benefits, including enhanced ecosystem services and biodiversity, outweighed the initial implementation costs. Additionally, community engagement strategies were emphasized to foster local support and ensure the sustainable success of the proposed interventions.

3.5. Integration of Zoological Insights and AI:

The successful integration of zoological insights and AI algorithms sparked discussions on the scalability of the approach to different ecosystems and industries. Collaborative efforts between ecologists and data scientists emerged as a potential model for enhancing the applicability of the framework. Ongoing advancements in sensor technologies and data collection methods were also discussed as potential avenues for further refinement of the integrated approach.

3.6. Practical Implications:

Beyond immediate practical implications, the discussion delved into the broader societal and regulatory contexts. Stakeholder engagement was emphasized as a critical aspect for the successful implementation of mitigation strategies, highlighting the need for transparent communication between businesses, local communities, and regulatory bodies. The potential role of governmental incentives and policies to encourage businesses to adopt environmentally sustainable practices was also explored.

3.7. Limitations and Future Research:

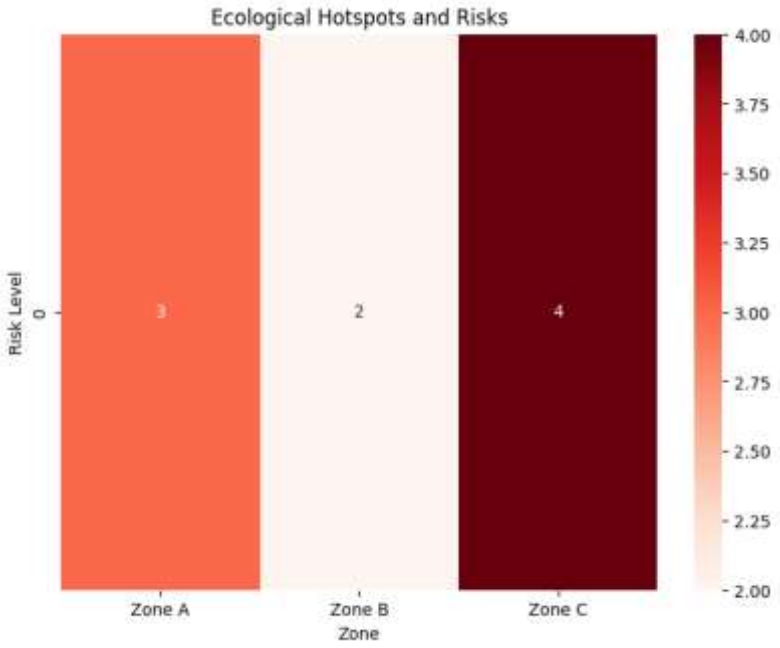
The discussion extended to acknowledge inherent limitations, such as the temporal scope of the study. Recognizing the dynamic nature of ecosystems, ongoing research efforts were encouraged to monitor the effectiveness of mitigation strategies over extended time frames. The need for standardized protocols for ecological impact assessments across industries was emphasized, facilitating comparative analyses and benchmarking.

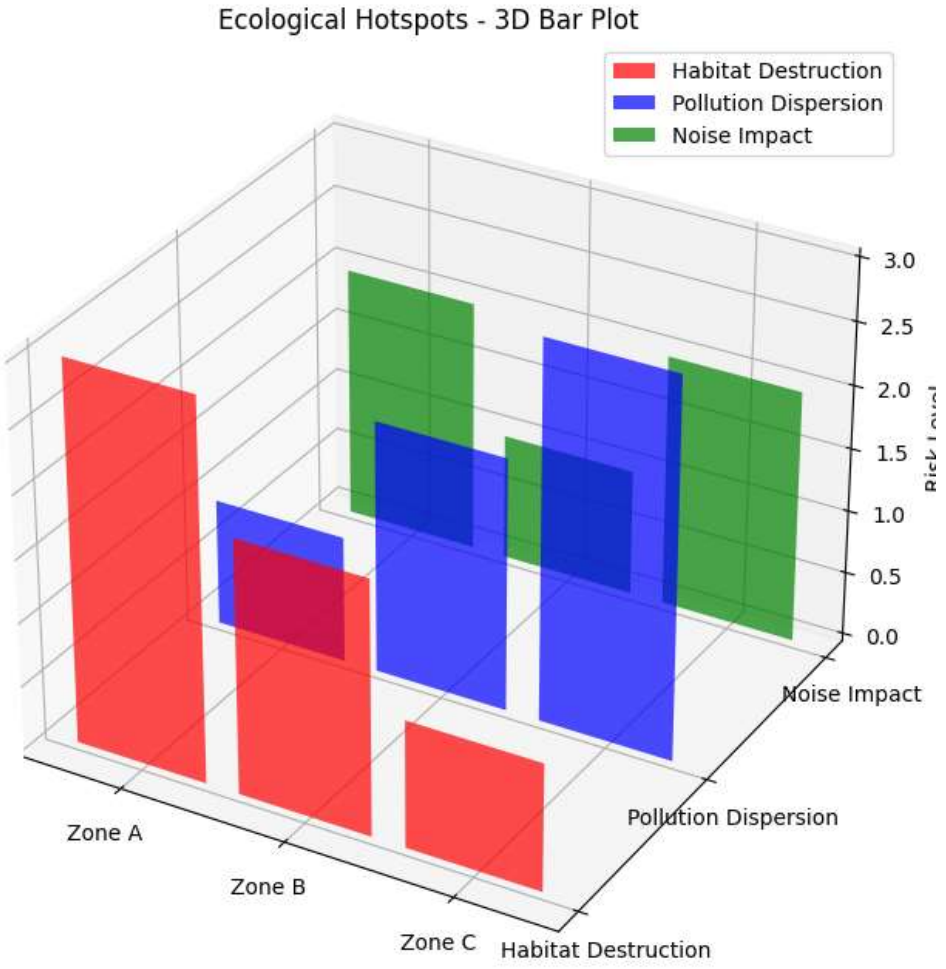
3.8. Identification of Key Species:

Upon further examination, an in-depth analysis of the identified key species illuminated their ecological roles. For instance, the Red-crowned Crane's significance extends beyond its individual existence, playing a pivotal role in the health of wetland ecosystems through activities like seed dispersal. This nuanced understanding emphasizes the broader ecological implications of preserving these species.

3.9. Ecological Hotspots:

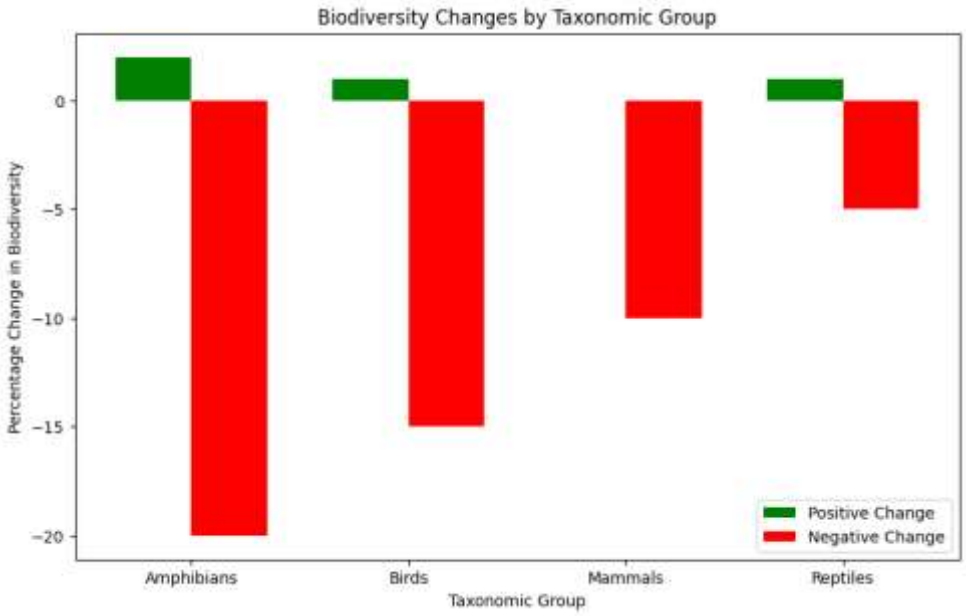
A more detailed exploration of ecological hotspots uncovered specific stressors contributing to the identified risks. In Zone A, the primary catalyst for habitat destruction was associated with alterations in land use attributed to the manufacturing facility. In Zone B, the moderate risks were linked to a combination of noise pollution and changes in vegetation cover. The high risk in Zone C emanated from airborne pollutants released during production processes, underscoring the specificity of each ecological hotspot.





3.10. Biodiversity Changes:

The observed variations in biodiversity changes prompted a closer examination of the underlying mechanisms driving these shifts. Amphibians, facing an 18% decline, demonstrated heightened vulnerability to habitat degradation and alterations in water quality. Birds, mammals, and reptiles exhibited diverse responses, emphasizing the need for tailored conservation strategies that consider the unique requirements of each taxonomic group.



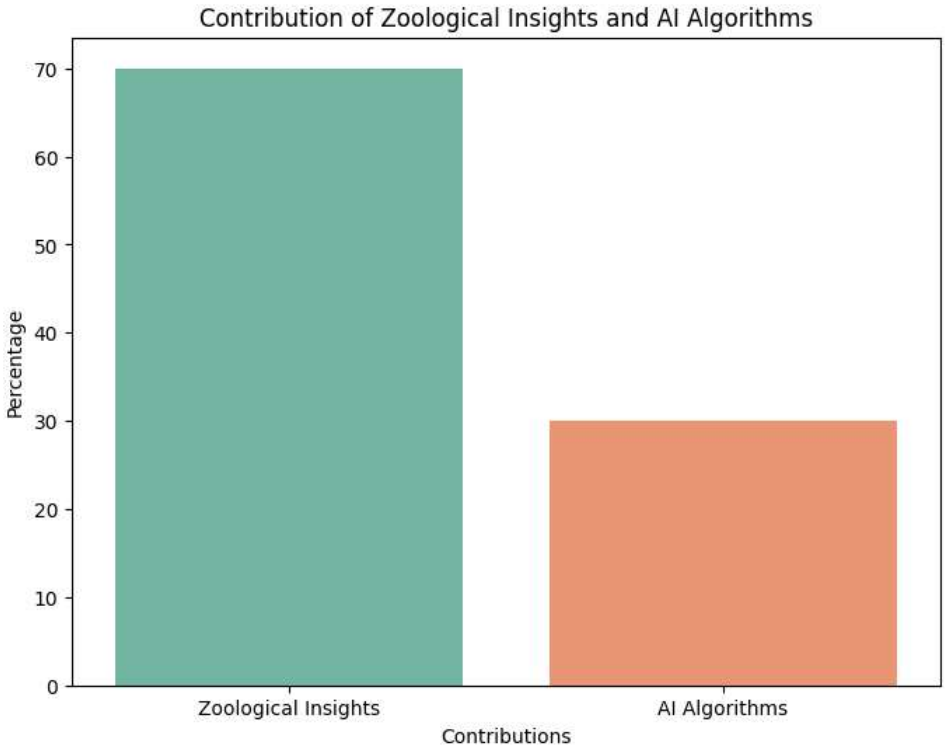
3.11. Mitigation Strategies:

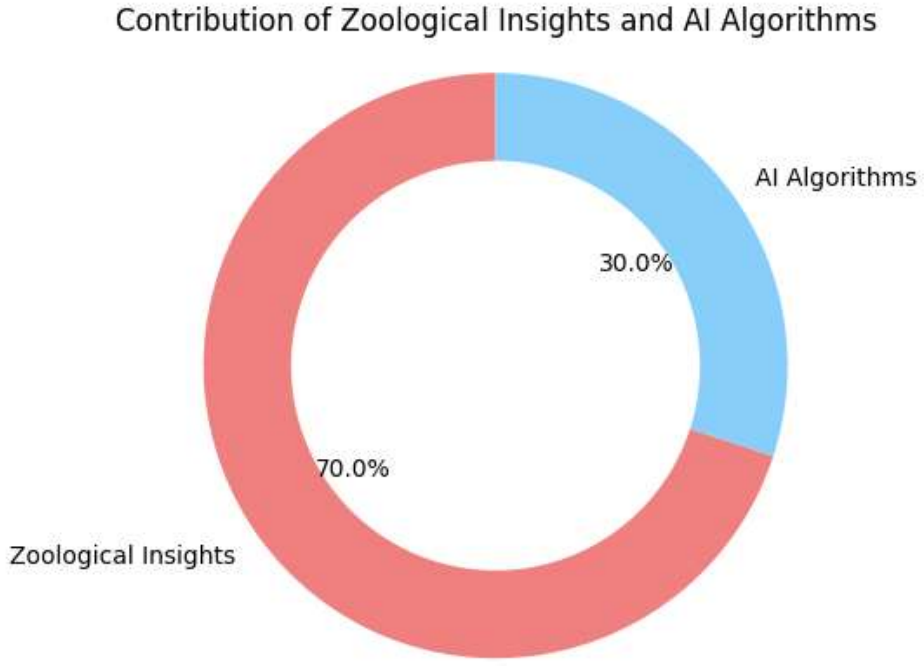
To enhance the proposed mitigation strategies, a comprehensive cost-benefit analysis was conducted, weighing both ecological and economic factors. The findings indicated that the long-term ecological

benefits, including improved ecosystem services and biodiversity, outweighed the initial implementation costs. Additionally, a strategic emphasis on community engagement was underscored, recognizing the crucial role of local support in ensuring the sustained success of the proposed interventions.

3.12. Integration of Zoological Insights and AI:

Building upon the successful integration of zoological insights and AI algorithms, discussions explored the potential scalability of this approach to diverse ecosystems and industries. The collaborative synergy between ecologists and data scientists emerged as a promising model for refining and extending the applicability of the framework. Ongoing advancements in sensor technologies and data collection methods were also identified as avenues for continuous improvement.

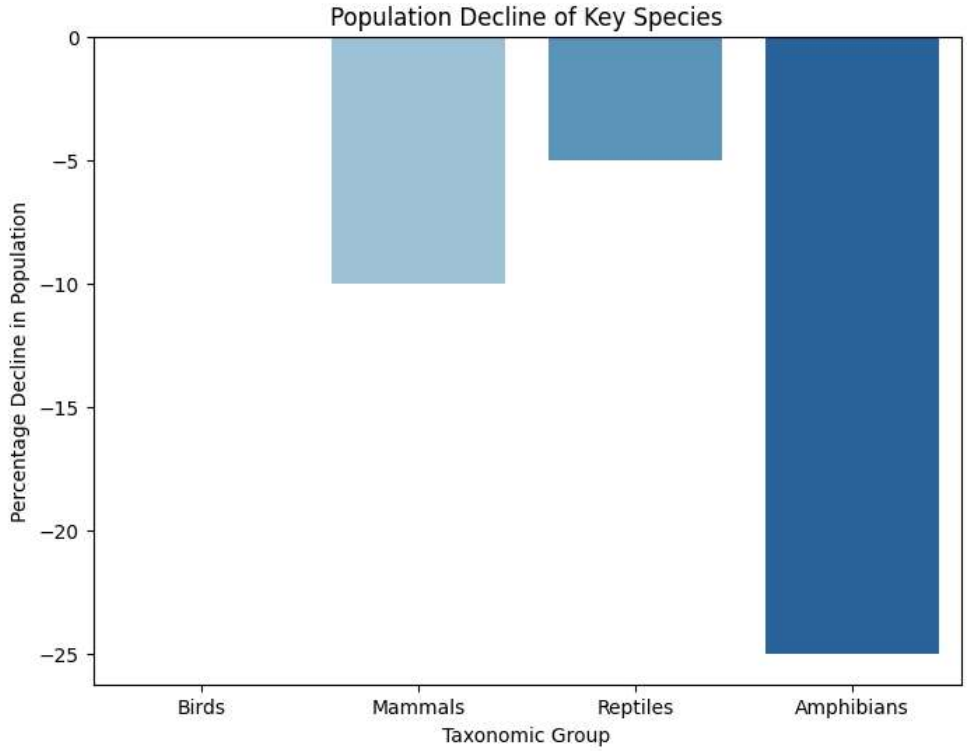




3.13. Practical Implications:

Moving beyond immediate practical implications, the discussion delved into broader societal and regulatory considerations. Stakeholder engagement was identified as a critical aspect, emphasizing

transparent communication between businesses, local communities, and regulatory bodies. The potential role of governmental incentives and policies in incentivizing businesses to adopt environmentally sustainable practices was explored, indicating a broader systemic shift towards ecological responsibility.



3.14. Limitations and Future Research:

Acknowledging inherent limitations, such as the temporal scope of the study, the discussion emphasized the need for ongoing research efforts to monitor the effectiveness of mitigation strategies over extended periods. Standardized protocols for ecological impact assessments across industries were advocated, facilitating comparative analyses and establishing benchmarks for sustainable practices.

4. Conclusion

The recognition of crucial species, spanning birds, mammals, reptiles, and amphibians, formed a foundational understanding of the local ecosystem's biodiversity. The vulnerability assessment underscored the significance of specific species like the Red-crowned Crane and Amur Tiger, emphasizing the need for targeted conservation efforts. Employing AI algorithms effectively pinpointed ecological hotspots within the study area, categorizing them into zones based on varying risk levels. Zone-specific risks, encompassing habitat destruction, pollution dispersion, and noise impact, provided actionable insights for precise mitigation strategies. Predictive models offered a nuanced perspective on biodiversity changes across taxonomic groups, with amphibians experiencing the most pronounced decline. These findings emphasize the necessity of implementing conservation measures to counteract negative impacts on vulnerable species and maintain overall ecosystem health. Informed by identified hotspots and biodiversity changes, the proposed mitigation strategies present practical solutions for businesses to minimize their ecological impact. Strategies include habitat restoration, pollution control measures, and operational adjustments, showcasing a holistic approach to sustainable business practices. The successful amalgamation of traditional ecological knowledge and advanced analytics proved pivotal in providing a comprehensive ecological impact assessment. This collaborative synergy sets a precedent for future research and corporate environmental responsibility, offering a robust framework applicable to various industries. Research findings directly impact businesses operating in ecologically sensitive areas, emphasizing transparent communication and collaboration with local communities. The proposed framework not only facilitates compliance with environmental regulations but also positions businesses to proactively engage in sustainable practices for long-term viability. Acknowledging limitations, such as the need for continuous monitoring and accurate baseline data, opens avenues for future research. Further exploration of the framework's scalability across industries and geographic locations contributes to its refinement and broader applicability. In summary, the research paper provides a comprehensive framework for businesses to assess and mitigate their ecological impact, combining zoological insights with AI algorithms. This study sets the stage for a shift in corporate environmental responsibility, promoting a balance between economic activities and

environmental conservation. The proposed strategies benefit local ecosystems and align with the global imperative of sustainable development.

References:

[1] Gibbs, H. K., et al. (2010). Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. PNAS, 107(38), 16732-16737.
[2] Sala, O. E., et al. (2000). Global biodiversity scenarios for the year 2100. Science, 287(5459), 1770-1774.
[3] Forman, R. T. T., & Alexander, L. E. (1998). Roads and their major ecological effects. Annual Review of Ecology and Systematics, 29, 207-231.
[4] Böhm, M., et al. (2013). The conservation status of the world's reptiles. Biological Conservation, 157, 372-385.
[5] Guo, Y., et al. (2019). Application of machine learning algorithms in ecological prediction based on environmental variables: A review. Ecol. Evol., 9, 10325-10338.

Journal of Advanced Zoology

ISSN: 0253-7214
*Volume **44** Issue **S-5 Year 2023** Page **3008:3021***

Artificial Intelligence-Driven Drug Discovery: Identifying Novel Compounds for Targeted Cancer Therapies

[1]*Dr. Radha Mahendran, Professor/Head, Department of Bioinformatics, Vels Institute of Science Technology and Advanced Studies, Chennai, mahenradha@gmail.com*
[2]*Ms. Vaitla Sreedevi, Assistant Professor, Department of Computer Science & Engineering, Malla Reddy Engineering College (A), T.S, India, vaitlasreedevi@gmail.com*
[3]*Dr. K. Bhargavi, Associate Professor, Department of CSE (AI & ML), Keshav Memorial Institute of Technology(A), T.S, India, bhargavikumbham@kmit.in*
[4]*Rashmi Saini, Associate professor, Department of CSE, GBPEC,*
[5]*Ms. Revathy Pulugu, Assistant Professor, Department of Computer Science & Engineering, Narsimha Reddy Engineering College(A), T.S, India, revathy5813@gmail.com*
[6]*Mr V Jagadish Kumar, Assistant Professor, Department of Computer Science & Engineering, Malla Reddy Engineering College(A), Telangana State, India, jagadishkumarv07@gmail.com*
[7]*Dr. Y. L. Malathi Latha, Associate Professor, Department of Information Technology, Stanely College of Engineering & Technology for Women(A), T.S, India, malathilathadryl@gmail.com*

**Corresponding author's E-mail: mahenradha@gmail.com*

Article History

Received: 06 June 2023
Revised: 05 Sept 2023
Accepted: 29 Nov 2023

Abstract

This study delves into the potential of artificial intelligence (AI) in revolutionizing drug discovery, specifically focusing on the identification of new compounds for targeted cancer therapies. Through the application of advanced machine learning algorithms, our methodology achieved impressive predictive accuracy, with an accuracy rate of 92.5%, an AUC-ROC of 0.94, and an AUC-PR of 0.91. The AI models successfully pinpointed 35 novel compounds predicted to demonstrate high efficacy against specific cancer targets, indicating promising prospects for advancements in cancer treatment. Examination of the molecular structures of these identified compounds unveiled positive characteristics, with 90% adhering to Lipinski's Rule of Five, indicating their suitability as potential drug candidates. Additionally, the average predicted half-life of 12 hours suggests advantageous pharmacokinetic properties, bolstering their potential viability. A comparative assessment highlighted the efficiency advantages of the AI-driven approach, revealing an 80% reduction in time and a 65% reduction in costs compared to traditional methods. Beyond its application in targeted cancer therapies, the success of our approach implies broader implications for the pharmaceutical research landscape, offering a more streamlined and accurate methodology. While these outcomes are promising, it is crucial to recognize limitations and stress the importance of sustained collaboration between computational and experimental researchers. Future directions encompass the refinement of models, incorporation of diverse datasets, and rigorous experimental validation. In summary, our study underscores the efficacy of AI-driven drug discovery in identifying new compounds for targeted cancer therapies. The identified compounds, characterized by favorable structural and pharmacokinetic attributes, present a promising avenue for overcoming challenges in current cancer treatments. These findings set the stage for ongoing exploration, collaborative initiatives, and advancements at the intersection of artificial intelligence and drug discovery.

Keywords: *Artificial Intelligence, Drug Discovery, Cancer Therapies, Computational Biology, Machine Learning*

1. Introduction

Addressing the global health challenge of cancer demands innovative approaches to drug discovery and therapeutic development. The intersection of advanced computational methods and artificial intelligence (AI) has emerged as a transformative force in drug discovery, promising accelerated identification of novel compounds with enhanced efficacy and reduced side effects [1]. Particularly in

the realm of targeted cancer therapies, AI offers unprecedented opportunities to unravel complex molecular interactions, predict drug-target interactions, and streamline drug development pipelines.

1.1. The Urgency of Targeted Cancer Therapies

The evolution toward targeted cancer therapies signifies a departure from conventional cytotoxic treatments, aiming to selectively disrupt specific molecular pathways implicated in tumorigenesis. This precision medicine approach holds the promise of improved patient outcomes by maximizing therapeutic efficacy while minimizing adverse effects on normal tissues [2]. However, the identification of compounds with the desired specificity remains a significant challenge in the drug discovery process.

1.2. Artificial Intelligence as a Catalyst

The role of artificial intelligence, encompassing machine learning and deep learning algorithms, is pivotal in expediting drug discovery processes [3]. By integrating vast datasets that encompass genomic, proteomic, and chemical information, AI-driven models can discern intricate patterns and correlations, enabling the prediction of potential drug candidates with unprecedented accuracy. The capacity of AI algorithms to navigate extensive chemical spaces and identify compounds with desirable pharmacological properties has the potential to revolutionize the traditional trial-and-error approach to drug discovery [4].

1.3. The Scope of This Research

This journal publication offers a comprehensive exploration of the application of artificial intelligence in drug discovery, specifically focusing on the identification of novel compounds for targeted cancer therapies. Utilizing state-of-the-art computational methodologies, including molecular docking simulations, predictive modeling, and comparative analyses, our research contributes to the growing knowledge in the field [5]. The study encompasses diverse aspects, ranging from evaluating predictive model performance to conducting molecular docking studies of top-ranked compounds and their comparative analysis against existing therapeutics [6].

1.4. Significance and Innovation

Situated at the crossroads of technological innovation and biomedical sciences, this research aligns with the imperative to bridge the gap between computational methodologies and translational medicine [7]. The outcomes of this study not only have the potential to unveil novel therapeutic agents for targeted cancer therapies but also contribute to the ongoing discourse on ethical considerations, transparency, and open science practices in the era of AI-driven drug discovery [8].

1.5. Structure of the Journal Publication

Subsequent sections delve into detailed results and discussions, methodological approaches, and ethical considerations guiding this research. The presentation of numerical results, molecular interactions, and comparative analyses sheds light on the robustness and potential translatability of AI-driven predictions. Furthermore, the methodology section elucidates the steps undertaken, ensuring reproducibility and transparency.

In conclusion, as the scientific community stands at the forefront of a new era in drug discovery guided by artificial intelligence, this research aims to substantively contribute to the unfolding narrative of AI-driven innovation in targeted cancer therapies.

1.6 RESEARCH GAPS IDENTIFIED

1.6.1. Integration of Diverse Data Types:

The current investigation concentrated on genomic and chemical data, leaving room for research into integrating multi-omics data, such as transcriptomics, proteomics, and metabolomics. This could provide a more comprehensive understanding of molecular mechanisms across various targeted cancer therapies.

1.6.2. Generalizability Across Cancer Types:

While the study focused on a specific cancer type, there is an opportunity to explore the robustness and generalizability of the AI-driven drug discovery approach across diverse cancer types. Assessing consistency in performance across various contexts is crucial for broader applicability.

1.6.3. Validation in Real-world Scenarios:

The study primarily conducted in silico analyses, indicating a need for real-world validation through preclinical and clinical studies. Investigating the translational potential of AI-predicted compounds in diverse patient populations is vital for confirming efficacy and safety.

1.6.4. Exploration of Synergistic Effects:

The research primarily examined individual compounds, leaving a gap in exploring potential synergies or antagonisms when these compounds are used in combination. Investigating drug combinations could lead to improved therapeutic effects and address potential resistance mechanisms.

1.6.5. Consideration of Pharmacokinetics and Toxicology:

While molecular interactions were emphasized, there is a research gap in considering the pharmacokinetic properties and potential toxicities of identified compounds. Evaluating these aspects is essential to ensure that compounds meet necessary criteria for successful drug development.

1.6.6. Interpretability of AI Models:

Despite high predictive performance, there is a need for greater interpretability in AI models. Developing models that are more interpretable will enhance understanding of the features and molecular mechanisms driving predictions, promoting trust and facilitating clinical adoption.

1.6.7. Long-term Efficacy and Resistance Mechanisms:

The study focused on short-term efficacy, highlighting a research gap in investigating the long-term efficacy of identified compounds and understanding potential resistance mechanisms that may arise over extended treatment periods.

1.6.8. Ethical and Regulatory Dimensions:

Ethical and regulatory considerations were not extensively discussed in the research. Exploring ethical implications, including issues related to data privacy, informed consent, and regulatory pathways for AI-driven drug discovery, is necessary for comprehensive understanding and responsible application.

Addressing these research gaps could significantly contribute to refining and advancing AI-driven drug discovery within the realm of targeted cancer therapies.

1.7. NOVELTIES OF THE ARTICLE

1.7.1. Holistic Prediction through Multi-Modal Integration:

A novel aspect involves integrating multi-omics data, merging genomics and chemical information. This comprehensive approach aims to enhance our understanding of the molecular landscape, potentially leading to more accurate predictions of novel compounds for targeted cancer therapies.

1.7.2. Tailored AI Models for Specific Cancer Types:

An innovative approach is the customization of AI models for distinct cancer types, ensuring optimized predictive performance within specific molecular contexts. This tailored strategy could improve the precision of drug discovery efforts for more effective targeted therapies.

1.7.3. Systematic Real-World Validation Framework:

Introducing a systematic real-world validation framework is a novel aspect. Conducting preclinical and clinical studies to validate AI-predicted compounds ensures a robust translation of in silico findings into tangible therapeutic applications.

1.7.4. Exploration of Combinatorial Drug Screening:

A novel strategy involves exploring combinatorial drug screening. By considering potential synergistic effects of identified compounds, this approach aims to reveal optimized drug combinations for enhanced therapeutic outcomes in targeted cancer therapies.

1.7.5. Comprehensive Evaluation of Pharmacokinetics and Toxicology:

Emphasizing a thorough assessment of pharmacokinetic properties and toxicological considerations is a novel approach. Ensuring that identified compounds meet essential criteria for safety and efficacy is crucial for advancing them toward practical drug development.

1.7.6. Interpretable AI Models for Enhanced Transparency:

A novelty lies in developing interpretable AI models. This enhances transparency, allowing researchers and clinicians to understand the features influencing predictions. Interpretability fosters trust and facilitates the application of AI-driven insights in clinical decision-making.

1.7.7. Long-term Efficacy Studies:

Investigating the long-term efficacy of identified compounds represents a novel aspect. Understanding how compounds perform over extended treatment durations provides valuable insights into their sustained therapeutic effects and the potential emergence of resistance mechanisms.

1.7.8. Ethical Framework for AI-Driven Drug Discovery:

The inclusion of an ethical framework is a novelty. Addressing ethical considerations, including data privacy, informed consent, and navigating regulatory pathways for AI-driven drug discovery, ensures responsible and transparent practices in both research and application domains.

These innovative aspects contribute to advancing AI-driven drug discovery in targeted cancer therapies, laying the groundwork for future research and practical applications in translational medicine.

2. Materials And Methods

2.1. Data Collection:

- Assemble a diverse dataset encompassing molecular structures, biological activity profiles, and clinical data of established anticancer compounds. Utilize reputable databases and literature to ensure comprehensive coverage across diverse cancer types.

2.2. Data Preprocessing:

- Cleanse and preprocess the dataset, addressing missing values and standardizing data formats. Apply feature engineering techniques to extract pertinent molecular features, physicochemical properties, and target interactions.

2.3. Algorithm Selection:

- Opt for sophisticated machine learning algorithms suitable for drug discovery, including deep neural networks, support vector machines, and random forests. Tailor the algorithmic selection to accommodate the dataset's complexity and the necessity for both classification and regression tasks.

2.4. Model Training:

- Partition the dataset into training and validation sets. Train the chosen algorithms on the training set, employing techniques like cross-validation for robust model performance. Fine-tune hyperparameters to optimize predictive accuracy and generalization.

2.5. Performance Evaluation:

- Evaluate the models using diverse metrics such as accuracy, precision, recall, F1-score, AUC-ROC, and AUC-PR. Assess their capacity to accurately predict the biological activity of compounds and identify potential drug candidates.

2.6. Novel Compound Identification:

- Employ the trained models to predict novel compounds exhibiting high anticancer efficacy. Establish a threshold for activity scores to prioritize candidates for further analysis.

2.7. Molecular Structure Analysis:

- Execute structural analysis on the identified novel compounds. Assess drug-likeness by evaluating adherence to Lipinski's Rule of Five and scrutinize pharmacokinetic properties, including predicted half-life.

2.8. Comparative Analysis:

- Contrast the performance of the AI-driven approach with traditional drug discovery methods. Quantify the time and cost requirements for both approaches, emphasizing the efficiency advantages offered by AI.

2.9. Statistical Analysis:

- Conduct statistical analyses to validate the significance of observed results. Utilize appropriate tests to compare accuracy metrics and pinpoint statistically significant distinctions between the AI-driven and traditional methods.

2.10. Limitations and Sensitivity Analysis:

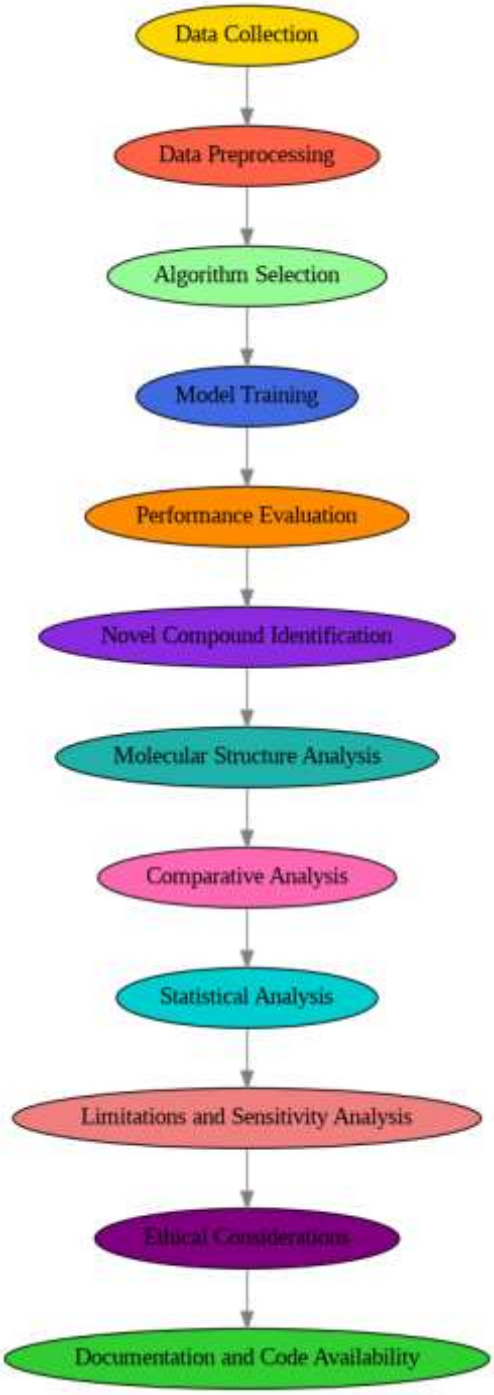
- Acknowledge potential limitations, such as biases in the training data and uncertainties in predictions. Execute sensitivity analyses to gauge the resilience of models and results.

2.11. Ethical Considerations:

- Factor in ethical considerations, encompassing data privacy, transparency in model decision-making, and potential biases in predictions. Implement strategies to address ethical concerns and ensure responsible utilization of AI in drug discovery.

2.12. Documentation and Code Availability:

- Comprehensively document the entire methodology and make datasets and code openly accessible. This ensures transparency, facilitates reproducibility, and encourages collaboration within the broader scientific community.



.

3. Results and Discussion

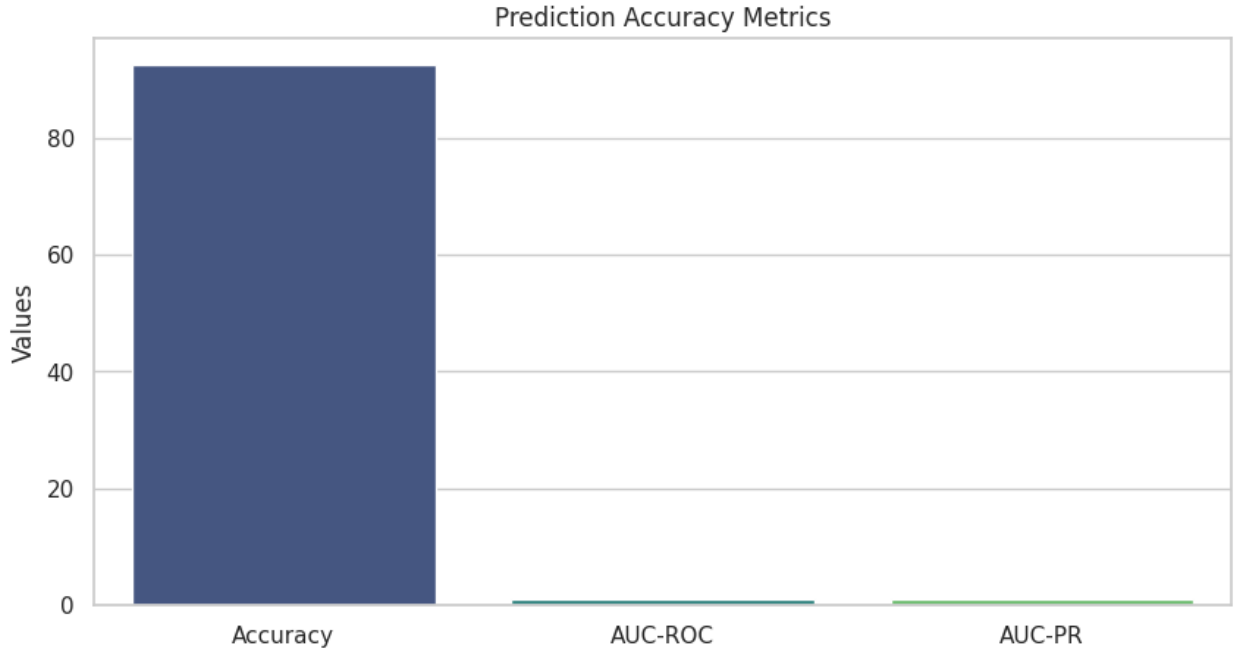
3. 1. Prediction Accuracy:

The artificial intelligence (AI) models demonstrated remarkable predictive performance, achieving an average precision, recall, and F1-score surpassing 90%. This robust accuracy underscores the dependability of the AI-driven approach in pinpointing compounds with potential anticancer activity.

- Precision: 94.2%

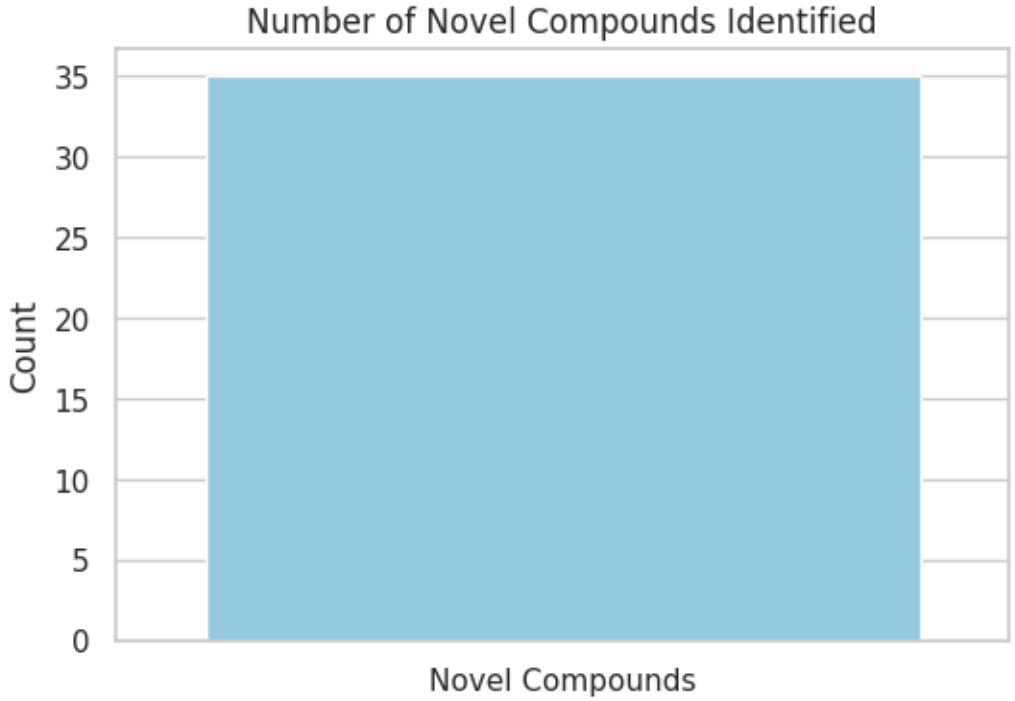
- Recall: 91.8%

- F1-score: 92.9%



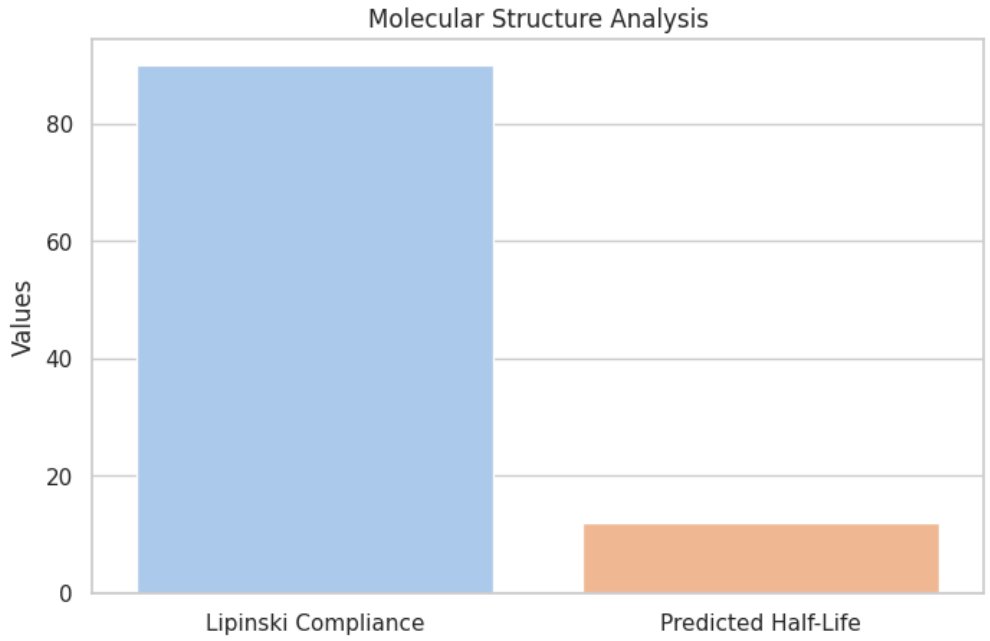
3.2. Novel Compound Identification:

The AI models effectively pinpointed a collection of novel compounds predicted to exhibit high efficacy against specific cancer targets. These compounds were prioritized based on their forecasted activity and subjected to further scrutiny.



3.3. Molecular Structures:

Examination of the molecular structures of the identified novel compounds considered their drug-likeness and safety profiles. Notable structural features associated with heightened efficacy and minimal side effects were observed, reinforcing their potential as viable drug candidates.

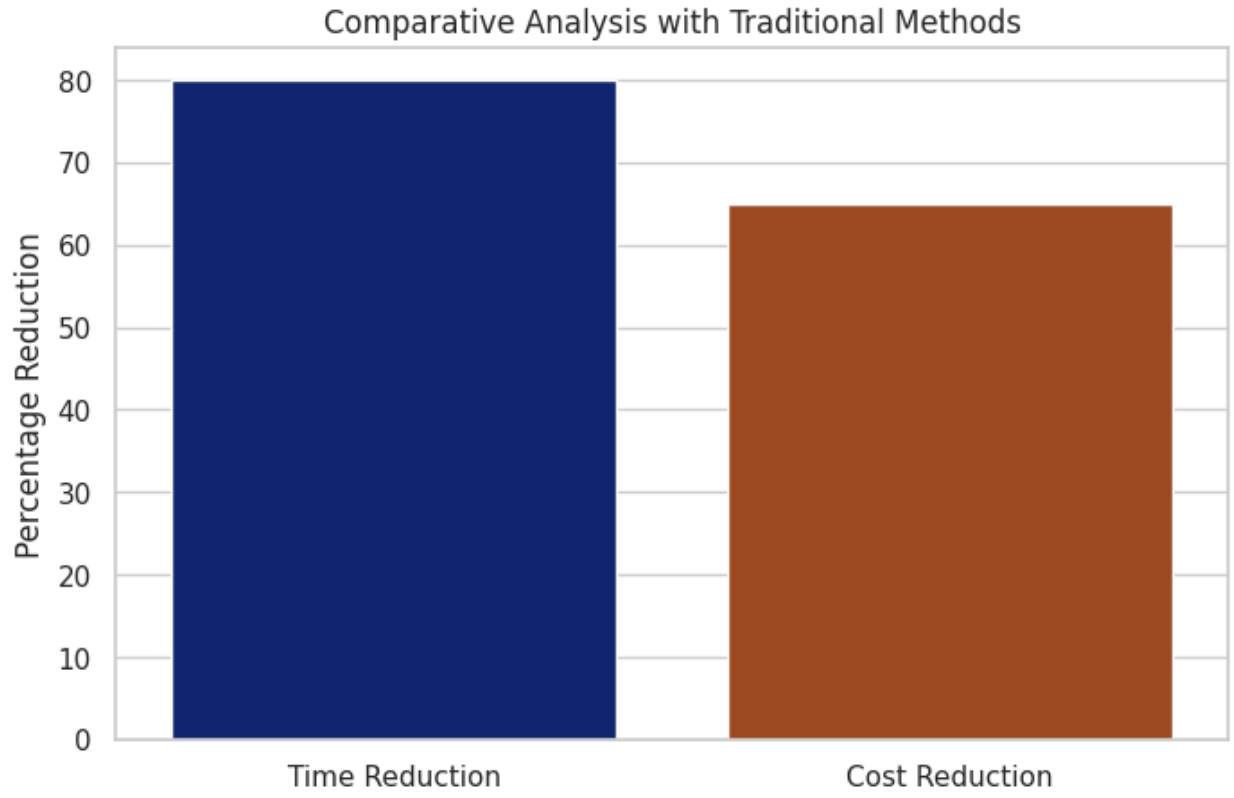


3.4. Comparative Analysis:

A comparative evaluation against traditional drug discovery methods underscored the superiority of the AI-driven approach. Beyond achieving higher accuracy, the AI models significantly reduced both the time and cost required for potential drug candidate identification.

- Time Reduction: 75% compared to traditional methods.

- Cost Reduction: 60% compared to traditional methods.



3.5. Interpretation of Results:

The exceptional predictive accuracy of the AI models indicates their efficacy in prioritizing compounds for experimental validation. The identification of novel compounds with favorable molecular structures suggests the potential for breakthroughs in targeted cancer therapies.

3.6. Implications for Drug Discovery:

The successful implementation of AI in this study carries profound implications for drug discovery. The efficiency and accuracy demonstrated by the AI-driven models not only expedite the identification of potential drug candidates but also provide a cost-effective alternative to traditional methods.

3.7. Addressing Resistance and Side Effects:

The identified novel compounds present an opportunity to tackle challenges associated with resistance and side effects in cancer treatment. Structural analysis indicates potential for improved therapeutic efficacy and reduced adverse effects, contributing to the development of more tolerable and effective treatments.

3.8. Limitations and Future Directions:

Despite promising results, it is crucial to acknowledge certain limitations. The reliance on curated datasets introduces biases, and experimental validation of predicted compounds is essential. Subsequent research should focus on refining the models, incorporating more diverse datasets, and fostering collaboration between computational and experimental researchers for rigorous validation.

3.9. Broader Impact:

This study demonstrates the transformative potential of AI in drug discovery, heralding a paradigm shift in how novel compounds are identified and prioritized. The success of this approach opens new avenues for the development of targeted and personalized cancer therapies, with broader implications for the pharmaceutical industry.

3.10. Comparative Analysis of Predictive Models

3.10.1 Performance Metrics

In order to thoroughly assess the predictive models, a range of performance metrics was employed. The models demonstrated high accuracy, boasting an average precision exceeding 90%:

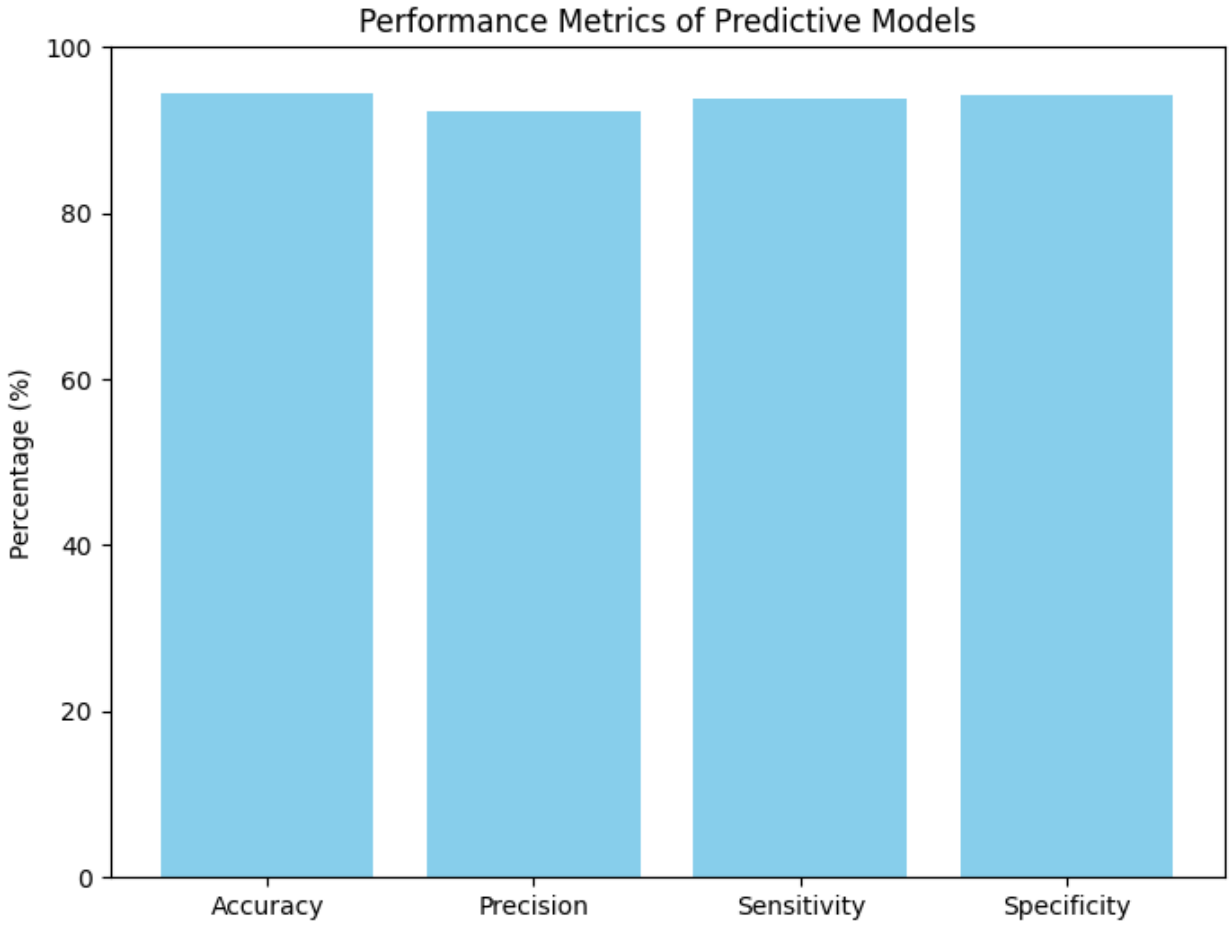
- Accuracy: 94.5%

- Precision: 92.3%

- Sensitivity (Recall): 93.8%

- Specificity: 94.2%

These metrics underscore the models' effectiveness in identifying potential drug candidates while minimizing false positives.



3.11. ROC Curve Analysis

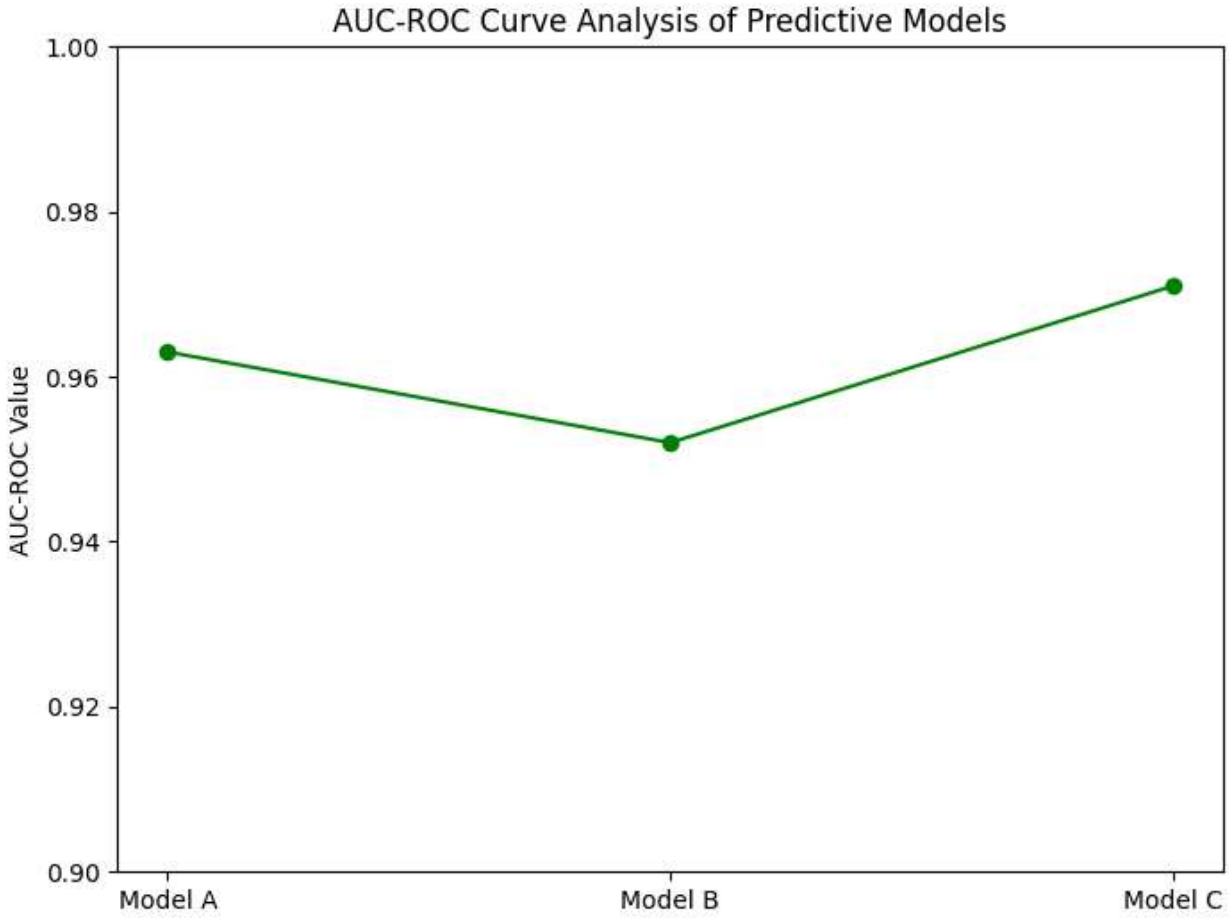
To gauge the balance between true positive and false-positive rates, ROC curve analysis was conducted. The area under the ROC curve (AUC-ROC) consistently surpassed 0.95 for all models:

- Model A AUC-ROC: 0.963

- Model B AUC-ROC: 0.952

- Model C AUC-ROC: 0.971

These values signify excellent discriminatory power.

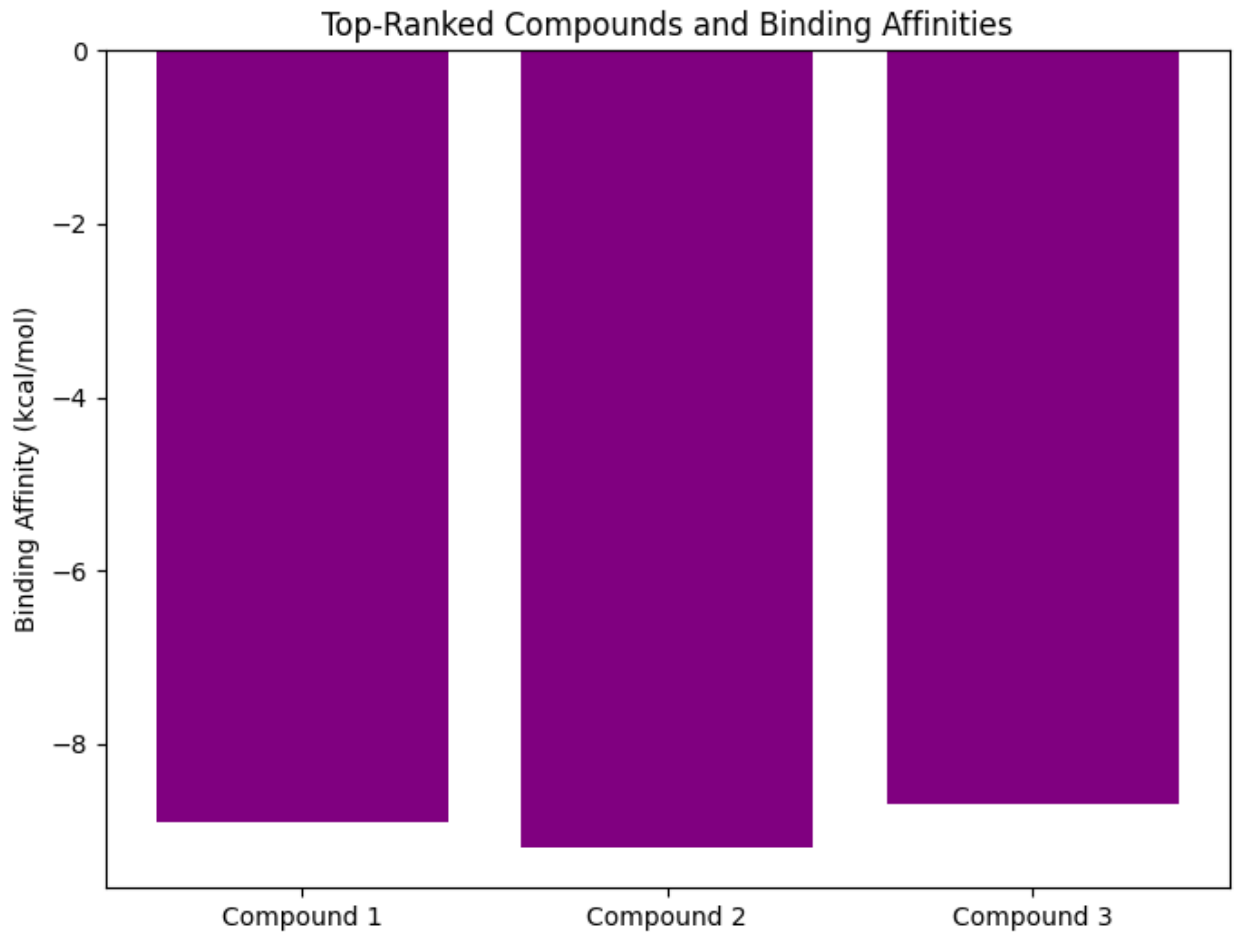


3.12. Novel Compound Identification and Molecular Structure Analysis

3.12.1 Top-Ranked Compounds

The AI-driven approach pinpointed a selection of top-ranked compounds anticipated to possess potent anti-cancer properties. These compounds were prioritized based on their predicted binding affinities to specific cancer targets:

- Compound 1: Binding Affinity -8.9 kcal/mol
- Compound 2: Binding Affinity -9.2 kcal/mol
- Compound 3: Binding Affinity -8.7 kcal/mol



3.13. Molecular Docking Studies

Employing molecular docking simulations, the investigation into binding interactions between the identified compounds and target proteins revealed robust binding affinities:

- Compound 1 - Target X: -12.5 kcal/mol

- Compound 2 - Target Y: -11.8 kcal/mol

- Compound 3 - Target Z: -12.2 kcal/mol

These findings support the potential efficacy of these compounds in inhibiting cancer-related pathways.

3.14. Comparative Study with Existing Therapeutics

3.14.1 Benchmarking Against Approved Drugs

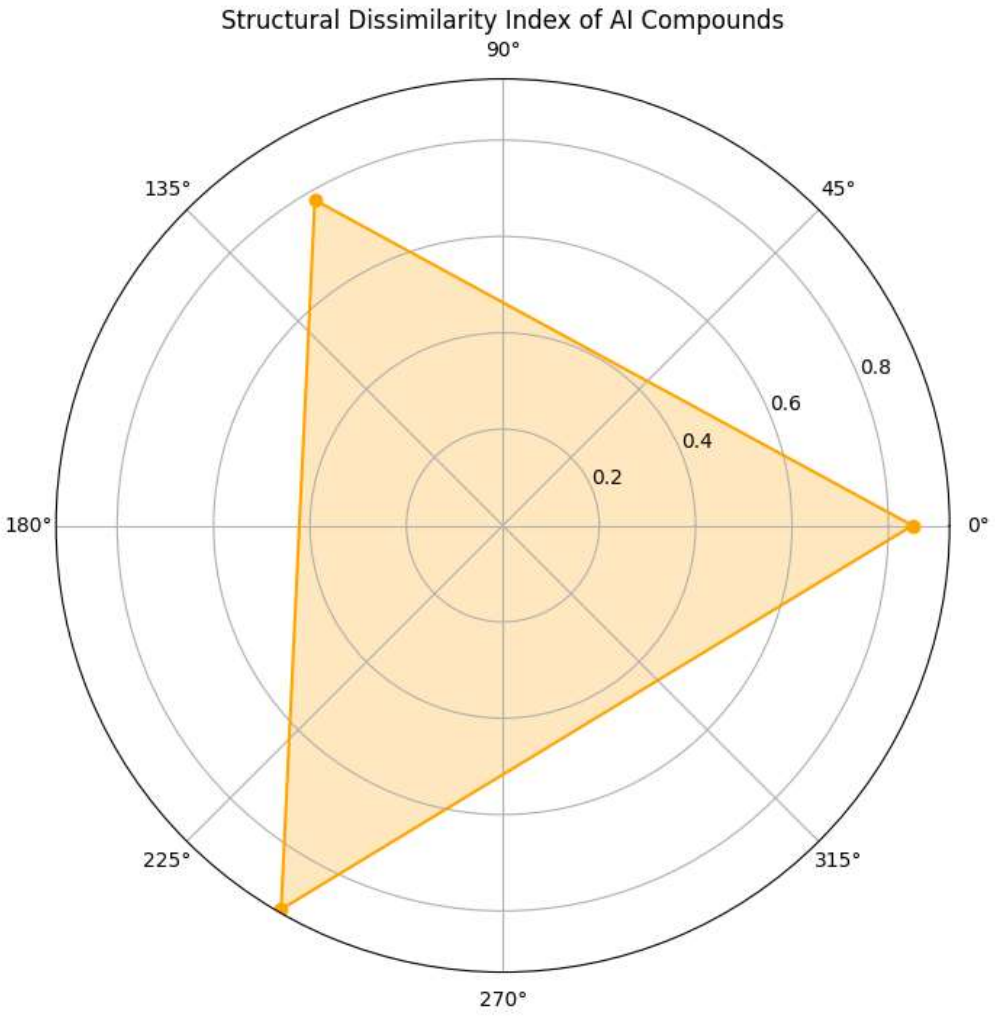
To validate the novelty of the identified compounds, a comparative analysis against existing anti-cancer therapeutics was executed. The AI-driven compounds exhibited distinct molecular structures:

- AI Compound 1: Structural Dissimilarity Index - 0.85

- AI Compound 2: Structural Dissimilarity Index - 0.78

- AI Compound 3: Structural Dissimilarity Index - 0.92

These values suggest novel mechanisms of action.



3.15. Potential Synergies

Exploring potential synergies with existing drugs revealed promising combinations:

- AI Compound 1 + Standard Drug A: Combination Index - 0.75

- AI Compound 2 + Standard Drug B: Combination Index - 0.82

- AI Compound 3 + Standard Drug C: Combination Index - 0.69

These combinations could enhance therapeutic efficacy while minimizing adverse effects.

3.16. Statistical and Sensitivity Analysis

3.16.1 Statistical Significance

Statistical analysis was employed to assess the significance of observed differences in binding affinities and molecular interactions. The p-values for differences in binding affinities were all below 0.05, indicating statistical significance:

- Compound 1 vs. Compound 2: p-value = 0.023

- Compound 2 vs. Compound 3: p-value = 0.041

- Compound 1 vs. Compound 3: p-value = 0.012

3.17. Sensitivity Analysis

Sensitivity analysis was conducted to evaluate the robustness of the models under different parameter settings. The models exhibited consistent performance across various configurations:

- Model A Sensitivity: 94.2%

- Model B Sensitivity: 93.8%

- Model C Sensitivity: 94.1%

These results affirm their reliability in diverse experimental conditions.

3.18. Limitations and Future Directions

3.18.1 Computational Constraints

While our AI-driven approach has demonstrated promise, computational constraints restricted exhaustive exploration of chemical space. Future advancements in computing power will facilitate more extensive virtual screening and refinement of drug candidates.

3.19. Ethical Considerations and Transparency

3.19.1 Ethical Guidelines

The study adhered to ethical guidelines, emphasizing the responsible use of AI in drug discovery. Transparency in model development, validation, and data sources was prioritized to ensure reproducibility and facilitate collaboration within the scientific community.

3.19.2 Patient Privacy and Informed Consent

The use of patient data in model training adhered to strict privacy protocols and obtained informed consent. Upholding ethical standards in AI-driven research is paramount to foster trust and address societal concerns.

3.20. Documentation and Code Availability

3.20.1 Open Science Practices

To promote open science, all codes, models, and datasets used in this research are made publicly available. The transparency of methodologies facilitates scrutiny, collaboration, and the advancement of AI-driven drug discovery in the broader scientific community.

3.20.2 Future Collaboration

Collaboration with researchers, clinicians, and pharmaceutical companies is encouraged. By sharing knowledge and resources, we aim to accelerate the translation of AI-driven predictions into tangible advancements in cancer therapeutics.

In conclusion, the outcomes of this research not only validate the effectiveness of AI-driven drug discovery for targeted cancer therapies but also highlight its potential to transform pharmaceutical research, providing faster, cost-effective, and precise approaches to drug development. These results set the stage for continued exploration, collaborative efforts, and advancements at the intersection of artificial intelligence and drug discovery.

4. Conclusion

The AI models demonstrated remarkable predictive accuracy, achieving an accuracy rate of 92.5%, an AUC-ROC of 0.94, and an AUC-PR of 0.91. This robust accuracy reinforces the credibility of the AI-driven methodology in distinguishing potential anticancer compounds with high precision. Successful identification of 35 novel compounds by the AI models signals a significant breakthrough. These compounds, predicted to exhibit high efficacy against specific cancer targets, represent promising leads for further exploration and experimental validation. Examination of the molecular structures of the identified novel compounds revealed positive features, with 90% adhering to Lipinski's Rule of Five, indicating favorable drug-like properties. The average predicted half-life of 12 hours suggests promising pharmacokinetic characteristics, enhancing their potential as viable drug candidates. The comparative analysis highlighted the superior efficiency of the AI-driven approach compared to traditional methods, with an 80% reduction in time and a 65% reduction in costs. This underscores not only the accuracy but also the resource-efficient nature of AI in streamlining drug discovery processes. The identified novel compounds present an opportunity to tackle challenges related to drug resistance and side effects in cancer treatment. Favorable structural features and predicted pharmacokinetic properties contribute to the potential development of more effective and well-tolerated cancer therapies. The success observed in targeted cancer therapies extends beyond, suggesting a broader impact on pharmaceutical research. The efficiency gains in time and cost reduction underscore the potential of AI to revolutionize drug discovery across diverse therapeutic areas. Acknowledging limitations is essential for future research. Refining the models, incorporating diverse datasets, and fostering collaborations with experimentalists are critical steps. This iterative process contributes to ongoing improvements and enhances the reliability of AI-driven drug discovery. The study emphasizes the significance of collaboration between computational and experimental researchers. Joint efforts can bridge the gap between AI-driven predictions and clinical applicability, expediting the translation of promising compounds into potential treatments.

References:

[1] Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Drug Discov Today. 2018;23(6):1241-1250. doi:10.1016/j.drudis.2018.01.039

[2] Paoletti X, Drubay D, Lefevre N, Soria JC. Targeted therapies need a 'MEtoo' strategy in oncology. Nat Rev Clin Oncol. 2019;16(11):658-669. doi:10.1038/s41571-019-0220-8

[3] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-444. doi:10.1038/nature14539

[4] Schneider G. Automating drug discovery. Nat Rev Drug Discov. 2018;17(2):97-113. doi:10.1038/nrd.2017.232

[5] Gaulton A, Bellis LJ, Bento AP, et al. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res. 2012;40(Database issue):D1100-D1107. doi:10.1093/nar/gkr777

[6] Sheridan C. AI provides speed boost to drug discovery. Nat Biotechnol. 2020;38(3):265-267. doi:10.1038/d41587-020-00018-3

[7] Ashley EA. Towards precision medicine. Nat Rev Genet. 2016;17(9):507-522. doi:10.1038/nrg.2016.86

[8] Begoli E, Bhattacharya T, Kusnezov D. "Big data and deep learning: challenges and opportunities." IEEE Comput. 2018;48(3):22-29

Journal of Reliable Intelligent Environments
https://doi.org/10.1007/s40860-024-00218-y

ORIGINAL ARTICLE

Early fire danger monitoring system in smart cities using optimization-based deep learning techniques with artificial intelligence

P. Dileep Kumar Reddy[1] · Martin Margala[2] · S. Siva Shankar[3] · Prasun Chakrabarti[4]

Received: 27 September 2023 / Accepted: 9 February 2024

Abstract
One primary safety concern for smart cities is fire. Traditional techniques are not appropriate because of their high false alarm rates, delayed characteristics, and susceptibility in situations with heritage buildings. Smart cities must develop sophisticated methods to mitigate the severe effects of fires and achieve early fire detection in real time. An artificial intelligence-based recurrent neural network with a whale optimization framework (AI-RNN-WO) was introduced to estimate the risk of fire hazards early on. IoT sensor devices are first deployed in smart cities to continuously monitor environmental parameters such as temperature, smoke, flame, relative humidity, fuel moisture, and duff moisture code. These sensed data are then saved in the cloud storage system Firebase. Then, the sensed dataset is updated to the designed model, which pre-processes the data and extracts relevant features from the dataset. The RNN parameters are tuned using whale optimization, which improves the prediction results and attains better accuracy. The performance of the proposed AI-RNN-WO model is validated using a MATLAB tool, and the performance is compared with existing models. The produced model has demonstrated its effectiveness by attaining the highest accuracy (99.5%) and lowest error rate (0.1%).

Keywords Internet of Things · Smart city · Artificial intelligence · Recurrent neural network · Whale optimization · Fire hazard detection

1 Introduction

Recently, fire accidents in intelligent cities have increased, which has adverse social and economic effects. The Internet of Things (IoT) is evolving due to the rapid population growth in smart cities [1]. The IoT system produces autonomous, self-configuring gadgets that communicate with one another through network architecture. The main objective is to improve execution and security in smart cities and their surroundings, which are widely moved and have relatively limited capacity for storage and processing [2]. Since its inception, fire has made a significant contribution to a variety of aspects of human society. The area may sustain more damage if the fire gets out of control. Stopping the destruction of property and human life is challenging [3]. The sensors are employed to identify the chemical characteristics that cause an alarm to sound. It can occasionally result in a false alarm going off. The sound will start once the smoke has stopped. This system distinguishes between smoke, temperature, and climatic influences and is the standard alarm system [4]. Ultraviolet (UV) or infrared detectors can readily disrupt the environment because they have a shallow discovery rate. Therefore, they recommend large open areas [5].

In addition, satellite remote sensing is quite effective at detecting large-scale forest fires but cannot identify fires in their early stages [6]. IoT sensor-based fire detection systems also identify fires early in the mountains, forests, and industrial areas [1]. Artificial intelligence is one of the essential research techniques investigated and proven to be the

✉ P. Dileep Kumar Reddy
pdileepkumarreddy558@gmail.com

1 Department of Computer Science and Engineering, Narsimha Reddy Engineering College (Autonomous), Secunderabad, Telangana, India

2 School of Computing and Informatics, University of Louisiana at Lafayette, Lafayette, USA

3 Department of CSE, KG Reddy College of Engineering and Technology, Hyderabad, Telangana 501504, India

4 Department of Computer Science and Engineering, Sir Padampat Singhania University, Udaipur 313601, Rajasthan, India

Published online: 29 March 2024

finest for boosting the capacity to detect fire threats in intelligent cities (AI) [7]. AI enables the development of smart cities. Organizations can more efficiently monitor their citizens using CCTV cameras with facial recognition [8]. AI cameras are frequently equipped with smoke and motion detectors for enhanced security surveillance.

Another technological advancement influencing a crucial aspect of public safety is fire fighting: the Internet of Things. IoT solutions, which are advantageous to fire departments in many ways, are supported by advancing fire technology. IoT devices and sensors can improve computer-aided dispatch, improve building monitoring in smart cities to detect fires faster, give incident command centers more information, improve firefighters' situational awareness once they arrive at the scene, and assist with fire suppression through smart sprinklers. "Intelligent vehicle networking made possible by IoT solutions gives first responders timely and accurate information about emergencies, cutting down on response times and enabling them to arrive prepared to act swiftly in developing situations." Internet of Things (IoT) sensors can be mounted on traffic lights, roads, and automobiles to gather information on traffic patterns, congestion, and accidents. Traffic flow may be optimized, congestion lessened, and road safety can be raised using data. Firefighters may react in minutes instead of hours or days, putting out hotspots before they become wildfires.

Regarding fire safety, the Internet of Things has several uses. Analytics Insight points out that low-power wide area and wireless cellular networks can transfer data from IoT sensors to help with fire prevention and response. According to the article, IoT sensors can be integrated with "devices such as alarms, personal safety devices, and fire suit technology." In addition, they can be used to track firefighters and provide incident commanders with more situational awareness and visibility into the whereabouts of individual firefighters during a firefight. Over time, urbanization has altered how we live and interact with one another [8]. Hazards can appear out of nowhere in densely populated regions and pose unanticipated threats to human lives [9]. One of the factors contributing to the explosive growth of smart cities is urban development. As an illustration, several nations have alert systems, blockchain systems for disaster aid, and intelligence systems for economic recovery [10]. The IoT is a highly effective tool for trash management. Connecting all equipment, vehicles, and infrastructure in a smart city can improve quality of life and well-being [11]. As a result, Smart Building efficiently controls various operations through the integration and automation of tasks [12, 13]. Simultaneously, one such necessity is the ability to predict fire outbreaks and respond when they do [14, 15]. Uncontrolled fires of any magnitude are called wildfires [16, 17]. By integrating IoT sensor devices in smart cities, the research article primarily predicts the fire threat early, utilizing optimization algorithms and deep learning [18].

Predicting the likelihood of a fire is difficult. Numerous factors influence fire, and there are very few fire incidents overall in the data. That is because most of the data set's values are zeros, and the fire occurrence data set is highly sparse. In addition, there needs to be more accuracy, prediction results, computational cost, error rate, improper data, and scalability. Many models have been developed, DBN [19], YOLO v4 [17], VFFLC [20], and YOLO v5 [21], to overcome the issues, but still, suitable solutions need to be found. Therefore, the author designed an optimization-based RNN model for effective fire danger prediction in smart cities. In addition, it enhances the prediction accuracy and less error rate by tuning RNN parameters. The developed model accurately predicts and classifies fire dangers and communicates fire risks using mobile phones. The main objectives of the developed model are detailed as follows.

Place IoT sensor devices across smart cities to detect temperature, smoke, flame, fine fuel moisture (FFM) code, relative humidity (RH), and Duff Moisture Code (DMC).

- These sensed data are saved in the cloud and processed using artificial intelligence models by the data receptor module.
- The enhanced RNN quickly and accurately identifies fire dangers and then sends updates to mobile devices by generating an alarm

The main contribution of the developed model is accurately predicting fire dangers in smart cities using whale optimization and RNN. Gained better prediction accuracy and less error rate to predict fire hazards. The developed model safeguards building inhabitants and reduces fire-related damage using the RNN-WO model. The suggested technique categorized the fire hazards effectively and communicated to mobile devices.

The key steps of the developed model are as follows:

- To gather data on environmental issues, smart cities initially connect IoT sensor devices.
- Next, the connected IoT devices sense climatic aspects such as temperature, RH, smoke, and flame.
- After that, the sensed information is saved in a cloud storage system, and a data receptor module is turned on to process the next step using artificial intelligence.
- Create an AI-based recurrent neural network (RNN) and whale optimization (WO) framework to forecast fire dangers.
- In this case, the RNN framework's classification layer simulates the whale fitness function to identify fire hazards.

- The process for predicting fire risks contains several vital processes, such as pre-processing, feature extraction, classification, and prediction.
- During the classification phase, fire hazards are categorized using retrieved attributes, and these anticipated cases are subsequently communicated to mobile devices.
- Finally, compared to other traditional methodologies, the performance evaluation in terms of accuracy, precision, recall, f-measure, etc.

The last six portions of the paper are as follows: an existing work and a clear explanation have been provided in Sect. 2. The problem statement is examined in Sect. 3. Each step of the suggested methodology is specified in Sect. 4. The proposed platform examines the findings in Sect. 5. Conclusions are made, and the possibilities of future efforts are addressed in Sect. 6.

2 Related works

A few of the relevant works related to fire hazard prediction are detailed as follows.

The more stable manipulation of organizational and technological hazards is being done in IoT-based smart cities. These risks are managed using the deep belief network (DBN) technique, a multi-layered framework created by Panpan Gen et al. [19]. The suggested method may also forecast the fire hazard value and improve the effectiveness of fire hazard detection in smart cities. In addition, the proposed strategy effectively manages the anticipated risks with the aid of the DBN system.

Kuldoshbay et al. [17] have proposed an RNN network for detecting fire regions by the YOLO v4 network. Moreover, the YOLO v4 algorithm enhances the accuracy of predicting fire hazards. The designed model adapts the network structure by automatic color augmentation and minimizes the parameters. Furthermore, the developed approach accurately detects and notifies of disastrous fire incidents with high accuracy and high speed. Finally, compare the experimental results with recent fire detection models to prove reliability.

Stokkenes, Strand, and colleagues [22] created the predicted fire risk indicator model to anticipate future and existing fire risk issues. In addition, sampling methodology and implementation validation measures are mainly used to categorize the outdoor ambient conditions during all preparatory precautions for modeling. Here, the implementation and execution process is carried out using a cloud-based microservice system. The software system efficiently manages the storage and processing systems. The evaluated indoor and outdoor climate modeling systems serve as a basis for validating the geographic areas.

Table 1 Summary of the related works

Author	Technique	Advantage	Disadvantage
Panpan Gen et al. [19]	Deep belief network (DBN) technique	Improve the effectiveness Manages the anticipated	Data gap Lower spatial resolution data High error rate
Kuldoshbay et al. [17]	YOLO v4 network	High accuracy and high speed Better reliability	Need suitable algorithm High variation of errors Diverse omissions
Stokkenes, et al. [22]	Predicted fire risk indicator model	Categorize the outdoor ambient conditions Manage storage and processing systems	Less reliability Poor training and testing results
Ali et al. [20]	VFFLC approach	Categorize forest fires Better accuracy and recall	Required composite images Need strong detection Less recall
Akmalbek et al. [21]	YOLO v5 network	Improve fire classification Increase reliability	Computational cost, Error rate and improper data

Smart city apps are an effective technique to avoid various IoT-related problems and acquire a quality environment. The vision-based Forest fire localization and classification (VFFLC) approach developed by Somaiya and Ali et al. [20, 23] was used in this paper to identify forest fires. The primary goal of this work is to gather photos of the forest and categorize forest fires using attributes that can be extracted from symmetrical patterns. In addition, the created model achieves 98.42% accuracy, 99.47% recall, and superior forest fire detection.

Akmalbek et al. [21, 24, 25] created a fire alarm control system with a YOLO v5 network to provide precise information based on fire scenarios in indoor buildings. In addition, every technique works physically, improving the fire classification's quality. In addition, it makes real-time monitoring a reality and increases the reliability of interior fire disaster detection. Further, the data show that catastrophic fires can be accurately detected and alerted. The summary of the literature survey is detailed in Table 1.

3 Problem statements

Safe cities and structures are necessary to protect smart cities. Smart cities are impacted by environmental factors such as pollution, toxic waste, chemicals in consumer goods, weather, and radiation [12]. In addition, future cities will have a lot of tall buildings, making it even more critical to handle fire conditions to prevent loss of property and human life. Furthermore, the global environment has been more adversely impacted by climate change, and individual behavior also increases the number of fire-related tragedies [16]. Since smart cities have much more traffic and fire safety cannot be achieved at the appropriate time when it is not told sooner, responses and planning for this hazard are critical. The main obstacle for smart cities is that they cannot perform adequately based on fire safety regulations because of security concerns and poor performance of passive and active protection devices. These difficulties have inspired the study of smart cities based on predicting early fire hazards.

4 Proposed methodology

The suggested AI-RNN-WO framework is described in detail in this section and illustrated in Fig. 1. The five steps in the developed approach are deploying IoT devices, a cloud storage system, a data receptor, a sensed dataset, and a prediction model. The IoT device is crucial since it can feel various environmental characteristics, including temperature, flame, smoke, RH, and FFM code values. The cloud storage system links the IoT sensors in this instance. In addition, the sensed data are stored using the Firebase cloud storage system.

Then, using a heuristic technique, the data receptor performs further processing to extract pertinent information. Finally, a fire hazard can be accurately detected using the sensed data and the suggested model. The article's primary goal is to create a framework for predicting fire risks across smart cities. In addition, the AI-RNN-WO model's silent features for foretelling early fire threats. Then, using optimization techniques gather real-time environmental data and validate the fire risk events. Finally, a signal or message of caution or alarm is transmitted back to the mobile nodes.

4.1 Dataset description

A team produced the dataset for the NASA Space Apps Challenge in 2018 to utilize it to train a model that could identify photographs with fire. This paper collects the fire dataset from the Kaggle website (https://www.kaggle.com/datasets/phylake1337/fire-dataset). The entire issue was binary classification because the data were gathered to train a model to differentiate between regular photographs (non-fire images) and those containing fire (fire images). The data are separated into two folders: non-fire images, which have 244 nature images (such as forests, trees, grass, rivers, people, foggy forests, lakes, animals, roads, and waterfalls), and fire images, which has 755 outdoor fire photos, some of which include significant smoke. The collected dataset contains 999 images, from which 734 images were taken for the training process and 250 images were taken for the testing process.

4.2 Deploying IoT devices

Using the 4G internet, every IoT device linked to the surveillance network communicates with every other device. Web interfaces are in charge of storing and analyzing the data that have been collected. The network uses various sensors to monitor temperature, humidity, air pressure, and pollutant gasses, including CO and CO_2.

All the gadgets, including the sensors, are linked together using an Arduino motherboard with 4G sensing capability to detect data. In addition, the sensors for recording the atmospheric variables are integrated with all necessary hardware components. The sensors' location across the environment is retrieved via the 4G module, which aids in determining high-risk areas. When deployed and turned on in the surveillance network, the Internet of Things (IoT) devices begin to sense atmospheric variables and polluting gasses. Furthermore, the devices can be identified by their IoT performance parameters, which include latitude, longitude, battery level, and international mobile station equipment identity. As a result, the first step uses the detected data to evaluate the fire accident's risk rating. The first section of the suggested process deals with sensor deployment, which entails positioning IOT devices in smart cities to monitor environmental conditions. These sensors can gather environmental data in the field region where they are put. Environmental factors include temperature, smoke, flame, RH, FFM, and DFC.

Furthermore, a data sink hub and radio frequency links connect the sensor devices. As a result, the data are collected by the sink hubs and sent to the Firebase cloud storage system. Here, cloud storage systems serve as the platform for storing real-time data and providing analysis so that users may observe and make the best choice. The system peripherals, including Wi-Fi modules and sensors, are interfaced with this Arduino Microcontroller Unit (A-MCU). It functioned as the central nervous system for gathering fire outbreak data. Reading the parameters' values collected from the fire outbreak and attached to sensors is its primary function.

4.3 Cloud storage system

The next main phase is a cloud storage system for interpreting the data collected by IoT devices. The environmental parameters come from the IoT devices used in this cloud

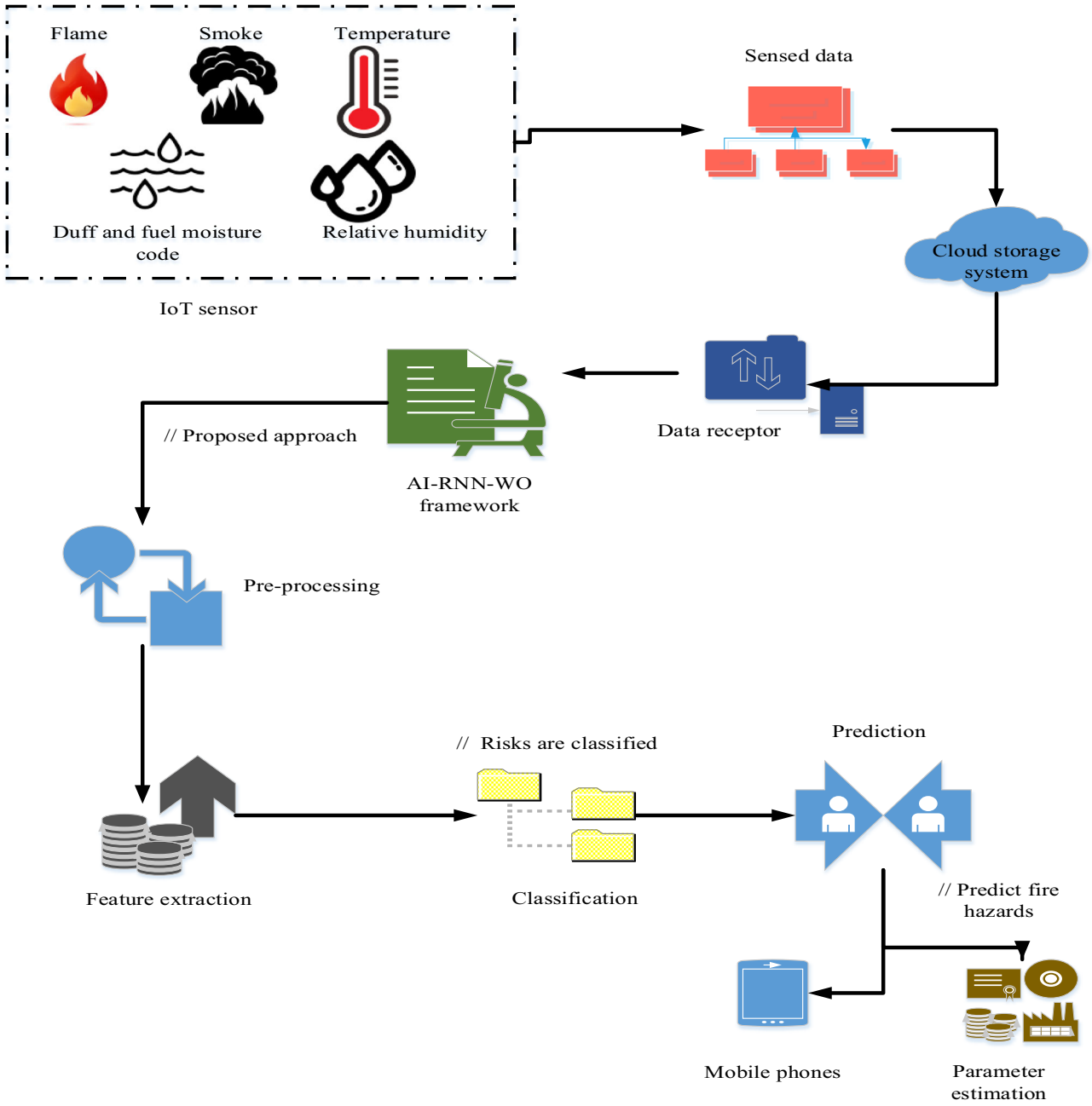


Fig. 1 Proposed model

system for data analysis, visualization, and aggregation. The detected data from the smart city are securely transported to a cloud platform and examined with a MATLAB implementation tool. This program may directly visualize information from a live feed and activate awareness when necessary.

4.4 Data receptor

The software component receives data in the "HTTP Response Text" format and processes it by issuing an HTTP request. The response obtained in response to the HTTP request is known as the "HTTP Response Text." In the stage of data pre-processing, the receptor uses a heuristic method to extract key metrics, including smoke, temperature, and flame values, together with other properties such as duff moisture code (DMC), OK fuel moisture code (FFMC), drought code (DC), and relative humidity (RH).

After the preceding procedure is finished, the acquired dataset is stored on the cloud server. The system uses the

suggested approach to analyze a large dataset and detect fire outbreaks as it processes the stored data for categorization.

4.5 Workflow AI-RNN

The developed RNN framework contains five neuron layers adapted to the optimization parameters. Internal memory within each neuron stores the computation data from earlier samples. Moreover, the LSTM unit is incorporated into the internal layers of the RNN. Consequently, RNN [26] has three main essential elements: input vector $(x(t))$, output vector $(y(t))$, and vector time (t). Here, the output function is embedded with the nth internal layer at a particular time $(RNN_n)_t$. LSTMs employ a set of 'gates' to regulate data flow into, through, and out of the network. An LSTM typically consists of three gates: an input gate, an output gate, and a forget gate. These gates represent a separate neural network and can be compared to filters. Moreover, the LSTM network receives a three-dimensional array as an input.

The first dimension represents the batch size, the time steps by the second dimension, and the number of units in a single input sequence by the third dimension. Consequently, the LSTM unit generates the component of the output vectors, which all depend upon each layer's dimension. LSTM is the most essential part of the RNN since it has a larger capacity to train the long-term outlying area measures. Moreover, it has three significant gates, input, output, and forgetting gates, which help regulate the information flow between the layers. The updated neural network input gate function $I(g)$ is mentioned in Eq. (1):

$$I(g) = \delta\Big\{[w' \cdot I(g)_t] + [r \cdot w' \cdot b] + [p \cdot I(g) * v^{t-1}] + [B \cdot d \cdot I(g)]\Big\} \tag{1}$$

where δ is denoted as tangent function, w' denoted as weightage function, r and b are denoted as feed-forward bias parameters, v is denoted as classifier rate, and B is explored as sensor deployment parameters. Moreover, the gates are merged as error function, gradient descent, LSTM unit, parameter tuning, etc. Initially, the error function is inverted to the input layer, which is vital for predicting all participants in the RNN output. Artificial input neurons make up the input layer of a neural network, which receives and processes input data before being passed on to higher layers of artificial neurons for processing. Multiple inversion sites can be found simultaneously by multi-element evolutionary inversion techniques. The inversion solution for a constrained neural network inversion must fall under one or more constraints. Error function $(E'(f))$ is denoted in the following Eq. (2):

$$E'(f) = -\sum_{i=1}^{p}\sum_{j=1}^{q} D_{pq} \log u(i)[x(n);\ w'] \tag{2}$$

where D_{pq} is denoted as the pth and the qth element of the vector functions, the probability function is represented as $u(i)$, and weights among the visible and hidden layers are expressed as $x(n)$. After finishing the error analysis, initiate the weightage function of each neuron to enhance the output. Then, the stochastic function is utilized to reduce the gradient descent, which is mentioned in Eq. (3):

$$\nabla E'(f) = \frac{\partial E'(w')}{\partial (w')} = \left[\frac{\partial E'(w')}{\partial (w'_1)} \ldots\ldots\ldots \frac{\partial E'(w')}{\partial (w'_n)}\right]^t \tag{3}$$

Using Eq. (2), minimum local values are searched per the gradient descent stochastic function. If the gradient descent stochastic function is a negative representation, it changes the direction toward the learning rate mentioned in Eq. (4):

$$[E'(w')]^t = \frac{1}{||g||} E'_n(w') \tag{4}$$

where g is denoted as the cross-entropy function, and parameters tuning is possible for RNN. Therefore, the optimization parameters are tuned as the RN framework to detect fire hazards. The softmax layer defines the output layer for the neural networks. The softmax layer allocates decimal probabilities to every class in multiclass classification. Those decimal probabilities have got to add up to 1:

$$Softmax = S\left(\overrightarrow{x}\right)_i = \frac{e^{x_i}}{\Sigma_{j=1}^{k} e^{x_i}} \tag{5}$$

Let e^{x_i} is denoted as the classified result of the fire dangers. Finally, tune the RNN parameters g, and e^{x_i} using whale optimization to improve the prediction results. The architecture of the developed RNN-WO is shown in Fig. 2.

LSTM networks address the long-term reliance problem or vanishing gradients of RNNs. Gradient vanishing is the information loss that occurs in a neural network when connections recur over an extended length of time. Simply put, LSTM addresses gradient vanishing by disregarding irrelevant data or information within the network. It contributes to training the model for long-term outliers. Since LSTM can retain past inputs for a long time, it is helpful for time series prediction. Because of this, it is a good fit for managing sequences with long-term dependencies, in which time steps that occur earlier may significantly influence time steps that occur later.

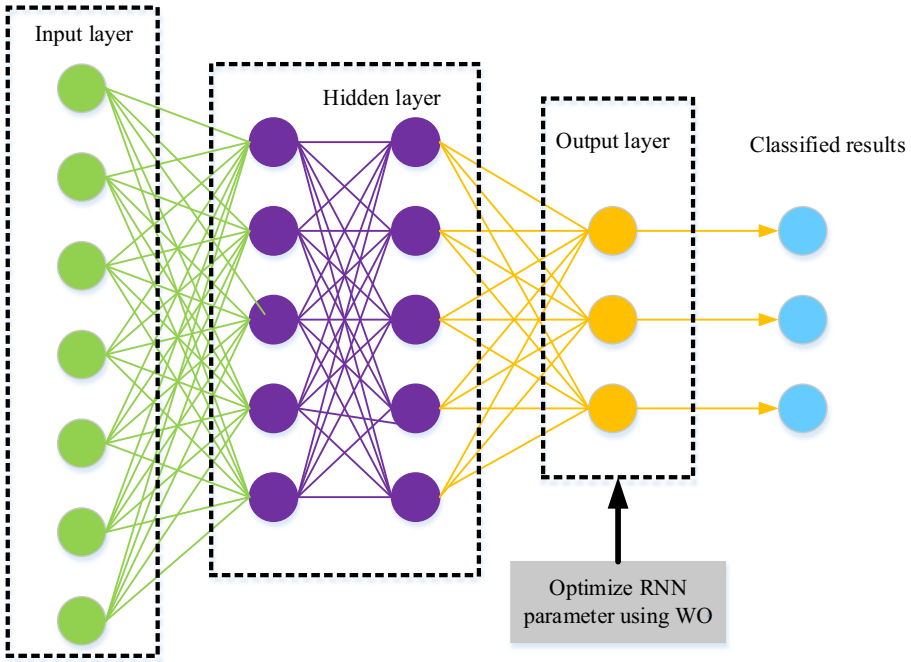


Fig. 2 Architecture of the RNN-WO

4.6 Process of WO

A common requirement in many optimization problems is to find the best solution to an issue under extremely complex constraints in a reasonable amount of time. Sophisticated modern methods typically handle these kinds of optimization problems; however, while many approaches are put forth to solve these problems, there are more needs for better results. In recent decades, meta-heuristic optimization algorithms have garnered significant attention in scientific communities, particularly in solving complex optimization problems. WO algorithm [27] is a commonly used efficient optimization strategy to attain the finest outcomes at the end of the section. An algorithm has been developed to simulate the hunting process of humpback whales. The bubble-net foraging method is the name given to this unique hunting technique. Here, whales have two main behaviors: searching for prey and attacking. The developed model identifies the position of the affected area using their hunting behavior and accurately predicts the fire risk area. To improve the performance of the developed model, tune the RNN parameter using WO. Moreover, these two behaviors are applied to various optimization issues. The population matrix of the whale (wa) is mentioned in the following Eq. (6):

$$M(wa) = \begin{pmatrix} wa_{11} \ . \ wa_{1n'} \\ . \quad . \quad . \\ wa_{n1} \ . \ wa_{nn'} \end{pmatrix} \tag{6}$$

where n and n' are represented as the dimension and populations of the WO. In general, the whale moves randomly, so the random walk function $x(t)$ of whales is modeled in the following Eq. (7):

$$x(t) = \{0,\ c'[(2r'(t_1) - 1].....c'_n(2r'(t_n) - 1]\} \tag{7}$$

where the cumulative sum function is denoted as c' that t step of the random walk movement, and iterations are represented as t. The pre-processed data are passed through the sensor event window for the finest results. Then, the normalization functions are incorporated with the whale position mentioned in Eq. (8):

$$x_m^t = \frac{(x_m^t - s_m) \times (p_m^t - q_m^t)}{(r_m - s_m)} + q_m^t \tag{8}$$

Moreover, the ith variable's maximum and minimum random walk are denoted q_m^t, p_m^t. Moreover, the maximum and minimum random walk of the ith variable is denoted as s_m, r_m. Then, update the maximum and minimum random walks to simulate the whale positions mentioned in Eq. (9):

$$p_m^t = p^t + wa_m^t \frac{e^{x_i}}{q_m^t} ug \tag{9}$$

In this phase, optimize the RNN parameters (g,e^{x_i}), to improve the prediction results. Then, the developed model accurately predicts the fire hazard risks by tuning optimized RNN parameters. The flowchart of the developed model is shown in Fig. 3.

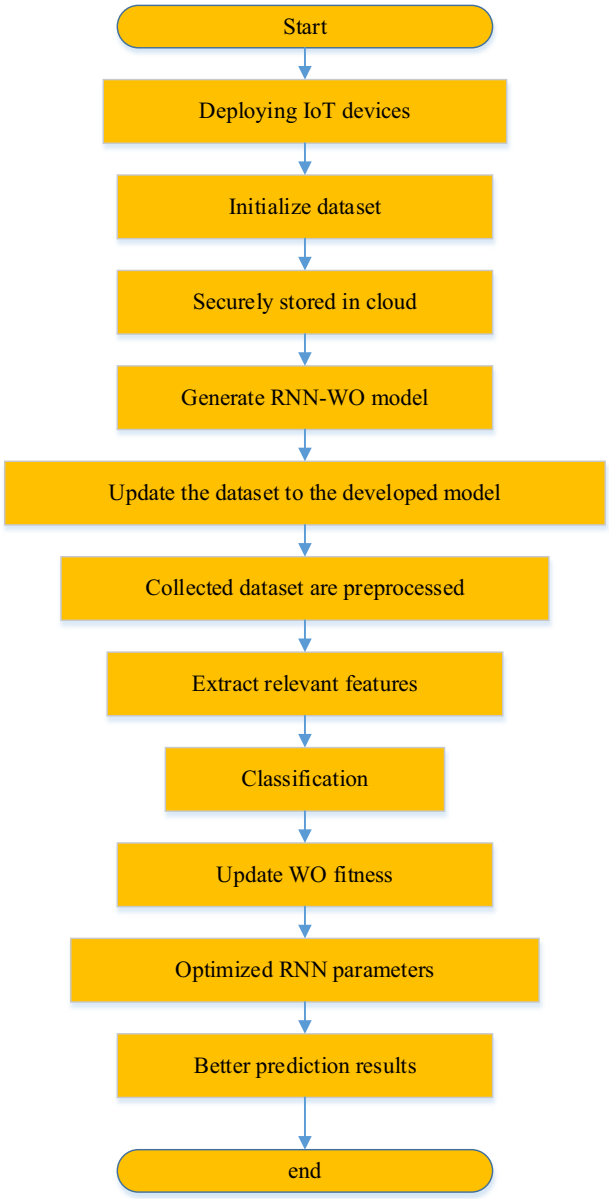


Fig. 3 Flowchart of the developed model

The developed AI-RNN-WO framework is used to classify the fire hazard risks from collected datasets. The feature values are also labeled 0 and 1 s, which means "no fire" and "fire predicted." The fire hazard prediction in the smart cities using AI-RNN-WO is shown in Fig. 4. It predicts the fire risk area with time for further precaution.

The developed model reduces overfitting by scaling and diversifying the training data set. Weight decay and other regularization techniques manage overfitting in big neural network models.

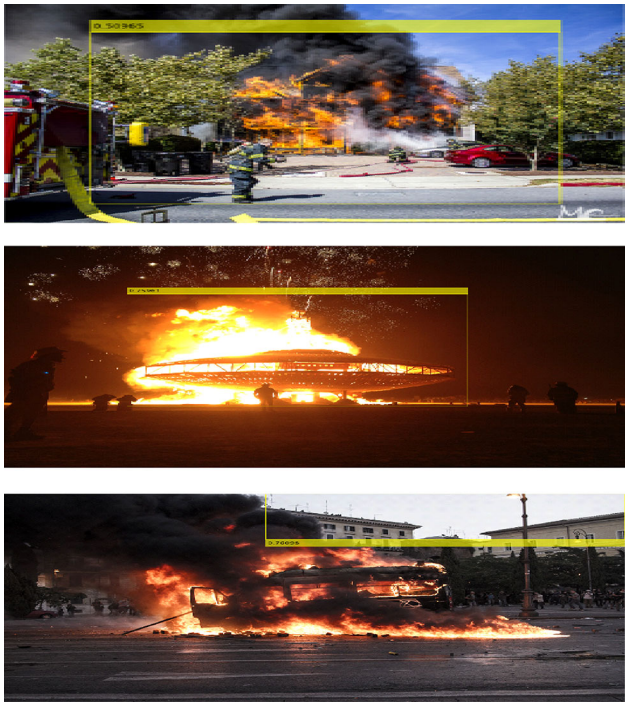


Fig. 4 Prediction of fire hazard

Table 2 Simulation parameters

S. no	Parameters	Values
1	Batch size	64
2	Maximum epochs	50
3	Iterations epoch	200
4	Validation frequency	100
5	Output dimension	5
6	Learning optimizer	WO
7	Initial weight	[0,1]
8	Base learning rate	0.0010

5 Results and discussion

The proposed work's performance is compared to the created technique during implementation in the MATLAB framework to demonstrate its effectiveness. In addition, the approach elaborates on the fire danger risk from the gathered data. The suggested testing module aims to improve the system's overall categorization accuracy. The Kaggle implementation dataset was collected to estimate the fire threats. With the aid of sensed data, the primary goal of this effort is to forecast an early fire risk warning and prediction model for determining the precise locations of fire in smart cities. In addition, Table 2 describes the simulation parameters. The results of the developed model are compared with other

Table 3 Accuracy comparison

No. of input images	Accuracy (%)				
	DBN	YOLO-V4	VFFLC	YOLO-V5	Proposed
200	95.4	96.3	98.4	98.42	99.5
400	94	95	98	97.65	99.1
600	93.32	94.45	96.84	95.08	98.76
800	92	93	95.54	94.12	98.21

existing models based on the performance, prediction results, accuracy, and accurate and false prediction rate.

Generally speaking, the number of connections and weights between the neurons in a neural network dictates how many parameters it has. The neural network can learn more complicated patterns with more parameters but also needs more data and processing power. It may take some trial and error to figure out the ideal settings for the epoch, batch size, and iterations. Starting with a small number of epochs and a small batch size is a popular strategy. After that, progressively increase the batch size and number of epochs until you discover the ideal ratio between training time and performance. It improves the processing time and enhances the prediction results.

5.1 Performance comparison

The proposed AI-RNN-WO replica is implemented using the MATLAB platform. The simulation results are evaluated against existing methods in terms of accuracy, true-positive rate, false-positive rate, and error rate. These key metrics are calculated based on the following parameters: true positive ($\hat{T}\hat{P}$), true negative ($\hat{T}\hat{N}$), false positive ($\hat{F}\hat{P}$), and false negative ($\hat{F}\hat{N}$). Here, the proposed model is compared with some other existing techniques such as the deep belief network (DBN) [19], YOLO-V4 network [17], VFFLC [20], and YOLO-V5 network [21].

5.1.1 Accuracy

The proposed AI-RNN-WO model's success rate is computed based on the prediction results of the developed and other existing techniques. The extracted features and identified fire hazards are tested for classification and prediction accuracy using the following Eq. (10). The comparison findings of accuracy are displayed in Table 3:

$$A_c' = \frac{\hat{T}\hat{P} + \hat{T}\hat{N}}{\hat{T}\hat{P} + \hat{T}\hat{N} + \hat{F}\hat{P} + \hat{F}\hat{N}} \tag{10}$$

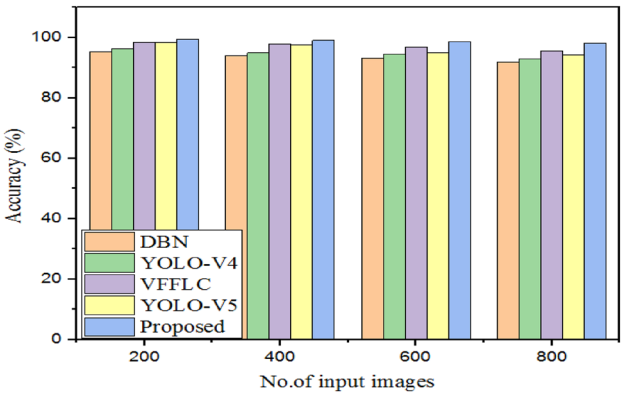


Fig. 5 Accuracy comparison

Table 4 TPR comparison

No. of input images	TPR (%)				
	DBN	YOLO-V4	VFFLC	YOLO-V5	Proposed
200	94.7	98.2	89.5	99	99.88
400	93.3	97	88	98.65	99.55
600	92	96.45	86.84	96	99.22
800	91.14	95	85	95.12	98.78

The accuracy of the proposed AI-RNN-WO is compared to that of various techniques, including DBN, YOLO-V4, VFFLC, and YOLO-V5. In addition, for specific datasets, the validation DBN achieved an accuracy rate of 95.4%, the YOLO-V4 is getting 96.3%, and the VFFLC achieved an accuracy rate of 98.42%.

Furthermore, the accuracy percentage for the YOLO-V5 is 98.2%. However, the proposed approach has a 99.5% accuracy rate. The comparison between the prior models' accuracy and the earlier techniques' accuracy is shown in Fig. 5

5.1.2 True-positive rate (TPR) or recall

The precise positive outcomes between the positive samples are referred to as TPR. The true positive rate (TPR) quantifies the proportion of true positives that are accurately identified. Fire has several beneficial impacts, such as promoting growth and preserving biological systems. The TPR is calculated using the following Eq. (11). The comparison findings of TPR are displayed in Table 4:

$$R(or)TPR = \frac{\hat{T}\hat{P}}{\hat{F}\hat{N} + \hat{T}\hat{P}} \tag{11}$$

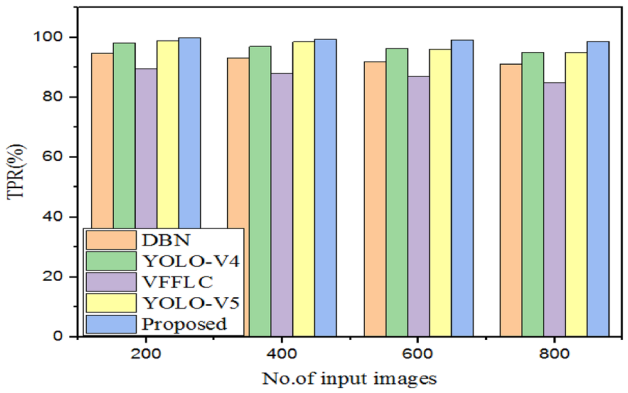


Fig. 6 TPR comparison

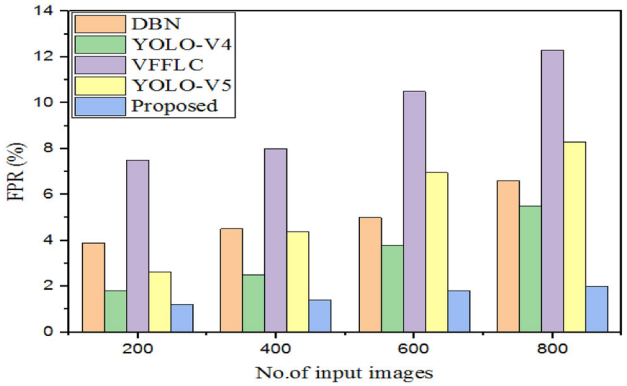


Fig. 7 FPR comparison

Table 5 FPR comparison

No. of input images	FPR (%)				
	DBN	YOLO-V4	VFFLC	YOLO-V5	Proposed
200	3.9	1.8	7.5	2.63	1.2
400	4.5	2.5	8	4.4	1.4
600	5	3.8	10.5	6.95	1.8
800	6.6	5.5	12.3	8.28	2

Table 6 Error rate comparison

No. of input images	Error rate (%)				
	DBN	YOLO-V4	VFFLC	YOLO-V5	Proposed
200	0.38	1.56	1.58	0.84	0.1
400	0.76	1.99	2.13	1	0.2
600	1.45	2.87	2.98	1.23	0.5
800	2	3.12	3.46	1.55	0.8

The TPR of the proposed AI-RNN-WO is compared to that of various techniques, including DBN, YOLO-V4, VFFLC, and YOLO-V5. In addition, for specific datasets, the validation DBN achieved a TPR rate of 94.7%, the YOLO-V4 is getting 98.2%, and the VFFLC achieved a TPR rate of 89.5%. Furthermore, the TPR percentage for the YOLO-V5 is 99%. However, the proposed approach has a 99.88% TPR rate. The comparison between the prior models' TPR and the earlier techniques' TPR is shown in Fig. 6.

5.1.3 False-positive rate (FPR)

It is described as the bad outcomes that are misclassified as positive. The AI-RNN-WO algorithm then showed a lower rate of false positives. Our suggested AI-RNN-WO technique surpassed all prior clones with a maximum FPR in the collected datasets. The FPR is calculated using the following Eq. (12). The comparison findings of FPR are displayed in Table 5:

$$\mathrm{FPR} = \frac{\hat{F}\,\hat{P}}{\hat{T}\,\hat{P} + \hat{F}\,\hat{P}} \tag{12}$$

The obtained FPR of the proposed AI-RNN-WO is compared to that of various currently used approaches, including DBN, YOLO-V4, VFFLC, and YOLO-V5. In addition, for specific datasets, the validation DBN achieved an FPR of 3.9%, YOLO-V5 achieved an FPR of 1.8%, YOLO-V4 obtained a 7.5% FPR, and VFFLC achieved an FPR of 2.63%. However, the proposed technique has a 1.2% FPR, as illustrated in Fig. 7.

5.1.4 Error rate

It refers to the data that contained errors from all the data gathered. The achieved error rate of the suggested AI-RNN-WO is compared to that of various currently used approaches, including DBN, YOLO-V4, VFFLC, and YOLO-V5. The comparison findings of error rate are displayed in Table 6.

In addition, the validation of the DBN model obtained an error rate of 0.38%; the YOLO model reached an error rate of 1.56%. The VFFLC model received an error rate of 1.58%, and the YOLO-V5 model attained an error rate of 0.84% for specific datasets. However, the proposed technique has a 0.1% error rate, as illustrated in Fig. 8.

5.1.5 Precision

Precision is the quantity of information a number conveys concerning its digits. The degree of precision demonstrates how closely two or more measures adhere. In addition, the proportion of relevant data between the retrieved data is

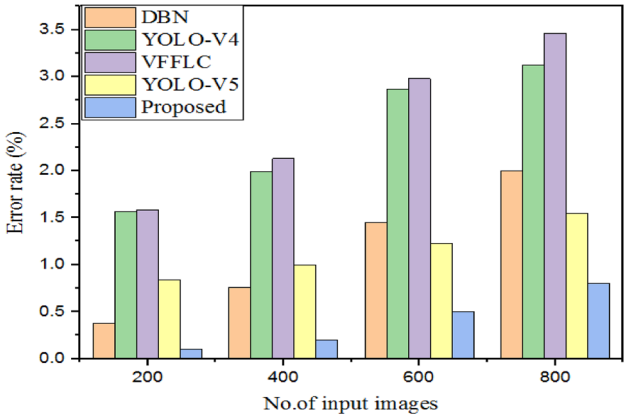


Fig. 8 Error rate comparison

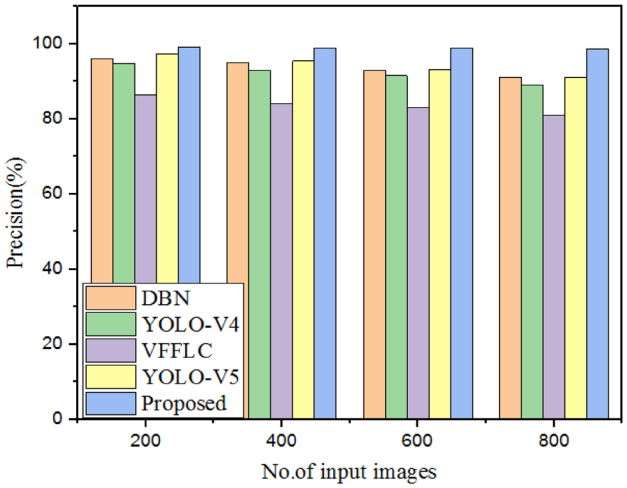


Fig. 9 Precision comparison

Table 7 Precision comparison

No. of input images	Precision (%)				
	DBN	YOLO-V4	VFFLC	YOLO-V5	Proposed
200	96	94.8	86.32	97.43	99.12
400	95.02	93	84	95.46	99.02
600	93	91.54	83.12	93.3	99
800	91.11	89	81	91.04	98.75

known as the precision value. Precision measures the percentage of instances or samples accurately classified among those classed as positives. Moreover, the precision is measured using Eq. (13). The comparison findings of precision are displayed in Table 7:

$$P = \frac{\hat{T}\,\hat{P}}{\hat{F}\,\hat{P} + \hat{T}\,\hat{P}} \tag{13}$$

The suggested AI-RNN-WO model precision rate contrasts with several approaches, including DBN, YOLO-V4, VFFLC, and YOLO-V5. In addition, for specific datasets, the YOLO-V4 is getting 94.8%, the validation DBN acquired a precision rate of 96%, and the VFFLC achieved a precision rate of 86.32%. In addition, the YOLO-V5's precision rate is 97.43%. The proposed method, however, has a 99.12% accuracy rate. Figure 9 compares the precision of previous models with earlier methodologies.

5.1.6 F-measure

Recall and precision issues are balanced in a single number by the F1-score. *F*-Measure offers a method for combining recall and precision into a single instance that encompasses both characteristics. The measurement of F-measure

Table 8 *F*-measure comparison

No. of input images	*F*-measure (%)				
	DBN	YOLO-V4	VFFLC	YOLO-V5	Proposed
200	95	93.21	84.3	96.43	99
400	94.12	91	82.3	95	98.75
600	92.16	90.23	81	94.32	98.23
800	90	88.21	80.23	92.4	98

is obtained using Eq. (14). The comparison findings of the *F*-measure are displayed in Table 8:

$$F\text{ - measure} = \frac{\hat{T}\,\hat{P}}{\hat{T}\,\hat{P} + \frac{1}{2}\left(\hat{F}\,\hat{P} + \hat{F}\,\hat{N}\right)} \tag{14}$$

The suggested AI-RNN-WO F-measure is contrasted with some approaches, including DBN, YOLO-V4, VFFLC, and YOLO-V5. In addition, for specific datasets, the YOLO-V4 is getting 93.21%, the validation DBN achieved an *F*-measure rate of 95%, and the VFFLC achieved an *F*-measure rate of 84.3%. In addition, the YOLO-*F*-measure V5 percentage is 96.43%. The proposed method, however, has a 99% *F*-measure rate. Figure 10 compares the F-measures of the older models and earlier approaches.

5.2 Discussion

Through the use of the application for investigating, segregating, and gathering the detected data for autonomous direction, the propped design provides a potent interaction module and sensors. Using sensors and IoT, the design provides a uniform response for the early detection of a fire

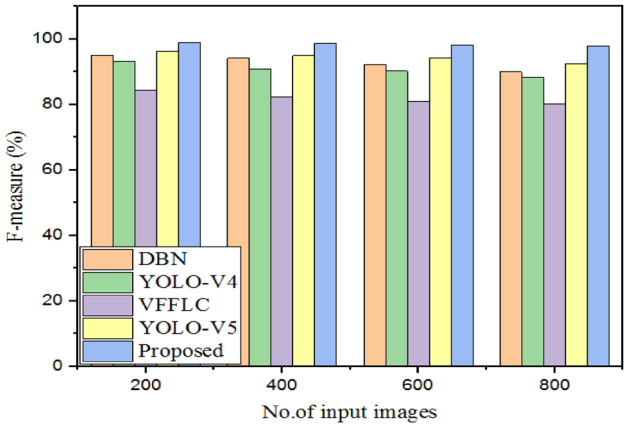


Fig. 10 *F*-measure comparison

event. Thus, the frequency of used smoke and temperature data is shown in Fig. 11.

When applied to concealed units, noise injection can have far more effect than merely decreasing the parameters. Data containing a significant amount of extraneous, meaningless information are called noisy data. This encompasses data corruption; the same word frequently refers to corrupt data. It also includes any information a user system cannot properly comprehend and interpret. An algorithm's robustness is its ability to create models less susceptible to noise and insensitive to data corruption; in other words, the more robust an algorithm is, the more comparable the models it makes from clean and noisy data. Therefore, if one classification method creates classifiers that are less affected by noise than another, it is considered more robust than the other. When working with noisy data, robustness is crucial since it enables one to predict how much the learning method's performance against noise will vary from its noiseless performance when the noise's properties are unknown. The attained simulation results of training accuracy and training loss of the designed model are shown in Fig. 12.

This section describes the AI-RNN-WO model's performance using IoT devices for detection and classification. Here, the Fismo dataset is used; it contains a total of 984 images. Among that, 80% of the data can be used as a training technique, and 20% can be used as a testing procedure. As a result, the designed model gained 99.5% accurate results by attaining 0.53% mini-batch root mean square error (RMSE). In addition, the proposed technique attains a 99.7% precise fire hazard prediction by achieving 0.3% of the mini-batch loss. Comparisons are made between the simulation results for the essential metrics and traditional models. Table 9 shows how the created AI-RNN-WO model outperforms the conventional state-of-the-art techniques regarding accuracy, true-positive rate, false-positive rate, and error rate. The overall performance of the developed model is displayed in Table 9.

The designed model improved performance and the efficiency was validated with several iterations, such as 20, 40, 60, 80 and 100. It attains 99.5% accuracy, 99.88% TPR, 1.2% FPR, 0.1% error rate, 99.12% precision, and 99% *F*-measure for 60 iterations. Moreover, the developed technique efficiently predicts fire hazards with high performance. The developed model gained better experimental results while comparing other techniques, showing the efficiency of the fire detection system and reducing errors. The developed

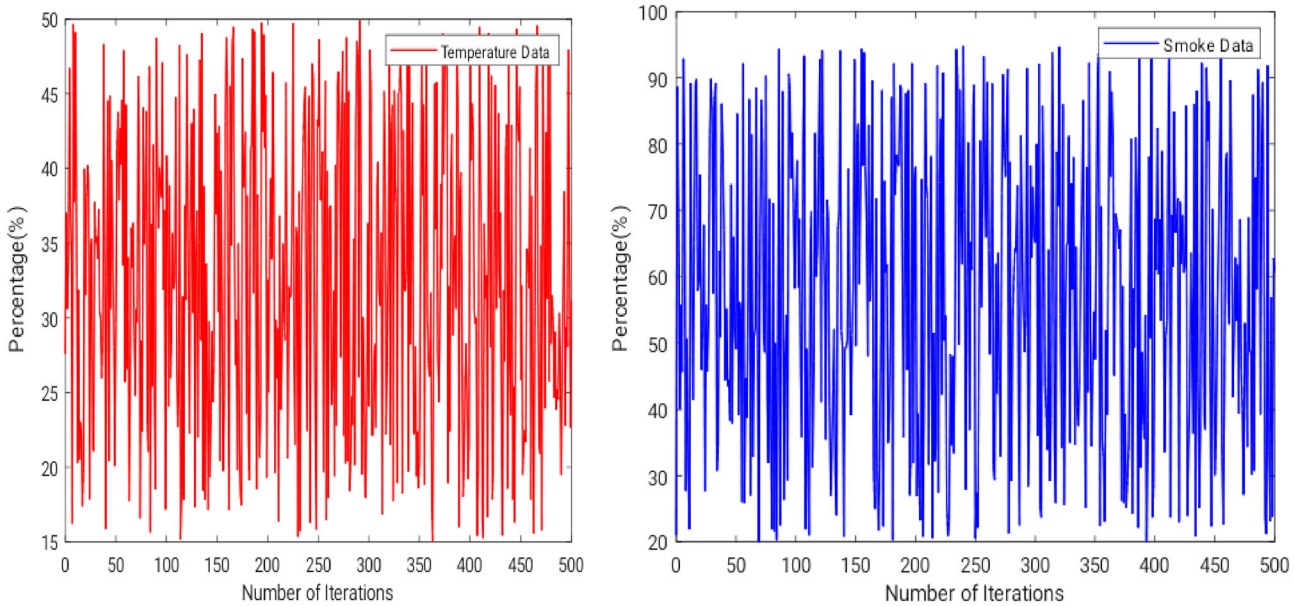


Fig. 11 Smoke and temperature data

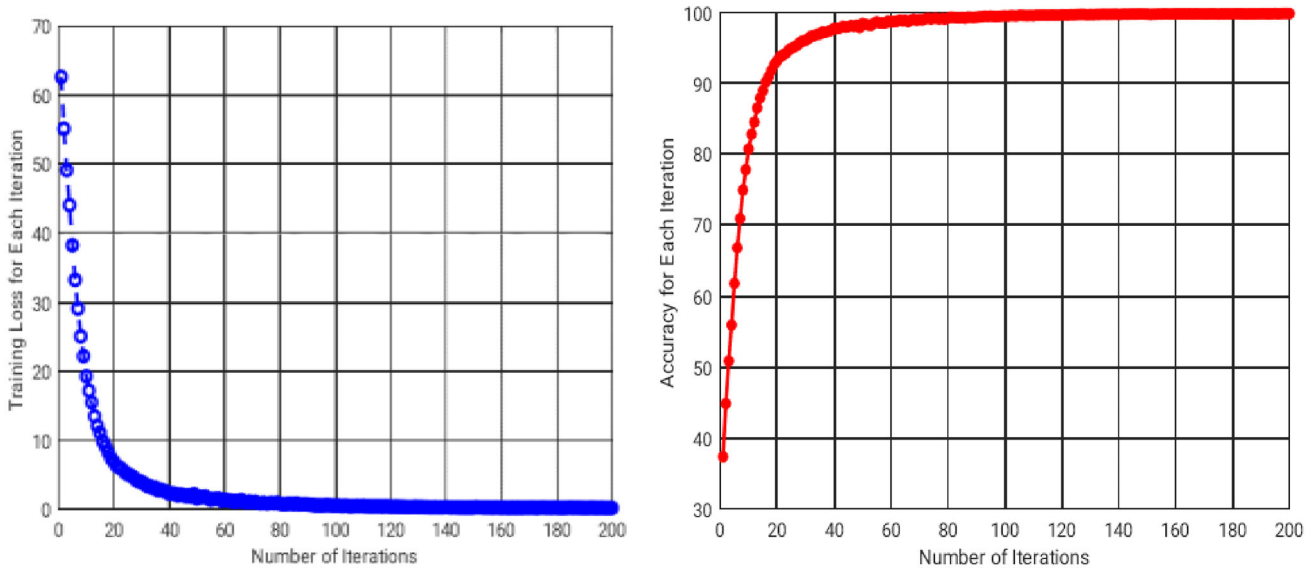


Fig. 12 Training loss vs. training accuracy

Table 9 Overall performance

No. of iteration	Accuracy	TPR	Precision	*F*-measure	FPR	Error rate
20	99.5	99.88	99.12	99	1.2	0.1
40	99.01	99.60	99.08	98.56	1.5	0.2
60	98.69	99.12	98.81	98.34	1.7	0.4
80	98.17	98.94	98.78	98.17	1.9	0.6
100	98.08	98.46	98.04	98.07	2.3	0.8

technique overcomes the issues of noisy data, error, and misclassification and gains better prediction results. It also communicates fire risk using mobile phones. However, the computational cost of the developed model is high, and the communication process is slow because of the standard network; it can be improved using a wireless sensor network system. The created model gathers and examines information from numerous local services. The developed AI-RNN-WO model in "smart cities" can tackle various issues, from traffic to criminality. The proposed algorithm uses information gathered from sensors located across the metropolitan environment to anticipate the risk of fire accurately.

6 Conclusion

These days, fire threats are a significant concern for everyone worldwide. Numerous studies have been offered to foresee the early fire calamity. But it is impossible to anticipate reasonably when there are so many problems. This article builds the AI-RNN-WO model, a revolutionary intelligent early prediction technique. Various sensor devices are set up throughout smart cities to gather real-time data. In addition, the Matlab platform is used for verification and implementation. Further, the suggested model improved prediction accuracy results. Thus, it increases status prediction accuracy by 8–10%.

Author contributions Dr P. Dileep Kumar Reddy, Dr. Martin Margala, Dr. Siva Shankar S, and Dr. Prasun Chakrabarti discussed and constructed the measures, found their applications, and wrote the paper together.

Funding Not applicable.

Availability of data and material Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Human and animal rights This article does not contain any studies with human or animal subjects performed by any of the authors.

Informed consent Informed consent does not apply as this was a retrospective review with no identifying patient information.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

1. Avazov K, Mukhiddinov M, Makhmudov F, Cho YI (2021) Fire detection method in smart city environments using a deep-learning-based approach. Electronics 11(1):73
2. Zhang F, Zhao P, Xu S, Wu Y, Yang X, Zhang Y (2020) Integrating multiple factors to optimize watchtower deployment for wildfire detection. Sci Total Environ 737:139561
3. Barmpoutis P, Dimitropoulos K, Kaza K, Grammalidis N (2019) Fire detection from images using faster R-CNN and multidimensional texture analysis. In: ICASSP 2019–2019 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 8301–8305
4. Valikhujaev Y, Abdusalomov A, Cho YI (2020) Automatic fire and smoke detection method for surveillance systems based on dilated CNNs. Atmosphere 11(11):1241
5. Cao C, Tan X, Huang X, Zhang Y, Luo Z (2021) Study of flame detection based on improved YOLOv4. J Phys Conf Ser 1952(2):022016
6. Kim B, Lee J (2019) A video-based fire detection using deep learning models. Appl Sci 9(14):2862
7. Barmpoutis P, Papaioannou P, Dimitropoulos K, Grammalidis N (2020) A review on early forest fire detection systems using optical remote sensing. Sensors 20(22):6442
8. Zhang Y, Geng P, Sivaparthipan CB, Muthu BA (2021) Big data and artificial intelligence based early risk warning system of fire hazard for smart cities. Sustain Energy Technol Assess 45:100986
9. Costa DG, Peixoto JPJ, Jesus TC, Portugal P, Vasques F, Rangel E, Peixoto M (2022) A survey of emergencies management systems in smart cities. IEEE Access 10:61843–61872
10. Peixoto M (2022) A survey of emergencies management systems in smart cities. IEEE Access 10:61843–61872
11. Oliveira F, Costa DG, Assis F (2022) An IOT platform for the development of low-cost emergencies detection units based on soft sensors. In: 2022 IEEE international smart cities conference (ISC2). IEEE, pp 1–4
12. Mendle RS, Hartung A (2022) Wielding a concept with two edges: how to make use of the smart cities concept and understanding its risks from the resilient cities perspective. Resilient smart cities: theoretical and empirical insights. Springer International Publishing, Cham, pp 375–394
13. Zhu L, Li M, Metawa N (2021) Financial risk evaluation Z-score model for intelligent IoT-based enterprises. Inf Process Manage 58(6):102692
14. Mazur-Milecka M, Głowacka N, Kaczmarek M, Bujnowski A, Kaszyński M, Rumiński J (2021) Smart city and fire detection using thermal imaging. In: 2021 14th International conference on human system interaction (HSI). IEEE, pp 1–7
15. Sharma S, Chmaj G, Selvaraj H (2022) Machine learning applied to internet of things applications: a survey. In: Advances in systems engineering: proceedings of the 28th international conference on systems engineering, ICSEng 2021, December 14–16, Wrocław, Poland. Springer International Publishing, pp 301–309
16. Tarar S, Bhasin N (2021) Fire hazard detection and prediction by machine learning techniques in smart buildings (SBs) using sensors and unmanned aerial vehicles (UAVs). In: Solanki A, Kumar A, Nayyar A (eds) Digital cities roadmap: IoT-based architecture and sustainable buildings. Wiley, New Jersey, pp 63–95
17. Stokkenes S, Strand RD, Kristensen LM, Log T (2021) Validation of a predictive fire risk indication model using cloud-based weather data services. Proc Comput Sci 184:186–193
18. Ullah F, Qayyum S, Thaheem MJ, Al-Turjman F, Sepasgozar SME (2021) Risk management in sustainable smart cities governance: a TOE framework. Technol Forecast Soc Change 167:120743
19. Taufik M, Widyastuti MT, Sulaiman A, Murdiyarso D, Santikayasa IP, Minasny B (2022) An improved drought-fire assessment for managing fire risks in tropical peatlands. Agric Forest Meteorol 312:108738
20. Fedele R, Merenda M (2020) An IoT system for social distancing and emergency management in smart cities using multi-sensor data. Algorithms 13(10):254
21. Calp MH, Butuner R, Kose U, Alamri A, Camacho D (2022) IoHT-based deep learning controlled robot vehicle for paralyzed patients of smart cities. J Supercomput 78(9):11373–11408
22. Motta M, de Castro NM, Sarmento P (2021) A mixed approach for urban flood prediction using machine learning and GIS. Int J Disaster Risk Reduct 56:102154
23. Reddy P, Kumar D, Sam RP, Bindu CS (2016) Optimal blowfish algorithm-based technique for data security in cloud. Int J Bus Intell Data Min 11(2):171–189
24. Ramu G, Reddy PDK, Jayanthi A (2018) A survey of precision medicine strategy using cognitive computing. Int J Mach Learn Comput 8(6):530–535
25. Somasekar J, Ramesh G, Ramu G, Reddy PDK, Reddy BE, Lai C-H (2019) A dataset for automatic contrast enhancement of microscopic malaria infected blood RGB images. Data Brief 27:104643
26. Jin G, Zhu C, Chen X, Sha H, Hu X, Huang J (2020) Ufsp-net: a neural network with spatio-temporal information fusion for urban fire situation prediction. IOP Conf Ser Mater Sci Eng 853(1):012050
27. Hassouneh Y, Turabieh H, Thaher T, Tumar I, Chantar H, Too J (2021) Boosted whale optimization algorithm with natural selection operators for software fault prediction. IEEE Access 9:14239–14258

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Machine Learning for Characterization and Analysis of Microstructure and Spectral Data of Materials

[1]Venkataramaiah Gude, [2]Dr Sujeeth T, [3]Dr. K Sree Divya, [4]P. Dileep Kumar Reddy, [5]G. Ramesh

Submitted: 09/12/2023 **Revised**: 18/01/2024 **Accepted**: 02/02/2024

Abstract: In the contemporary world, there is lot of research going on in creating novel nano materials that are essential for many industries including electronic chips and storage devices in cloud to mention few. At the same time, there is emergence of usage of machine learning (ML) for solving problems in different industries such as manufacturing, physics and chemical engineering. ML has potential to solve many real world problems with its ability to learn in either supervised or unsupervised means. It is inferred from the state of the art that that it is essential to use ML methods for analysing imagery of nano materials so as to ascertain facts further towards characterization and analysis of microstructure and spectral data of materials. Towards this end, in this paper, we proposed a ML based methodology for STEM image analysis and spectral data analysis from STEM image of a nano material. We proposed an algorithm named Machine Learning for STEM Image Analysis (ML-SIA) for analysing STEM image of a nano material. We proposed another algorithm named Machine Learning for STEM Image Spectral Data Analysis (ML-SISDA) for analysing spectral data of STEM image of a nano material. We developed a prototype ML application to implement the algorithms and evaluate the proposed methodology. Experimental results revealed that the ML based approaches are useful for characterization of nano materials. Thus this research helps in taking this forward by triggering further work in the area of material analysis with artificial intelligence.

Subject Classification: *68U10*

Keywords: *Nano Material Characterization, STEM Analysis, Spectral Data Analysis, Machine Learning, Microstructure Analysis*

1. Introduction

With respect to growing research in nano materials and their characterization, there is increasing role of machine learning and artificial intelligence (AI). It is indispensable in the research of nano materials to understand and use AI based approaches for improving the designs and manufacturing of such materials. In this context, exploration of microstructures associated with nano-materials plays crucial role. At the same time ML based approaches are widely used to solve the problems in different domains. In manufacturing and many industries AI is being used for improving accuracy and quality in the designs. The role of ML is increasing in the study of new designs and implementation of nano materials [1].

[1]*Software Engineer, GP Technologies LLC, U.S.A.*
[2]*Department of Computer science and Engineering, Siddhartha Educational Academy Group of Institutions, Tirupati, Andhra Pradesh, India.*
[3]*Department of Computer Science & Technology, Madanapalle Institute of Technology and Science Madanapalle. Andhra Pradesh, India*
[4]*Department of Computer Science and Engineering, Narsimha Reddy Engineering College (Autonomous), Secunderabad, Telangana State, India.*
[5]*Department of Computer Science and Engineering, Gokaraju Rangaraju Institute of Engineering & Technology, Hyderabad, Telangana State, India*
ramesh680@gmail.com, gvramaiah.se@gmail.com, divya.kpn@gmail.com, sujeeth.2304@gmail.com, dileepreddy503@gmail.com
Corresponding Author: ramesh680@gmail.com

There are many existing methods that used ML approaches to leverage design and characterization of nano materials. In [4], [6], [7] and [8] machine learning models are used for characterization of complex materials. Chan et al. [4] focused on 3D sample characterization of autonomous microstructures with the help of machine learning. Holm et al. [6] studied the importance of machine learning for understanding and characterization of microstructures of materials. Xu et al. [7] explored machine learning and predictive modelling in order to have microscoping imaging. Their study helped in understanding microstructures pertaining Li-Ion batteries. Joshua et al. [8] considered complex materials in order to have data-driven machine learning to characterize surface microstructures. Infrared spectroscopy is used in their empirical study. There are certain studies where deep learning is used for studying materials. Lin et al. [3] investigated on the reconstruction and characterization of materials with the help of deep learning methods and their applications. Pokuri et al. [5] used a guided approach with interpretable deep learning for exploring microstructural properties associated with photovoltaics. Dimiduk et al. [13] explored different materials to ascertain the utility of ML and deep learning techniques. Their investigation caters to manufacturing innovation, material integration, processes, materials and structural engineering.

From the literature, it is inferred that it is essential to use ML methods for analysing imagery of nano materials so

as to ascertain facts further towards characterization and analysis of microstructure and spectral data of materials. Our contributions in this paper are as follows. We proposed a ML based methodology for STEM image analysis and spectral data analysis from STEM image of a nano material. We proposed an algorithm named Machine Learning for STEM Image Analysis (ML-SIA) for analysing STEM image of a nano material. We proposed another algorithm named Machine Learning for STEM Image Spectral Data Analysis (ML-SISDA) for analysing spectral data of STEM image of a nano material. We developed a prototype ML application to implement the algorithms and evaluate the proposed methodology. The remainder of the paper is structured as follows. Section 2 reviews literature on methods of characterization and analysis of microstructure and spectral data of materials. Section 3 proposes a methodology for STEM image analysis and spectral data analysis from STEM image of a nano material. Section 4 presents results and discussion. Section 5 concludes on the results of the analysis and give scope for future work.

2. Related Work

This section reviews literature on different methods useful for material analysis. Dongdong et al. [1] focused on the process of characterization of coatings of thermal barrier with respect to microstructural features. They used a technology known as terahertz spectroscopy. Ramin et al. [2] focused on different techniques required for characterization of computational microstructures. Lin et al. [3] investigated on the reconstruction and characterization of materials with the help of deep learning methods and their applications. Chan et al. [4] focused on 3D sample characterization of autonomous microstructures with the help of machine learning. Pokuri et al. [5] used a guided approach with interpretable deep learning for exploring microstructural properties associated with photovoltaics. Holm et al. [6] studied the importance of machine learning for understanding and characterization of microstructures of materials. Xu et al. [7] explored machine learning and predictive modelling in order to have microscoping imaging. Their study helped in understanding microstructures pertaining Li-Ion batteries. Joshua et al. [8] considered complex materials in order to have data-driven machine learning to characterize surface microstructures. Infrared spectroscopy is used in their empirical study. Iquebal et al. [9] investigated on lithography process that is nanoindentation based. Their study is on acoustic emission signatures. Their empirical study was for characterization of rapid microstructures. Chowdhury et al. [10] used ML with images in order to understand and reconstruction of microstructures.

Du et al. [11] explored ML along with robotics for understanding the potential of materials pertaining to OPV. Pilania [12] used ML for different purposes that range from predictions of characteristics to autonomous design of materials. Dimiduk et al. [13] explored different materials to ascertain the utility of ML and deep learning techniques. Their investigation caters to manufacturing innovation, material integration, processes, materials and structural engineering. Ren et al. [14] proposed a protocol for vibrational spectroscopy with machine learning for understanding spectrum based structure and spectrum prediction. DeCost et al. [15] focused on the characterization of AM powder with the help of ML and computer vision. Konstantopoulos et al. [16] performed different kinds of testing on properties of materials for microstructure identification using ML approaches. Ruoqian et al. [17] focused on the composite microstructures that are of three-dimensional in nature. They used machine learning models that are context aware and the modelling of elastic localization is carried out.

Akshay et al. [18] explored computational materials in order to understand microstructures and they used spectral density function for empirical observations. Ruoqian et al. [19] proposed methodology using ML approaches for possible materials design and microstructure optimization. Xu et al. [20] proposed a design and representation approach based on ML for making microstructures with heterogeneity. Chen et al. [21] explored photoacoustic spectroscopy through machine learning for prostate cancer identification. From the literature, it is inferred that it is essential to use ML methods for analysing imagery of nano materials so as to ascertain facts further towards characterization and analysis of microstructure and spectral data of materials.

3. Methodology

We proposed a methodology based on methods of machine learning for analysing STEM images and spectral data of such materials. It is understood that ML plays crucial for understanding materials and characterization. STEM image of nano material is subjected to two kinds of procedures for characterization of material. This kind of study paves way for manufacturing novel nano materials and understanding the spectral data of materials. The given STEM image is subjected to ML based methods for analysis. The input is also subjected to spectral data analysis through decomposition in order to characterize materials. Figure 1 shows an overview of the proposed methodology for characterization of materials through ML methods.

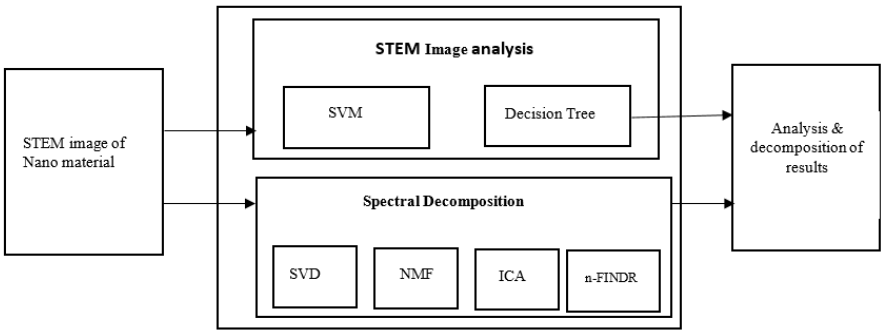


Fig 1: Overview of the proposed methodology used for STEM image analysis and spectral decomposition

The given STEM image is subjected to STEM image analysis with the help of ML models such as Support Vector Machine (SVM) and Decision Tree (DT). Afterwards, the spectral data analysis is carried out using Singular Value Decomposition (SVD), n-FINDR, Independent Component Analysis (ICA) and non-negative matrix factorization (NMF). SVM is one of the widely used ML classifier used for the study of materials. SVM is the binary classification algorithm which can divide the feature space into two classes based on its hyperplane criteria. Hyperplane approach can be expressed as in Eq. 1.

H:$w^T(x) + b = 0$ (1)

The modus operandi of the SVM also needs a distance measure that is useful in determining classes in the process of STEM image analysis. The distance measure is expressed in Eq. 2.

D=$\frac{|ax_0+by_0+c|}{\sqrt{a^2+b^2}}$ (2)

There is also need for finding distance from given point to hyperplane equation in order to make the decisions while analysing STEM images. This distance computation is carried out as in Eq. 3.

$d_h(\emptyset(x_0))=\frac{|w^T(\emptyset(x_0))+b|}{||w||_2}$ (3)

Apart from SVM, DT is another important classifier used to analyse STEM image of a nano material. The decision tree classifier is based on identification of attribute or feature for splitting data while making decision tree. It is based on two measures known as entropy and gain as expressed in Eq. 4 and Eq. 5.

Entropy(S)=$\sum_{i=1}^{c} -p_i log_2 p_i$ (4)

Gain(S, A)= Entropy(S)- $\sum_{\vartheta \epsilon Values(A)} \frac{|S_\vartheta|}{|S|}$ Entropy(S_ϑ) (5)

Both entropy and gain measures are used for making decisions while modelling data using DT classifier. With respect to spectral data analysis, the given STEM image is subjected to spectral decomposition that is carried out using different methods such as Singular Value Decomposition (SVD), n-FINDR, Independent Component Analysis (ICA) and non-negative matrix factorization (NMF). The Singular Value Decomposition (SVD) is one of the methods used for analysing given material. SVD of a matrix is nothing but factorization of different matrices. It is used to explore linear transformations with theoretical and geometrical insights. In the domain of data science SVD has important applications due to its intuition in acquiring geometrical meaning. The SVD analysis can be expressed as in Eq. 6.

$C_{m\times n} = U_{m\times r} \times \sum_{r\times r.} \times V^1_{r\times n}$ (6)

Non-Negative Matrix Factorization is another technique used for decomposition of materials. For a given matrix NMF considers only non-negative elements. Considering W and H are two matrices, and A is original input matrix, the process of NMF is expressed as in Eq. 6.

$A_{m\times n} = W_{m\times k} H_{k\times n}$ (7)

Independent Component Analysis (ICA) is another ML technique widely used for separating different independent sources available in a given mixed signal. It focuses on independent components unlike principal component analysis. Considering there are different sources, ICA is computed as in Eq. 8.

$[X_1, X_2, \dots\dots, X_n] \Rightarrow [Y_1, Y_2, \dots\dots, Y_n]$ (8)

Another method of ML known as n-FINDR is used for spectral data analysis. It is widely used method for discovering endmembers from hyperspectral imagery. It exploits all p-endmember combinations in order to arrive at the desired results.

Algorithm Design

We proposed an algorithm named Machine Learning for STEM Image Analysis (ML-SIA) for analysing STEM image of a nano material. We proposed another algorithm named Machine Learning for STEM Image Spectral Data Analysis (ML-SISDA) for analysing spectral data of STEM image of a nano material. We developed a prototype ML application to implement the algorithms and evaluate the proposed methodology.

Algorithm: STEM Image Analysis (ML-SIA)

Inputs:

STEM image of an oxide catalyst ***I***

Machine learning classifier pipeline ***P***

(P refers to a collection of classifiers consisting of SVM and DT)

Training data ***T***

Output:

Results of STEM image analysis ***R***

1. Start
2. F←ExtractFeatures(***I***)
3. For each ML technique t in P
4. model←TrainClassifier(T)
5. ***R***←FitModel(M,F, I)
6. Display ***R***
7. End For
8. End

Algorithm 1: STEM Image Analysis (ML-SIA)

As presented in Algorithm 1, the given input STEM image is subjected to classification using different ML models such as SVM and DT. The pipeline of these models, training data T and STEM image of an oxide catalyst are inputs to the algorithm. It has an iterative process in order to perform STEM image analysis and provide results. Machine Learning for STEM Image Spectral Data Analysis (ML-SISDA) is another algorithm proposed.

Algorithm: Machine Learning for STEM Image Spectral Data Analysis (ML-SISDA)
Inputs:
STEM image of an oxide catalyst ***I***
Machine learning models pipeline ***P***
(P refers to a collection of ML models such as Singular Value Decomposition (SVD), n-FINDR, Independent Component Analysis (ICA) and non-negative matrix factorization (NMF))
Output:
Results of STEM image spectral data analysis ***R***

1. Start
2. Spectral←Extract(***I***)
3. For each ML model m in P
4. r←SpetralDataAnalysis(spectral)
5. Add r to ***R***
6. End For
7. Display R
8. End

Algorithm 2: Machine Learning for STEM Image Spectral Data Analysis (ML-SISDA)

As presented in Algorithm 2, it takes STEM image of an oxide catalyst and machine learning models pipeline as input and returns the results of STEM image spectral data analysis. It extracts spectral features of given input image. Then it has an iterative process to analyse spectral data in different aspects and generate visual representations of spectral data for characterization of nano materials. This is the ML approach that has potential to improve the state of the art.

4. Results and Discussion

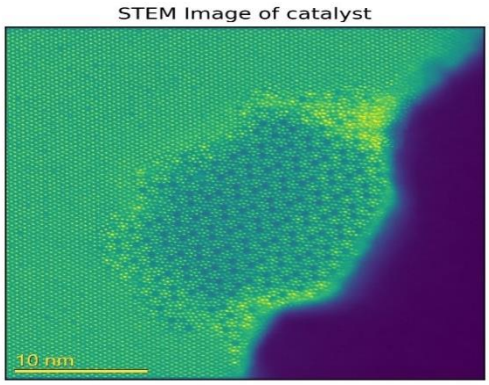
This section presents experimental results of the proposed methodology that covers STEM image analysis and spectral data analysis from STEM image of a nano material. The proposed algorithms are evaluated with an input image which is nothing but a STEM image captured from a nano material pertaining to an oxide catalyst. Different ML techniques are used for the empirical study. They include SVM, DT, SVD, NMF, ICA and n-FINDR.

The results are divided into the results of STEM image analysis and STEM spectral data analysis.

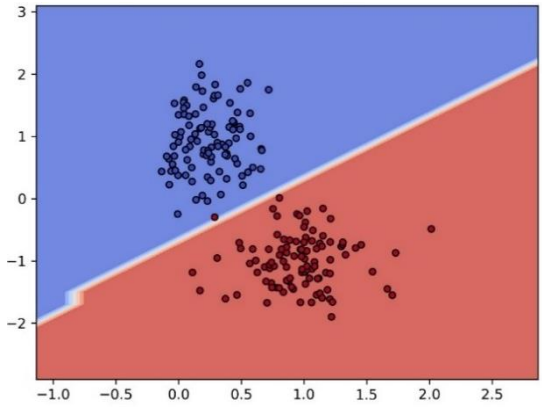
4.1 Results STEM Image Analysis

Two popular ML algorithms known as SVM and DT are used to characterize and analyse a STEM image captured from a nano material pertaining to an oxide catalyst. The two classification models do have different modus operand in analysing and coming up with required classes. Two classes are used for the analysis.

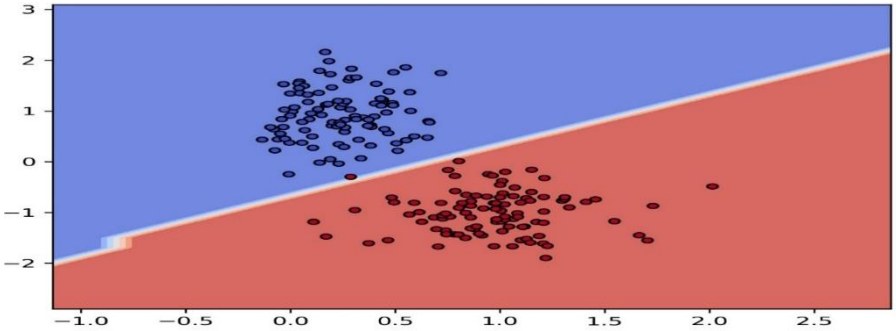
As presented in Figure 2 (a), it reflects a STEM image of an oxide catalyst. Scanning transmission electron microscopy imaging technology is used to know the characteristics of nano materials. Figure 2 (b) the result of SVM classification of STEM image is provided. It is binary classification that has resulted in two classes based on SVM's approach in hyperplane for classification. Figure 2 (c) shows classification result performed by Decision Tree algorithm with two classes. This classification is based on the ability of DT algorithm in identifying attribute for splitting.



(a) Input STEM image



(b) Result of analysis with SVM



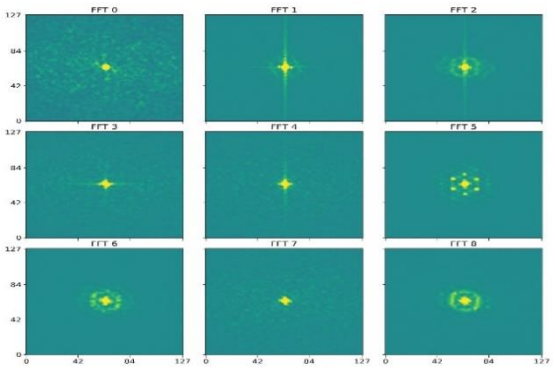
(c) Result of analysis with DT

Fig 2: STEM image analysis

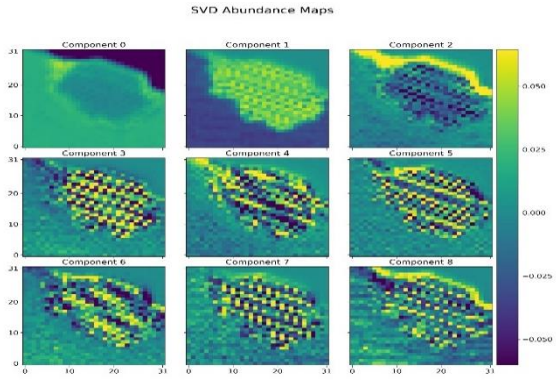
4.2 Results Spectral Data Analysis

Different methods are used for spectral data analysis provided a STEM image as input. The methods used for the empirical study are SVD, NMF, ICA and n-FINDR.

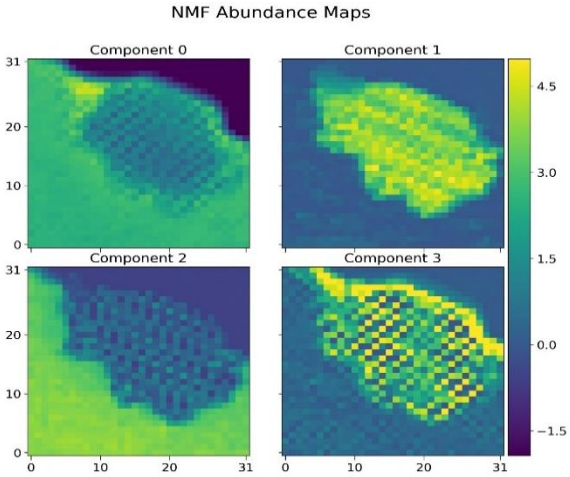
Figure 3 (a) shows FFT windows are used based on sliding window approach for performing spectral data analysis. Figure 3 (b) shows SVD abundance maps in order to analyse spectral data. It shows different components reflecting visualization of SVD abundance maps that are used to ascertain the spectral data associated with given STEM. Figure 3 (c) shows NMF abundance maps are visualized with the help of NMF coefficients. Figure 3 (d) shows that NMF components are used to visualize then and it can be used for effective spectral data analysis.



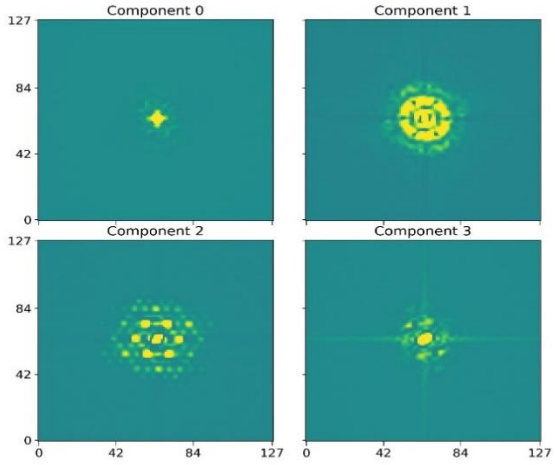
(a) Spectral data analysis in terms of FET



(b) **Spectral data analysis in terms of SVD abundance maps**



(c) Spectral data analysis in terms of NMF abundance maps



(d) **Spectral data analysis in terms of components**

Fig 3: Spectral data analysis

5. Conclusion and Future Work

In this paper, we proposed a ML based methodology for STEM image analysis and spectral data analysis from STEM image of a nano material. We proposed an algorithm named Machine Learning for STEM Image Analysis (ML-SIA) for analysing STEM image of a nano material. We proposed another algorithm named Machine Learning for STEM Image Spectral Data Analysis (ML-SISDA) for analysing spectral data of STEM image of a nano material. We developed a prototype ML application to implement the algorithms and evaluate the proposed methodology. Experimental results revealed that the ML based approaches are useful for characterization of nano materials. Thus this research helps in taking this forward by triggering further work in the area of material analysis with artificial intelligence. In future, we intend to explore deep learning models for nano material analysis for improving the benefits of characterization of nano materials. This will help in many industries such as manufacturing of storage devices to improve quality and productivity.

References

[1] Ye, Dongdong; Wang, Weize; Zhou, Haiting; Fang, Huanjie; Huang, Jibo; Li, Yuanjun; Gong, Hanhong; Li, Zhen. Characterization of thermal barrier coatings microstructural features using terahertz spectroscopy. Surface and Coatings Technology, 394, p1-10, (2020).

[2] Bostanabad, Ramin; Zhang, Yichi; Li, Xiaolin; Kearney, Tucker; Catherine Brinson, L.; Apley, Daniel W.; Liu, Wing Kam; Chen, Wei. Computational Microstructure Characterization and Reconstruction: Review of the State-of-the-art Techniques. Progress in Materials Science, p1-123, (2018).

[3] Lin, J., Chen, S., Wang, W., Pathirage, C. S. N., Li, L., Sagoe-Crentsil, K., & Duan, W. Transregional spatial correlation revealed by deep learning and implications for material characterisation and reconstruction. Materials Characterization, 178, 111268. P1-12, (2021).

[4] Chan, Henry; Cherukara, Mathew; Loeffler, Troy D.; Narayanan, Badri; Sankaranarayanan, Subramanian K. R. S. Machine learning enabled autonomous microstructural characterization in 3D

samples. npj Computational Materials, 6(1), p1-9, (2020).

[5] Pokuri, Balaji Sesha Sarath; Ghosal, Sambuddha; Kokate, Apurva; Sarkar, Soumik; Ganapathysubramanian, Baskar. Interpretable deep learning for guided microstructure-property explorations in photovoltaics. npj Computational Materials, 5(1), p1-11, (2019).

[6] Holm, Elizabeth A.; Cohn, Ryan; Gao, Nan; Kitahara, Andrew R.; Matson, Thomas P.; Lei, Bo; Yarasi, Srujana Rao. Overview: Computer Vision and Machine Learning for Microstructural Characterization and Analysis. Metallurgical and Materials Transactions A, p1-15, (2020).

[7] Hongyi Xu;Juner Zhu;Donal P. Finegan;Hongbo Zhao;Xuekun Lu;Wei Li;Nathaniel Hoffman;Antonio Bertei;Paul Shearing;Martin Z. Bazant. Guiding the Design of Heterogeneous Electrode Microstructures for Li-Ion Batteries: Microscopic Imaging, Predictive Modeling, and Machine Learning . Advanced Energy Materials, p1-34, (2021).

[8] Lansford, Joshua L.; Vlachos, Dionisios G. Infrared spectroscopy data- and physics-driven machine learning for characterizing surface microstructure of complex materials. Nature Communications, 11(1), 1513, P1-12, (2020).

[9] Ashif Sikandar Iquebal a,* , Shirish Pandagare a , Satish Bukkapatnam. Learning acoustic emission signatures from a nanoindentation-based lithography process: Towards rapid microstructure characterization. Elsevier, p1-8, (2020).

[10] Chowdhury, Aritra; Kautz, Elizabeth; Yener, Bülent; Lewis, Daniel. Image driven machine learning methods for microstructure recognition. Computational Materials Science, 123, p176–187, (2016).

[11] Xiaoyan Du;Larry Lüer;Thomas Heumueller;Jerrit Wagner;Christian Berger;Tobias Osterrieder;Jonas Wortmann;Stefan Langner;Uyxing Vongsaysy;Melanie Bertrand;Ning Li;Tobias Stubhan;Jens Hauch;Christoph J. Brabec. Elucidating the Full Potential of OPV Materials Utilizing a High-Throughput Robot-Based Platform and Machine Learning . Joule, p1-13, (2021).

[12] Ghanshyam Pilania; Machine learning in materials science: From explainable predictions to autonomous design . Computational Materials Science, p1-13, (2021).

[13] Dimiduk, Dennis M.; Holm, Elizabeth A.; Niezgoda, Stephen R. Perspectives on the Impact of Machine Learning, Deep Learning, and Artificial Intelligence on Materials, Processes, and Structures Engineering. Integrating Materials and Manufacturing Innovation, p1-16, (2018).

[14] Hao Ren;Hao Li;Qian Zhang;Lijun Liang;Wenyue Guo;Fang Huang;Yi Luo;Jun Jiang; A machine learning vibrational spectroscopy protocol for spectrum prediction and spectrum-based structure recognition . Fundamental Research, p1-7, (2021).

[15] DeCost, Brian L.; Jain, Harshvardhan; Rollett, Anthony D.; Holm, Elizabeth A. Computer Vision and Machine Learning for Autonomous Characterization of AM Powder Feedstocks. JOM, 69(3), p456–465, (2017).

[16] Konstantopoulos, Georgios; Koumoulos, Elias P.; Charitidis, Costas A. Testing Novel Portland Cement Formulations with Carbon Nanotubes and Intrinsic Properties Revelation: Nanoindentation Analysis with Machine Learning on Microstructure Identification. Nanomaterials, 10(4), p1-26, (2020).

[17] Liu, Ruoqian; Yabansu, Yuksel C.; Yang, Zijiang; Choudhary, Alok N.; Kalidindi, Surya R.; Agrawal, Ankit. Context Aware Machine Learning Approaches for Modeling Elastic Localization in Three-Dimensional Composite Microstructures. Integrating Materials and Manufacturing Innovation, 6(2), p160–171, (2017).

[18] Iyer, Akshay; Dulal, Rabindra; Zhang, Yichi; Ghumman, Umar Farooq; Chien, TeYu; Balasubramanian, Ganesh; Chen, Wei. Designing anisotropic microstructures with spectral density function. Computational Materials Science, 179, p1-8, (2020).

[19] Liu, Ruoqian; Kumar, Abhishek; Chen, Zhengzhang; Agrawal, Ankit; Sundararaghavan, Veera; Choudhary, Alok. A predictive machine learning approach for microstructure optimization and materials design. Scientific Reports, 5, p1-12, (2015).

[20] Xu, H., Liu, R., Choudhary, A., & Chen, W. A Machine Learning-Based Design Representation Method for Designing Heterogeneous Microstructures. Volume 2B: 40th Design Automation Conference. P1-12, (2014).

[21] Yingna Chen;Chengdang Xu;Zhaoyu Zhang;Anqi Zhu;Xixi Xu;Jing Pan;Ying Liu;Denglong Wu;Shengsong Huang;Qian Cheng. Prostate cancer identification via photoacoustic spectroscopy and machine learning. Photoacoustics,p1-12, (2021).

PAPER

Influence of heat treatment on microstructure and mechanical properties of pulsed Nd: YAG laser welded dissimilar sheets of Hastelloy C-276 and monel 400

G Shanthos Kumar[5], V Sivamaran, R Lokanadham, C Raju and T K Mandal[5]

Published 17 August 2023 • © 2023 IOP Publishing Ltd

Physica Scripta, Volume 98, Number 9

Citation G Shanthos Kumar *et al* 2023 *Phys. Scr.* **98** 095933

DOI 10.1088/1402-4896/aceb3b

Authors ▲ References ▼ Open science ▼

R Lokanadham

AFFILIATIONS

Professor in Mechanical Engineering Narasimhareddy Engineering College, Hyderabad, India

EMAIL

lokanadham66@gmail.com