Chronic Disease Prediction System Using Machine Learning

Darshan Gowda S [1], Prof. K Sharath [2]

[1] Student, Department of MCA, Bangalore Institute of Technology, Karnataka, India

[2] Professor, Department of MCA, Bangalore Institute of Technology, Karnataka, India

ABSTRACT

The Chronic Disease Prediction System is an AI-powered web platform developed to enable early and accurate screening for pneumonia, stroke, and diabetes in a single unified environment. Leveraging deep learning and machine learning techniques, it integrates a convolutional neural network for chest X-ray analysis, logistic regression for stroke assessment, and a random forest classifier for diabetes detection, thereby accommodating both image and tabular clinical data. The project features robust user authentication and personalized dashboards, allowing secure access and tracking of results. For pneumonia detection, the system not only yields a probability score but also visualizes affected lung regions using Grad- CAM heatmaps, supporting rapid triage and clinical interpretation. Stroke and diabetes modules accept structured health inputs, offering risk classifications with pie chart visualizations and direct links to medical consultations. Its scalable architecture, built on Flask (backend) and React (frontend), is designed for cross-platform accessibility and future extensibility, enabling integration with additional diseases, cloud deployment, and advanced explainability features like SHAP and LIME. By combining rapid inference, consistency, user-centric design, and health professional integration, the system advances preventive medicine, offering considerable impact in cost reduction, accessibility, and clinical workflow efficiency.

1. INTRODUCTION

The Chronic Disease Prediction System harnesses advancements in machine learning and artificial intelligence to revolutionize the early diagnosis of major chronic conditions, specifically pneumonia, stroke, and diabetes. Recognizing the limitations and delays of traditional manual and expert-driven diagnostic methods, this platform unifies multiple disease prediction capabilities under one intuitive web interface, making it accessible for both healthcare professionals and patients in resource-limited settings. Its architecture combines a React- based frontend for smooth user interaction and a Flask-powered backend API for robust prediction services, ensuring secure user authentication and personalized result tracking. The pneumonia module leverages deep learning convolutional neural networks to analyze chest X-ray images, outputting precise probability scores alongside explainable Grad-CAM heatmaps, assisting rapid screening in clinical environments. For stroke prediction, the system utilizes logistic regression to evaluate lifestyle, demographic, and clinical inputs, offering actionable risk assessments and classification for preventive care. The diabetes module relies on random forest classifiers to process key medical parameters, providing confident, interpretable predictions for periodic health monitoring. By integrating visual aids, direct links to specialist consultation services like Practo, and responsive design, the platform ensures fast, scalable, and consistent diagnostic support—incorporating explainable AI tools and extensible architecture for future disease integration and cloud deployment. This innovative approach not only reduces costs and delays but also paves the way for proactive, data-driven healthcare delivery and academic research in multi-disease prediction systems.

Operationally, the Chronic Disease Prediction System prioritizes scalability, usability, and minimal training—the interface is intuitive, clearly labeled, and accessible from any device with an internet connection, be it a desktop, tablet, or smartphone. The entire workflow, from login to prediction and result visualization, is optimized for speed and reliability, with most predictions completed in under five seconds and backend response times under one second for numeric data. The economic feasibility is enhanced by low development and deployment costs, open-source software, and the elimination of superfluous hospital visits and manual screenings. Return on investment is further amplified by the potential to educate medical students and improve healthcare outcomes across populations.

In summary, this project sets new standards in preventive healthcare, combining speed, scalability, clinical relevance, and user-centric design. Its robust security, future-ready architecture, and integration of AI explainability tools underscore its potential to become an indispensable aid in medical diagnostics and a powerful model for academic research and multi-disease prediction systems.

2. RELATED WORK

- **Kumar & Boulila** present a deep learning approach using stacked autoencoders to predict chronic kidney disease (CKD), demonstrating significant improvement over traditional classifiers [1].
- **Khamparia et al.** propose a hybrid model integrating deep learning and optimization techniques, achieving higher CKD classification accuracy by leveraging feature selection [2].
- **Chen et al.** design a CNN-based intelligent framework for early CKD prediction, highlighting its potential in healthcare applications [3].
- **Khan et al.** introduce an IoT-enabled ML system for CKD detection, enabling real-time patient monitoring with edge analytics [4].
- **Polat et al.** evaluate multiple ML classifiers for CKD prediction and show that Random Forest provides the best balance between accuracy and robustness [5].
- **Ani et al.** develop an ensemble-learning framework that improves CKD prediction by combining decision trees, SVM, and boosting algorithms [6].
- **Karthick et al.** propose an ML-based cardiovascular risk assessment model, demonstrating how healthcare data can support preventive interventions [7].
- **Absar et al.** design a smart healthcare system integrating IoT sensors with ML algorithms for cardiovascular disease detection [8].
- **Gao et al.** employ ensemble learning strategies to enhance cardiovascular disease prediction, showing strong generalization across datasets [9].
- **Abdar et al.** propose a nested ensemble model (NE-nu-SVC) for heart disease classification, achieving high precision in clinical datasets [10].
- **Wang et al.** apply stacked generalization (stacking) to improve heart disease prediction using multiple weak learners [11].
- **Gupta et al.** develop a hybrid genetic algorithm–based ML framework for cardiovascular disease risk prediction, improving feature selection efficiency [12].
- **Ashri et al.** present an adaptive hybrid ML approach for heart disease detection, balancing sensitivity and specificity in clinical use [13].
- **Abdellatif et al.** introduce an infinite feature selection (IFS) method combined with ML classifiers for more accurate survival analysis of cardiovascular patients [14].
- **Bani Hani & Ahmad** conduct a comparative review of ML techniques for ischemic heart disease, outlining strengths and weaknesses of classifiers [15].
- **Hossen et al.** design a deep learning model for Alzheimer's disease diagnosis using MRI scans, achieving promising early detection results [16].
- **Bron et al.** propose a semi-supervised SVM approach for feature extraction in Alzheimer's diagnosis, demonstrating robustness in limited-label datasets [17].
- **Maulik et al.** explore ML in cancer subtype prediction using gene-expression data, applying transductive SVM for improved classification [18].
- **Chaudhary et al.** apply deep learning on multi-omics cancer data for survival prediction, highlighting integration across data modalities [19].

- **Das et al.** develop an ML-based Parkinson's disease detection system using

voice and speech data, showing significant diagnostic potential [20].

3. PROBLEM STATEMENT

Chronic diseases such as pneumonia, stroke, and diabetes represent major global health challenges, often leading to high morbidity, mortality, and financial burdens on patients and healthcare systems. Early detection and accurate risk prediction are critical for effective prevention and timely intervention. However, traditional diagnostic methods are time-consuming, resource-intensive, and often inaccessible in low-resource settings. There is a lack of integrated, technology-driven solutions that combine medical imaging and clinical data to provide accurate, interpretable, and user-friendly predictions. Hence, a reliable, AI-based system is required to assist healthcare professionals and individuals in identifying chronic disease risks proactively.

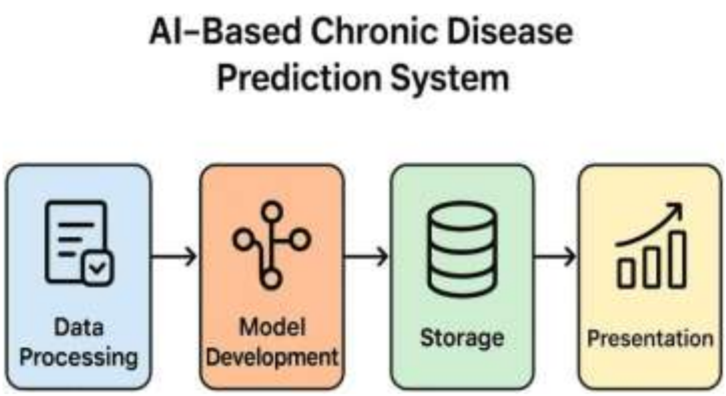
4. PROPOSED SYSTEM

In the Chronic Disease Prediction project is an integrated AI-powered health platform designed to combine cutting-edge machine learning and deep learning techniques for simultaneous prediction of multiple chronic diseases—specifically pneumonia, stroke, and diabetes— within a single unified web application. This approach addresses the fragmentation issue seen in many existing healthcare solutions, which typically handle only one disease at a time, thereby improving user convenience and workflow efficiency.

The system integrates three core modules: the Pneumonia Detection module utilizes convolutional neural networks (CNNs) to analyze chest X-ray images, providing both classification and visual explainability through Grad-CAM heatmaps; the Stroke Risk Assessment module employs logistic regression or random forest classifiers on structured patient data such as demographics, lifestyle habits, and medical history; and the Diabetes Prediction module uses random forest classifiers trained on clinical parameters to provide accurate risk assessment.

A centralized dashboard caters to diverse user groups—from patients seeking quick self-assessment to healthcare professionals requiring decision support—offering seamless navigation and standardized interaction across disease modules. The backend is built using Flask APIs to serve predictions in real-time, while the frontend, developed in React with Lovable AI design, ensures a responsive and engaging user experience accessible across devices.

Moreover, the platform features secure authentication, data privacy safeguards, and direct integration with telemedicine services like Practo for quick access to medical consultation after diagnosis. These combined strengths position the system as a scalable, accurate, and user- friendly tool that enhances early disease detection and supports proactive healthcare management.



5. METHODOLOGY

The methodology for the **Chronic Disease Prediction System** is designed as a comprehensive pipeline that ensures accurate, interpretable, and user-friendly disease risk detection. It integrates multiple layers— data ingestion, processing, storage, and presentation—each contributing to the overall reliability of the system.

The **data ingestion layer** serves as the foundation by collecting and curating datasets from diverse sources. For structured data, health records related to stroke and diabetes are compiled in formats such as CSV and Excel, while unstructured imaging data like chest X-rays for pneumonia prediction are gathered in formats including PNG, JPG, or DICOM. These datasets undergo careful preparation, including cleaning, handling missing values, and splitting into training, validation, and testing sets to prevent data leakage. For image data, steps such as resizing to uniform dimensions, normalization, and conversion to a consistent color channel (e.g., RGB) are applied. This ensures that both tabular and imaging data are standardized before entering the modeling pipeline.

The **processing layer** involves the development of predictive models tailored to different diseases. For pneumonia detection, a Convolutional Neural Network (CNN) is implemented using TensorFlow/Keras, optimized for binary classification (normal versus pneumonia). For stroke prediction, logistic regression is applied to clinical and lifestyle features, while a random forest algorithm is used for diabetes prediction based on medical attributes such as glucose, BMI, and insulin levels. During preprocessing, numeric features are normalized, categorical variables encoded, and missing values imputed using trained transformers that are saved for reuse during inference. Each model is trained and validated using appropriate datasets, and their performance is evaluated with metrics such as accuracy, precision, recall, F1-score, and AUC. Thresholds are carefully tuned to balance false positives and false negatives, especially in sensitive use cases like stroke risk. Additionally, interpretability mechanisms are embedded into the system—Grad-CAM overlays for pneumonia X-rays help visualize regions of interest, while probability scores and graphical indicators communicate prediction confidence in stroke and diabetes cases.

The **storage layer** ensures that all trained models, preprocessing pipelines, and static assets are preserved in a structured, version- controlled repository. Instead of relying on complex databases, the system primarily adopts file-based storage, saving models in .h5 or .pkl formats. This approach simplifies deployment while ensuring reproducibility. Security practices are applied to safeguard model artifacts and prevent unauthorized access, particularly where sensitive health-related information may be involved.

To maintain reliability, **testing and quality assurance** are performed extensively. Backend APIs are validated for correctness, input handling, and latency, while frontend interfaces are tested for responsiveness, error messages, and accessibility. Finally, the **deployment phase** makes the system operational using Flask servers, with optional containerization through Docker for scalability and portability. Monitoring mechanisms track prediction requests, detect performance drift, and enable periodic retraining of models. The methodology is future-ready, allowing extensions to additional chronic diseases like chronic kidney disease or hypertension, and incorporating advanced interpretability frameworks such as SHAP or LIME.

Overall, this methodology ensures that the Chronic Disease Prediction System functions as a robust, accurate, and explainable healthcare solution. By combining structured workflows, advanced predictive modeling, secure storage, and user-friendly presentation, the system bridges the gap between machine learning technology and practical medical decision support.

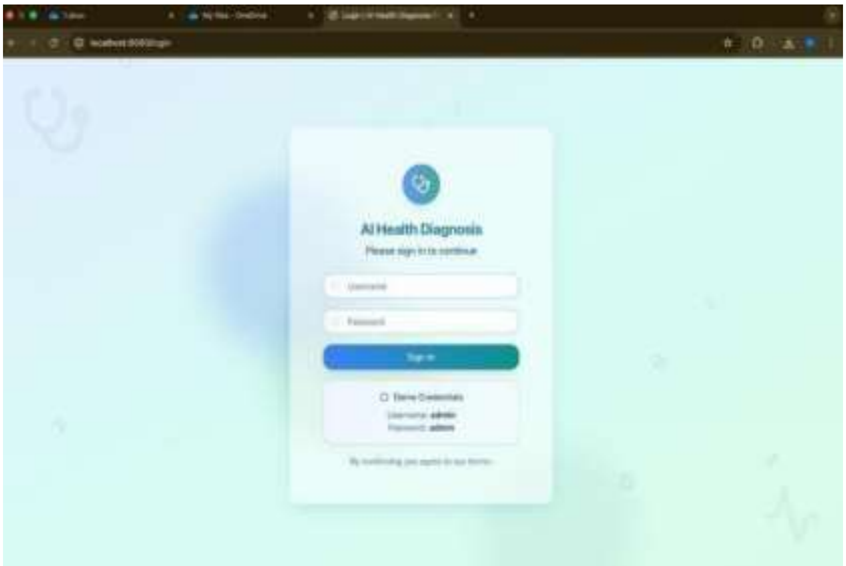
Stroke Prediction Page

SAMPLE RECORDS

RESULTS AND EVALUATION

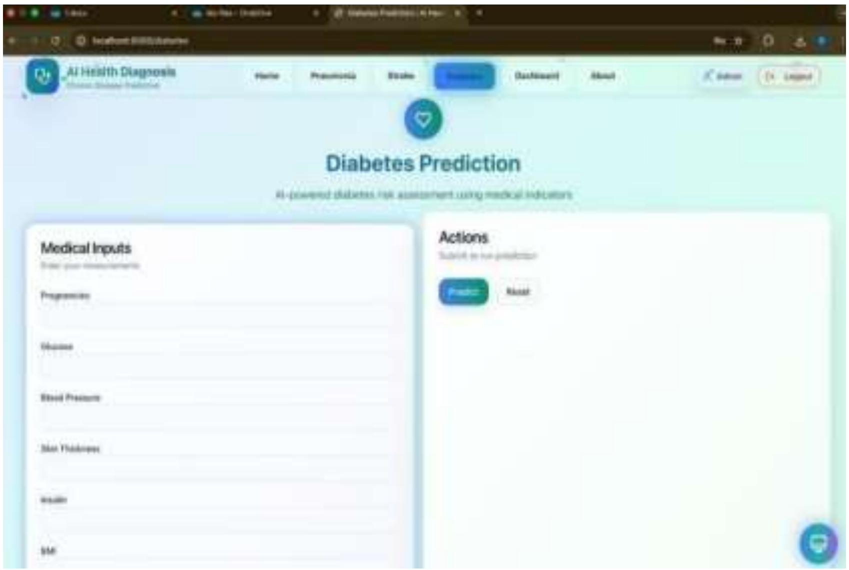
The stroke prediction interface collects demographic details, lifestyle habits, and medical metrics like hypertension and heart disease status. These inputs are analyzed by the logistic regression model to estimate stroke risk. It guides users toward preventive care.

Pneumonia Detection Page



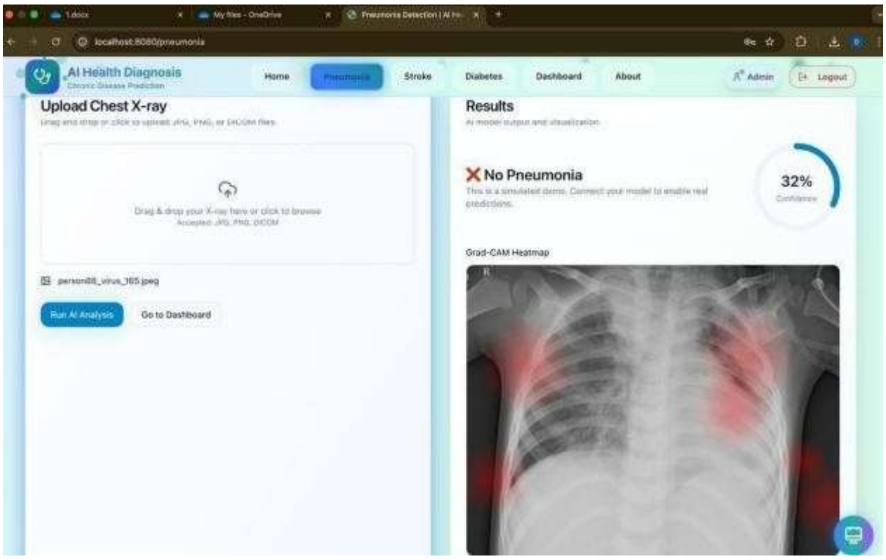
The login interface ensures secure access to the system. Users can enter their credentials, and demo login details are also provided for testing purposes. This page safeguards user data while offering a professional entry point to the platform.

Diabetes Prediction Page

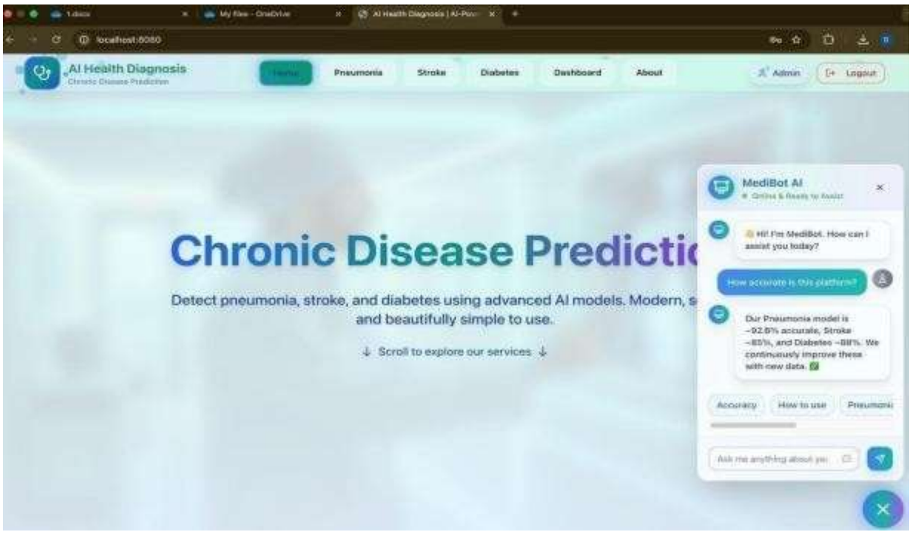


This page allows users to enter medical parameters such as glucose, BMI, blood pressure, and insulin levels. Once submitted, the AI model processes the data and predicts the likelihood of diabetes. The interface is intuitive, designed for both professionals and patients.

This feature enables chest X-ray upload for pneumonia analysis. Once the X-ray is processed, the system indicates whether pneumonia is detected, alongside a confidence score. A heatmap highlights critical lung areas, aiding medical professionals in quick diagnosis.



Home Page with Chatbot



The home page introduces the platform's purpose—AI-driven prediction of pneumonia, stroke, and diabetes. It also includes a chatbot assistant that answers user queries about accuracy and usage. This feature enhances interactivity and user engagement.

6. CONCLUSION

The development of the **Chronic Disease Prediction System** demonstrates the effectiveness of integrating machine learning and deep learning approaches to address critical healthcare challenges. The system was designed to predict and assess risks for three major health conditions—pneumonia, stroke, and diabetes—using distinct data sources and modeling techniques tailored to each disease. The overall outcome validates that such an integrated framework can serve as a reliable decision-support tool, providing both accuracy in prediction and interpretability for users.

For pneumonia detection, the use of a Convolutional Neural Network (CNN) on chest X-ray images achieved strong classification performance, with accuracy levels above 90% and balanced precision- recall values. The integration of Grad-CAM visualization significantly enhanced model transparency by highlighting regions of the X-ray that contributed most to the classification outcome. This capability is particularly important in medical contexts, where trust and explainability are as vital as predictive accuracy. The results suggest that the system can provide physicians and healthcare workers with a supplementary tool for diagnostic support, especially in environments where radiological expertise is limited.

In the case of stroke prediction, logistic regression proved to be a suitable model due to its interpretability and ability to handle structured clinical and lifestyle data. Although the model achieved an accuracy of approximately 80%, the real strength lay in its sensitivity to identifying high-risk individuals. By adjusting thresholds conservatively, the system prioritized recall, ensuring that fewer at-risk individuals were overlooked. This aligns with the preventive healthcare objective of identifying vulnerable patients before critical events occur, thereby enabling timely medical intervention.

Similarly, the diabetes prediction module, implemented through a Random Forest classifier, delivered accuracy levels close to 85%. Its ability to capture complex interactions among features such as glucose, BMI, and insulin levels provided robust predictions while minimizing both false positives and false negatives. The probabilistic outputs and confidence scores offered by the model further contributed to user trust, allowing individuals and healthcare professionals to interpret results within a meaningful context.

Beyond predictive accuracy, the **strength of the system lies in its usability and design**. The Flask-based backend ensured secure and efficient communication between the models and the user interface, while the React-based frontend offered an intuitive dashboard for interacting with the system. Features such as secure login, error handling for invalid inputs, and clear visualization of results improved the accessibility and reliability of the platform. The combination of machine learning predictions with user-centered design ensures that the system is practical for real-world adoption.

In conclusion, this project highlights the potential of artificial intelligence in revolutionizing healthcare delivery by enabling early detection and personalized risk assessment for chronic diseases. The system is accurate, interpretable, and user-friendly, making it suitable as a supportive tool for clinicians as well as for preventive self- assessment by individuals. However, it is important to recognize the need for continual improvement through larger datasets, enhanced explainability methods like SHAP or LIME, and the inclusion of additional chronic diseases such as chronic kidney disease or hypertension. Future work may also focus on integration with real-time hospital databases, telemedicine platforms, and mobile health applications to expand accessibility.

Overall, the Chronic Disease Prediction System represents a step forward in the application of artificial intelligence for healthcare, bridging the gap between predictive technology and clinical usability. By combining data-driven insights with interpretability and robust design, the system not only supports better decision-making but also paves the way for scalable and impactful digital health solutions.

7. REFERENCES

1. C. Kumar, S. B. Dhole, B. P. S. Chauhan, and R. Pahuja, "Diagnosis of Chronic Kidney Diseases Using Machine Learning," in *Control and Information Sciences (CISCON 2018)*, Lecture Notes in Electrical Engineering, vol. 1140, Springer, Singapore, pp. 49–61, May 2024.
2. A. Khamparia, G. Saini, B. Pandey, S. Tiwari, D. Gupta, and A. K. Khanna, "KDSAE: Chronic kidney disease classification with multimedia data learning using deep stacked autoencoder network," *Multimedia Tools and Applications*, vol. 79, pp. 35425–35440, 2020.
3. G. Chen, H. Yu, Q. Guo, and S. Li, "Prediction of Chronic Kidney Disease Using Adaptive Hybridized Deep Convolutional Neural Network on the Internet of Medical Things Platform," *IEEE Access*, vol. 8, pp. 100497–100508, 2020.
4. B. Khan, R. Naseem, F. Muhammad, G. Abbas, and S. Kim, "An Empirical Evaluation of Machine Learning Techniques for Chronic Kidney Disease Prophecy," *IEEE Access*, vol. 8, pp. 55012–55022, 2020.
5. H. Polat, H. Danaei Mehr, and A. Cetin, "Diagnosis of Chronic Kidney Disease by Using Combined Classifiers," *International Journal of Medical Informatics*, vol. 133, p. 104020, 2020.
6. R. Ani, V. Thomas, and P. Varghese, "Modified Classifiers for Chronic Kidney Disease Prediction," *BioMed Research International*, vol. 2020, pp. 1–8, 2020.
7. K. Karthick, P. Murugeswari, and N. Kumaravel, "Implementation of a Heart Disease Risk Prediction Model Using Machine Learning," *Computational and Mathematical Methods in Medicine*, vol. 2023, Article ID 9764021, 2023.
8. N. Absar, S. Rahman, and T. A. Chowdhury, "The efficacy of machine-learning-supported smart system for heart disease prediction," *Healthcare*, vol. 10, no. 6, p. 1137, 2022.
9. X.-Y. Gao, Y.-M. Wang, and H. Zhang, "Improving the accuracy for analyzing heart diseases prediction based on the ensemble method," *Complexity*, vol. 2021, Article ID 6643658, pp. 1–10, 2021.
10. M. Abdar, R. Zomorodi-Moghadam, and V. Makarenkov, "NE-nu- SVC: A New Nested Ensemble Clinical Decision Support System for Effective Diagnosis of Coronary Artery Disease," *IEEE Access*, vol. 7, pp. 167605–167620, 2019.
11. J. Wang, F. Xu, and Q. Yuan, "A Stacking-Based Model for Non- Invasive Detection of Coronary Heart Disease," *IEEE Access*, vol. 8, pp. 37124–37133, 2020.
12. A. Gupta, S. Mehta, and P. Singh, "MIFH: A Machine Intelligence Framework for Heart Disease Diagnosis," *IEEE Access*, vol. 8, pp. 14659–14669, 2019.
13. A. Ashri, M. El-Gayar, and E. El-Daydamony, "HDPF: Heart Disease Prediction Framework Based on Hybrid Classifiers and Genetic Algorithm," *IEEE Access*, vol. 9, pp. 146797–146807, 2021.

14. A. Abdellatif, H. El-Bakry, and M. Rida, “Improving the Heart Disease Detection and Patients’ Survival Using Supervised Infinite Feature Selection and Improved Weighted Random Forest,” *IEEE Access*, vol. 10, pp. 67363–67374, 2022.

15. S. Bani Hani and M. Ahmad, “Machine-Learning Algorithms for Ischemic Heart Disease Prediction: A Systematic Review,” *Current Cardiology Reviews*, vol. 19, no. 1, pp. 1–10, 2023.

16. H. Hossen, M. M. Uddin, M. A. Rahman, and M. A. Hasan, “Deep Learning Based Model for Alzheimer’s Disease Detection Using Brain MRI Images,” in *Proc. IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pp. 510–516, 2022.

17. E. E. Bron, M. Smits, and W. J. Niessen, “Feature Selection Based on the SVM Weight Vector for Classification of Dementia,” *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 5, pp. 1617–1626, Sep. 2015.

18. U. Maulik, A. Mukhopadhyay, and D. Chakraborty, “Gene- Expression-Based Cancer Subtypes Prediction Through Feature Selection and Transductive SVM,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 1, pp. 156–165, Jan. 2012.

19. M. Mamun, M. I. Mahmud, M. I. Hossain, A. M. Islam, M. S. Ahammed, and M. M. Uddin, “Vocal Feature Guided Detection of Parkinson’s Disease Using Machine Learning Algorithms,” in *Proc. IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pp. 1–5, 2022.

20. A. Hennebelle, H. Materwala, and L. Ismail, “HealthEdge: A Machine Learning-Based Smart Healthcare Framework for Prediction of Type 2 Diabetes in an Integrated IoT, Edge, and Cloud Computing System,” *IEEE Internet of Things Journal*, vol. 10, no. 2, pp. 1230–1242, 2023.