© DEC 2025 | IRE Journals | Volume 9 Issue 6 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I6-1712914

Addressing Cardiovascular Health Disparities: From Modifiable Risks to Primary Care Interventions

SARANYA RAJ[1], NIKHIL TIWARY[2], AMAN KUMAR[3], ANKUR PAL[4]
[1, 2, 3,4]*Department of Computer science and Engineering, KCC Institute of technology and Management, Greater Noida, UP*

Abstract- heart disease, also known as CVD, is the biggest reason people lose their lives in India. These complications may show clear signs or remain silent without any visible signs. Stopping them, identifying them early, and treating them on time are required to avoid poor quality of life and reduce the risk of death. A lot of research has shown that AI models can accurately predict stroke and coronary artery disease. These improvements for AI and machine learning are meant to do more than just help with early detection; they are also meant to make health interventions more personal. Researchers are still working on improving these technologies. There are also no studies that look at the causal links between the lifestyle factors that were used to make predictions. So, the goal of this study was to look at a lot of different causal links to see how certain lifestyle habits affect the risk of CVD. We used data from metabolic syndrome screening in the general population to build a prediction model for CVD in this study.

Index terms: Cardiovascular Disease, AI Models, Machine Learning, Prediction, Metabolic Syndrome

I. INTRODUCTION

Cardiovascular disease (CVD) is one of the paramount causes of death in India, and it's grave because it causes various types of health problems like arrhythmias, coronary heart disease (CHD), and heart failure that lead to painful death. Various health organizations of India, like the National Institute of Health (NHI) (.GOV), claim that cardiovascular diseases like stroke cause 17.7 million deaths globally and are one of the prominent issues. [1, 2]

Cardiovascular diseases are not only untreatable but also last for half of life in the body and cause diseases like heart attack, angina, and heart failure. Various factors are there that cause CVD, which include an uneven lifestyle, an unhealthy diet, smoking, high blood pressure, obesity, use of tobacco, and a family history of heart disease. There is no organized and structured data collection method that defines cardiac death for mainland India; also, the majority of deaths happen at home without knowing the exact cause of death. As we are aware, India is a developing country; the residents of the country have to face many challenges in the diagnosis and treatment of CVD due to limited facilities. But in the field of technology, machine learning provides several clinical decision-making tools that help in the early detection of CVD. Technologies that can pull medical data can find important information and perhaps save lives. Our technique could save millions of lives by predicting cardiovascular illness and allowing more individuals to get treatment sooner and faster. [1, 2]

II. RELATED WORK

Several researchers Work done in the past on predicting heart disease has used many different computer-based methods, and each study has tried to improve how early the condition can be identified. Earlier research mostly depended on basic statistical tools like logistic regression. Earlier models were quite straightforward and easy to use, but they often missed the deeper relationships hidden in medical data. Over time, researchers began experimenting with more capable techniques—methods like SVM, KNN, decision trees, Random Forest, and even XGBoost—because these approaches handled the data patterns much better. These newer models. generally performed better and handled clinical data more effectively. [3,4] Many recent papers also point out that the way a model is tuned plays a big role in its final accuracy. Techniques for choosing the right set of parameters and selecting meaningful features have helped make predictions more reliable. Another important direction has been model transparency. Tools like SHAP are now commonly used so that doctors can see which factors are influencing a prediction instead of receiving a “black box” answer. Some studies have even packed these models into

© DEC 2025 | IRE Journals | Volume 9 Issue 6 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I6-1712914

simple apps or web platforms so that people in rural or understaffed medical centers can benefit from them.[6]

III. DATASET DESCRIPTION

The dataset used in this study consists of entries collected from a health screening form with 303 records and 14 key attributes. These include age, sex, resting blood pressure, cholesterol level, fasting blood sugar, chest pain type, ECG characteristics, maximum heart rate, exercise induced angina, ST depression, ST slope, number of major vessels, and thalassemia type. The target variable indicates whether the subject showed signs of heart disease. The dataset provides a balanced combination of numerical and categorical variables suitable for classical supervised learning models. [3, 4, 5, 6]

IV. SYSTEM ARCHITECTURE

The architecture developed for this study follows a structured pipeline.s First, raw data is cleaned, imputed, and encoded. Next, numerical features undergo standardization to ensure uniform scale. After preprocessing, several machine learning models—including logistic regression, KNN, decision tree, SVM, random forest, and XGBoost—are trained and evaluated. Hyperparameters are tuned using cross validation, and the best model is selected based on ROC AUC. SHAP explainability is incorporated to interpret global and individual predictions. The final model is packaged into a Streamlit interface for real time usage. [7, 8] In below Figure 1 show Pipeline of Machine Learning Models.

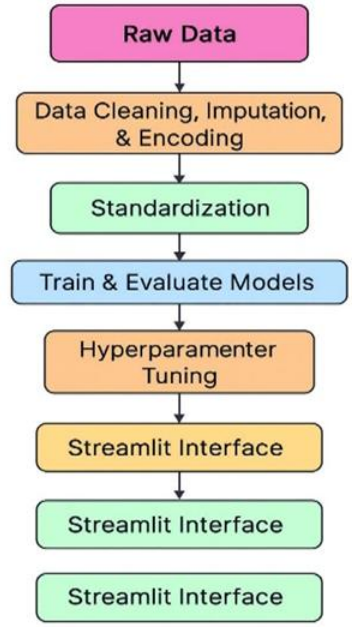

Figure 1: Pipeline of Machine Learning Models

V. MODEL TRAINING

The dataset was split into training and testing sets using an 80–20 stratified approach to retain the original class proportions. Missing numerical values were filled using the median of each feature, and categorical variables were transformed into numerical form through One-Hot Encoding to make them suitable for model training. Outliers were detected using clinically validated threshold ranges and subsequently removed to ensure greater data integrity. Several machine-learning techniques were examined as part of the study, but the Random Forest model proved to be the most dependable, performing well across precision, recall, and AUC metrics. After its parameters were fine-tuned through GridSearchCV, the model showed greater stability and produced more uniform results during cross-validation. [9,10]

VI. DEPLOYMENT STRATEGY

A Streamlit tool was built that allows users to enter patient information and receive an immediate estimate of heart-disease risk, along with explanations generated through SHAP. The system can be expanded by adding cloud support, API connections, or container-based setups to make it easier to use at larger scales. With these improvements, the application can help remote clinics and online healthcare services, especially in places where heart-care specialists are not easily available.

VII. RESULT

Machine learning is a rapidly advancing field of computational algorithms that aim to imitate human intelligence by learning from data and the surrounding environment. These algorithms play a vital role in processing and analyzing large-scale datasets, often referred to as "big data," and have proven effective across various domains, including pattern recognition, computer vision, aerospace engineering, and biomedical applications. Machine learning automates model building by enabling machines to discover hidden patterns and insights without explicit instructions, allowing continuous improvement and adaptation as new data is encountered. As shown in figure No 2. [11,12,13,20]

Machine learning generally follows two major approaches:
supervised learning and unsupervised learning. In supervised learning, the model is trained on data that already includes the correct answers, allowing it to learn the relationship between the inputs and their expected outputs. Learning occurs by comparing the machine's predictions with accurate results to identify errors, making this approach suitable for predicting future events based on past data. [14,15,19] In contrast, unsupervised learning allows the machine to explore data independently, discovering patterns or structures without guidance. This approach is commonly used to detect outliers or clusters in unlabeled datasets. The distribution of major continuous variables is illustrated in Figure 3 [16,17,18]

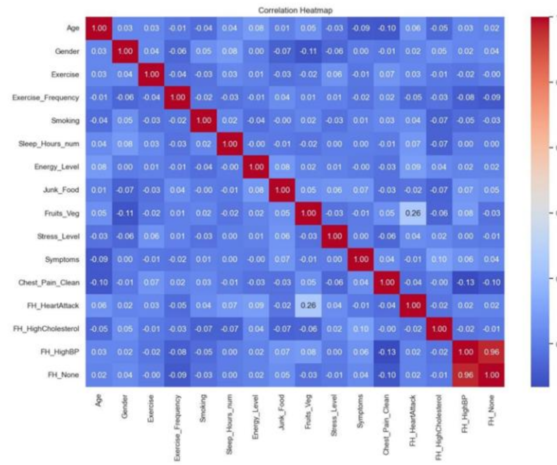

Figure 2. Correlation heatmap of input features used for machine learning analysis.
Machine learning is a rapidly advancing field of computational algorithms that aim to imitate human intelligence by learning from data and the surrounding environment. These algorithms play a vital role in processing and analyzing large-scale datasets, often referred to as big data, and have proven effective across various domains, including biomedical applications. Correlation analysis supports automated model learning by revealing relationships among variables and enabling discovery of hidden patterns without explicit programming.

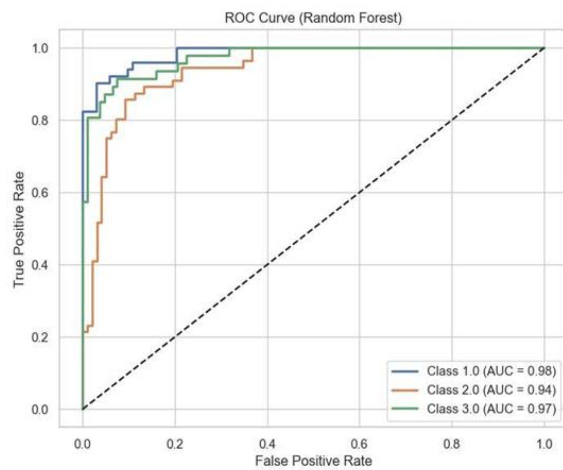

Figure 3. Receiver Operating Characteristic (ROC) curves for the Random F
The ROC curves illustrate the trade-off between true positive rate and false
for each class. High AUC values indicate strong discriminative performance
across multiple classes, demonstrating the robustness and reliability of the
classifier in predicting heart disease outcomes.

VIII. CONCLUSION

This study demonstrates that machine learning can serve as an effective early-screening mechanism for cardiovascular diseases when combined with explainability techniques. The Random Forest classifier offered strong predictive capabilities, and the user-friendly interface showcased real-world feasibility. With further enhancement, such systems can supplement clinical examinations and improve preventive care across diverse communities

REFERENCES

[1] El-Sofany, H., Bouallegue, B., & Abd El-Latif, Y.M. (2024). A proposed technique for predicting heart disease using feature selection and multi-model ensemble. Scientific Reports, 14, 23277.

[2] Ahmad, A.A., & Polat, H. (2023). Prediction of heart disease using ML with jellyfish optimization. Diagnostics (Basel).

[3] Bhatt, C.M., Patel, P., Ghetia, T., & Mazzeo, P.L. (2023). Effective heart disease prediction using machine learning techniques. Algorithms, 16(2), 88.

[4] Subramani, S., et al. (2023). Cardiovascular diseases prediction by ML: A comprehensive review. Frontiers in Medicine.

[5] Hajiarbabi, M. (2024). Heart disease detection using machine learning: A narrative review. JMAI.

[6] Hossain, S., et al. (2024). ML approach for predicting cardiovascular disease. BMC Cardiovascular Disorders.

[7] Banerjee, T. (2025). Systematic review of ML in heart disease prediction. PMC.

[8] Azimi Lamir, A., et al. (2025). Comprehensive ML framework for heart disease prediction. arXiv:2505.09969.

[9] Yi, J., Yu, P., Huang, T., et al. (2024). Transformer model for heart disease prediction using PSO. arXiv:2412.02801.

[10] Karmakar, R., Ghosh, U., Pal, A., et al. (2024). Data balancing expert system for heart disease prediction. arXiv:2407.18606.

[11] Muhammad, Y., Tahir, M., Hayat, M., & Chong, K.T. (2020). Early detection of heart disease using intelligent computational models. Scientific Reports.

[12] Ahsan, M.M., Luna, S.A., & Siddique, Z. (2022). ML-based disease diagnosis: A review. Healthcare (Basel).

[13] Logabiraman, G., et al. (2024). Heart disease prediction using ML algorithms. MATEC Web of Conferences.

[14] Ramalingam, V.V., Dandapath, A., & Karthik Raja, M. (2018). Heart disease prediction using ML techniques. IJET.

[15] Nagavelli, U., Samanta, D., & Chakraborty, P. (2022). Heart disease detection models using ML. Journal of Healthcare Engineering.

[16] Maiga, J., & Hungilo, G.G. (2019). Comparison of ML models for cardiovascular prediction. ICIMCIS Proceedings.

[17] Ingole, B.S., et al. (2024). Advancements in heart disease prediction using ML. arXiv:2410.14738.

[18] Ye, Z., An, S., Gao, Y., et al. (2023). Prediction of mortality in CKD patients with CAD using ML. European Journal of Medical Research.

[19] Dinov, I.D. (2023). Data Science and Predictive Analytics: Biomedical Applications. Springer.

[20] Singh, A., et al. (2024). Heart disease detection using machine learning models. Procedia Computer Science.