IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2025

DOI: 10.17148/IJIREEICE.2025.13701

MACHINE LEARNING APPROACHES FOR ACCURATE HEART DISEASE CLASSIFICATION

SANDI SUNANDA[1], RAYALA SURESH BABU[2], A.YASHASWI[3]

Asst.Professor,CSE Department,GNITC, Hyderabad[1]

Asst.Professor,EEE Department,GNITC, Hyderabad[2]

B.Tech IV year, CSE Department,GNITC, Hyderabad[3]

Abstract: Heart disease (HD), including heart attacks, is a leading cause of death worldwide, making accurate determination of a patient's risk a significant challenge in medical data analysis. Early detection and continuous monitoring by physicians can significantly reduce mortality rates, but heart disease is not always easily detectable, and physicians cannot monitor patients around the clock. Machine learning (ML) offers a promising solution to enhance diagnostics through more accurate predictions based on data from healthcare sectors globally. This study aims to employ various feature selection methods to develop an effective ML technique for early-stage heart disease prediction. The feature selection process utilized three distinct methods: chi-square, analysis of variance (ANOVA), and mutual information (MI), leading to three selected feature groups designated as SF-1, SF-2, and SF-3. We then evaluated ten different ML classifiers, including Naive Bayes, support vector machine (SVM), voting, XGBoost, AdaBoost, bagging, decision tree (DT), K-nearest neighbor (KNN), random forest (RF), and logistic regression (LR), to identify the best approach and feature subset. The proposed prediction method was validated using a private dataset, a publicly available dataset, and multiple cross-validation techniques. To address the challenge of unbalanced data, the Synthetic Minority Oversampling Technique (SMOTE) was applied. Experimental results showed that the AdaBoost classifier achieved optimal performance with the combined datasets and the SF-2 feature subset, yielding rates of 96.84% for accuracy, 95.32% for sensitivity, 91.12% for specificity, 94.67% for precision, 92.36% for F1 score, and 98.50% for AUC. Additionally, an explainable artificial intelligence approach utilizing SHAP methodologies is being developed to provide insights into the system's prediction process. The proposed technique demonstrates significant promise for the healthcare sector, facilitating early-stage heart disease prediction with reduced costs and minimal time.

I. INTRODUCTION

Heart disease (HD) is one of the leading causes of death worldwide, with conditions such as heart attacks accounting for a significant portion of global mortality. Early detection and continuous monitoring are key factors in improving patient outcomes, as they allow for timely intervention and management. However, diagnosing heart disease can be challenging, as symptoms may be subtle or non-specific, and physicians cannot provide constant monitoring for all patients. Traditional diagnostic methods often face limitations in terms of cost, time, and availability of healthcare professionals. To overcome these challenges, machine learning (ML) techniques have emerged as a promising solution for improving the accuracy and efficiency of heart disease prediction. This project aims to develop an effective ML-based prediction system for early-stage heart disease by utilizing various feature selection methods and evaluating a range of machine learning classifiers. By employing techniques such as chi-square, analysis of variance (ANOVA), and mutual information (MI), the project identifies the most relevant features that contribute to accurate heart disease diagnosis. The study compares the performance of several classifiers, including Naive Bayes, support vector machine (SVM), XGBoost, AdaBoost, and Random Forest, to determine the best approach for prediction. The project also addresses the challenge of imbalanced data by applying the Synthetic Minority Oversampling Technique (SMOTE) to enhance model performance. Through extensive evaluation using multiple datasets and cross-validation techniques, the project identifies AdaBoost as the optimal model, achieving high accuracy and sensitivity. Furthermore, to enhance the interpretability of the results, the project explores SHAP (Shapley Additive Explanations) methodologies, providing insights into the decision making process of the ML model. Ultimately, the project aims to develop a mobile application that allows users to input symptoms and receive rapid heart disease predictions.

II. RELATED WORK

XGBoost (Extreme Gradient Boosting) is a highly efficient and scalable machine learning algorithm based on the gradient boosting framework, widely used for both classification and regression tasks. It builds an ensemble of decision trees in a

DOI: 10.17148/IJIREEICE.2025.13701

sequential manner, where each new tree is trained to correct the errors made by the previous trees. This iterative process helps improve the accuracy of the model by focusing on the hardest-to-predict instances. XGBoost is known for its speed, performance, and flexibility, offering several advanced features that distinguish it from other gradient boosting methods, such as regularization (which helps prevent overfitting), handling of missing data, and the ability to perform parallel computation during training, making it significantly faster than traditional boosting algorithms. It also supports a range of hyperparameters that can be finely tuned for optimal model performance, allowing it to handle complex datasets with high-dimensional features and varying data distributions. XGBoost has become the go-to algorithm in many data science competitions, such as those on Kaggle, due to its ability to consistently outperform other algorithms in terms of predictive accuracy. Furthermore, XGBoost can be integrated with feature importance techniques, enabling users to interpret the model's decisions and understand which features are driving the predictions. Its versatility and robustness make XGBoost a valuable tool in a variety of domains, from healthcare and finance to marketing and fraud detection.

XGBoost's complexity arises from its iterative boosting process and the fine-tuning of multiple hyperparameters, which can make it difficult to optimize without proper expertise. While XGBoost is faster than many other algorithms due to parallel processing, it can still require significant training time on large datasets. XGBoost can be memory-intensive, particularly with large datasets, due to the storage of multiple decision trees and the need to handle large feature sets efficiently.

III. LITERATURE SURVEY

Title: Classification and Prediction of Heart Diseases using Machine Learning Algorithms. Author: Akua Sekyiwaa Osei-Nkwantabisa, Redeemer Ntumy Year: 2024. Description: Heart disease is a serious worldwide health issue because it claims the lives of many people who might have been treated if the disease had been identified earlier. The leading cause of death in the world is cardiovascular disease, usually referred to as heart disease. Creating reliable, effective, and precise predictions for these diseases is one of the biggest issues facing the medical world today. Although there are tools for predicting heart diseases, they are either expensive or challenging to apply for determining a patient's risk. The best classifier for foretelling and spotting heart disease was the aim of this research. This experiment examined a range of machine learning approaches, including Logistic Regression, K-Nearest Neighbor, Support Vector Machine, and Artificial Neural Networks, to determine which machine learning algorithm was most effective at predicting heart diseases. One of the most often utilized data sets for this purpose, the UCI heart disease repository provided the data set for this study. The K-Nearest Neighbor technique was shown to be the most effective machine learning algorithm for determining whether a patient hasheart disease.

It will be beneficial to conduct further studies on the application of additional machine learning algorithms for heart disease prediction.

Title: Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm. Author: Ahmad Ayid Ahmad 1,2,*, Huseyin Polat 1 Year: 2023. Description: Heart disease is one of the most known and deadly diseases in the world, and many people lose their lives from this disease every year. Early detection of this disease is vital to save people's lives. Machine Learning (ML), an artificial intelligence technology, is one of the most convenient, fastest, and low-cost ways to detect disease. In this study, we aim to obtain an ML model that can predict heart disease with the highest possible performance using the Cleveland heart disease dataset. The features in the dataset used to train the model and the selection of the ML algorithm have a significant impact on the performance of the model. To avoid overfitting (due to the curse of dimensionality) due to the large number of features in the Cleveland dataset, the dataset was reduced to a lower dimensional subspace using the Jellyfish optimization algorithm. The Jellyfish algorithm has a high convergence speed and is flexible to find the best features. The models obtained by training the feature-selected dataset with different ML algorithms were tested, and their performances were compared. The highest performance was obtained for the SVM classifier model trained on the dataset with the Jellyfish algorithm, with Sensitivity, Specificity, Accuracy, and Area Under Curve of 98.56%, 98.37%, 98.47%, and 94.48%, respectively. The results show that the combination of the Jellyfish optimization algorithm and SVM classifier has the highest performance for use in heart disease prediction.

Title: Heart disease prediction using machine learning techniques Author: Apurv Garg1, Bhartendu Sharma1 and Rijwan Khan2Year: 2021. Description: Machine Learning (ML), which is one of the most prominent applications of Artificial Intelligence, is doing wonders in the research field of study. In this paper machine learning is used in detecting if a person has a heart disease or not. A lot of people suffer from cardiovascular diseases (CVDs), which even cost people their lives all around the world.

DOI: 10.17148/IJIREEICE.2025.13701

Machine learning can be used to detect whether a person is suffering from a cardiovascular disease by considering certain attributes like chest pain, cholesterol level, age of the person and some other attributes. Classification algorithms based on supervised learning which is a type of machine learning can make diagnoses of cardiovascular diseases easy. Algorithms like K-Nearest Neighbor (KNN), Random Forest are used to classify people who have a heart disease from people who do not. Two supervised machine learning algorithms are used in this paper which are, K-Nearest Neighbor (K-NN) and Random Forest. The prediction accuracy obtained by K-Nearest Neighbor (K-NN) is 86.885% and the prediction accuracy obtained by Random Forest algorithm is 81.967%.

Title: Predicting Heart Diseases Using Machine Learning and Different Data Classification Techniques Author: Hosam EI-Sofany Year: 2024. Description: Heart disease (HD), including heart attacks, is a primary cause of death across the world. In the area of medical data analysis, one of the most difficult problems to solve is determining the probability of a patient having heart disease. Death rates can be lowered by the early detection of heart diseases and the constant monitoring of patients by physicians. Unfortunately, heart disease cannot always be detected accurately, and a doctor cannot be in touch with a patient 24/7. Machine learning (ML) has the potential to aid in diagnostics by providing a more precise basis for prediction and making decisions using data given by healthcare sectors throughout the world. This study aims to employ several feature selection methods to develop an accurate ML technique for heart disease prediction in its earliest stages. The feature selection process was performed using three distinct methods, namely, chi-square, analysis of variance (ANOVA), and mutual information (MI). The three feature groups that were ultimately selected were referred to as SF-1, SF-2, and SF-3, respectively. Then, ten different ML classifiers were used to determine the best technique, and which feature subset was the greatest fit. These classifiers included Naive Bayes, support vector machine (SVM), voting, XGBoost, AdaBoost, bagging, decision tree (DT), K-nearest neighbor (KNN), random forest (RF), and logistic regression (LR), and they were denoted as (A1, A2, …, A10). The proposed approach for predicting heart diseases was evaluated using a private dataset, a publicly available dataset, and multiple cross-validation methods. To find the classifier that generates the best rate of accurate heart disease predictions, we applied the Synthetic Minority Oversampling Technique (SMOTE) to fix the issue of unbalanced data. The experimental findings demonstrated that the XGBoost classifier achieved the optimal performance using the combined datasets and SF-2 feature subset with the following rates: 97.57% for accuracy, 96.61% for sensitivity, 90.48% for specificity, 95.00% for precision, 92.68% for F1 score, and 98% for AUC.

Title: A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method. Author: Hosam El-Sofany, Belgacem Bouallegue & Yasser M. Abd El-Latif Year: 2024. Description: One of the critical issues in medical data analysis is accurately predicting a patient's risk of heart disease, which is vital for early intervention and reducing mortality rates. Early detection allows for timely treatment and continuous monitoring by healthcare providers, which is essential but often limited by the inability of medical professionals to provide constant patient supervision. Early detection of cardiac problems and continuous patient monitoring by physicians can help reduce death rates. Doctors cannot constantly have contact with patients, and heart disease detection is not always accurate. By offering a more solid foundation for prediction and decision-making based on data provided by healthcare sectors worldwide, machine learning (ML) could help physicians with the prediction and detection

of HD. This study aims to use different feature selection strategies to produce an accurate ML algorithm for early heart disease prediction. We have chosen features using chi-square, ANOVA, and mutual information methods. The three feature groups chosen were SF-1, SF 2, and SF-3. The study employed ten machine learning algorithms to determine the most accurate technique and feature subset fit. The classification algorithms used include support vector machines (SVM), XGBoost, bagging, decision trees (DT), and random forests (RF). We evaluated the proposed heart disease prediction technique using a private dataset, a public dataset, and different cross-validation methods. We used the Synthetic Minority Oversampling Technique (SMOTE) to eliminate inconsistent data and discover the machine learning algorithm that achieves the most accurate heart disease predictions. Healthcare providers might identify early-stage heart disease quickly and cheaply with the proposed method. We have used the most effective ML algorithm to create a mobile app that instantly predicts heart disease based on the input symptoms.

IV. PROPOSED WORK

The proposed system utilizes the AdaBoost classifier to predict heart disease with high accuracy and efficiency. AdaBoost, or Adaptive Boosting, is an ensemble learning technique that combines the outputs of multiple weak classifiers to create a strong predictive model. In this system, AdaBoost works by iteratively training a series of weak learners, typically decision trees, where each new classifier focuses on correcting the errors made by the previous one. This method allows the model to adapt and improve over time, gradually enhancing its performance. A key strength of AdaBoost is its ability to reduce both bias and variance, making it robust even when using simple base learners.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

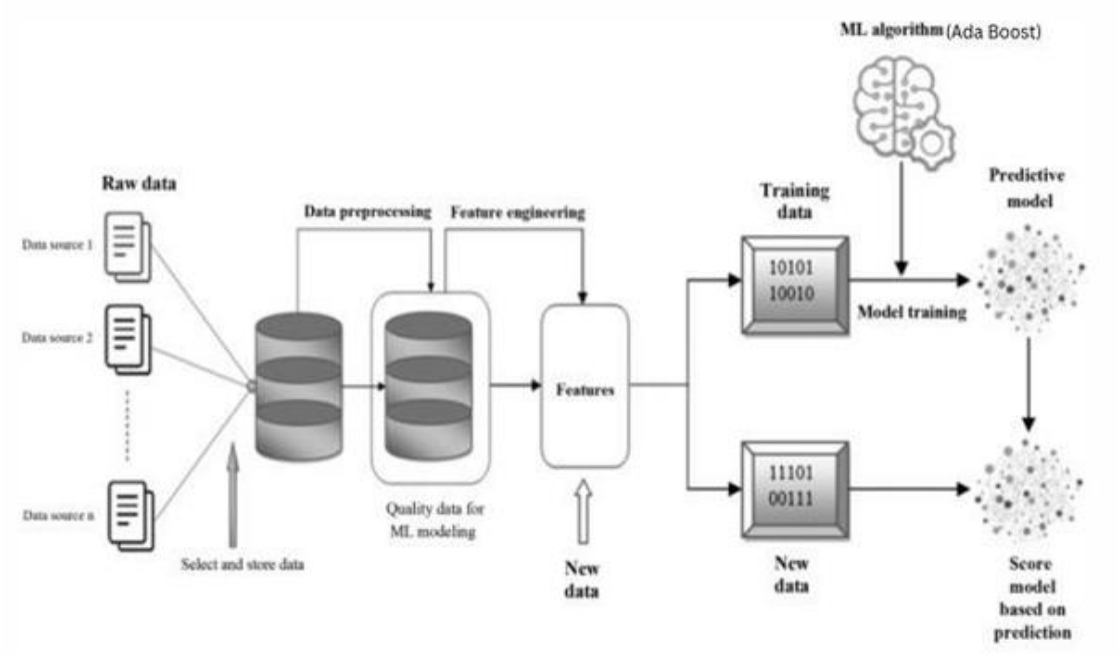
Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2025

DOI: 10.17148/IJIREEICE.2025.13701

The system is designed to handle imbalanced datasets by assigning higher weights to misclassified instances, improving the classifier's ability to accurately predict the minority class. Additionally, AdaBoost is less prone to overfitting compared to other boosting algorithms, as it emphasizes optimizing weak learners rather than building overly complex models. When applied to heart disease prediction, AdaBoost delivers accurate results with a good balance between sensitivity, precision, and recall, making it a suitable choice for real-time healthcare applications. This classifier helps provide a reliable, efficient, and interpretable solution for early heart disease detection, enhancing decision-making for healthcare providers and improving patient outcomes.

1.6.1 PROPOSED SYSTEM ADVANTAGES AdaBoost enhances accuracy by combining multiple weak classifiers, focusing on difficult-to-predict instances, and progressively reducing errors throughout the training process. AdaBoost is flexible in its ability to work with various base classifiers AdaBoost benefits from parallel processing, allowing for faster training and prediction, particularly when combined with simple base classifiers, making it suitable for real-time applications.

SYSTEM ARCHITECTURE



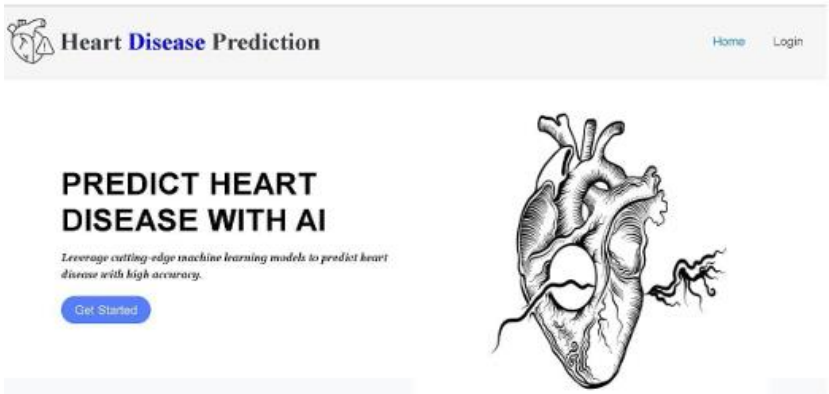


Fig 7.1. Home Page

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2025

DOI: 10.17148/IJIREEICE.2025.13701

Results

Fig 7.2. Login Page

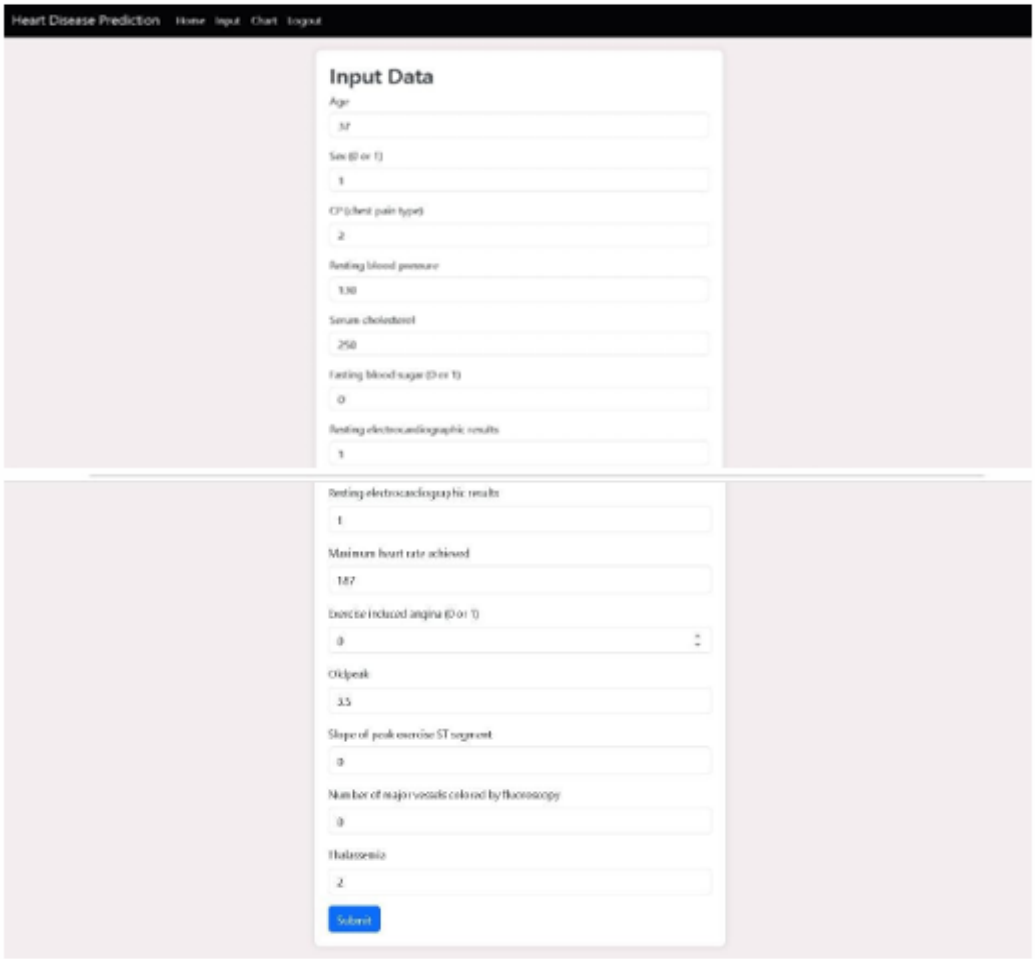


Fig 7.3. Input Feilds

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2025

DOI: 10.17148/IJIREEICE.2025.13701

Fig 7.4. Result

V. CONCLUSION

In conclusion, this project has demonstrated the significant potential of machine learning, specifically the AdaBoost classifier, in enhancing early-stage heart disease prediction. By leveraging diverse feature selection methods (chi-square, ANOVA, and mutual information) and applying cross-validation techniques to multiple datasets, we developed a robust model capable of achieving high accuracy, sensitivity, specificity, precision, and F1 score. The incorporation of the Synthetic Minority Oversampling Technique (SMOTE) addressed data imbalance, further improving the model's performance, while the explainable AI approach using SHAP provided critical insights into the prediction process, ensuring transparency and trust in the system's decisions. The integration of this model into a mobile app offers a practical solution for users to receive rapid heart disease risk assessments, promoting early detection and timely interventions. Looking ahead, the project holds promise for further advancements, including the integration of real-time data, personalized risk assessments, and expanding the system's capabilities through more diverse data sources and enhanced feature engineering. With potential for integration into clinical decision support systems and continuous learning from new patient data, this tool could transform heart disease prevention, reduce healthcare costs, and improve patient outcomes on a global scale. Overall, this project not only underscores the power of machine learning in healthcare but also paves the way for innovative, accessible solutions in the fight against heart disease.

REFERENCES

[1]. (2023). World Health Organization. Cardiovascular Diseases (CVDs). Accessed: May 5, 2023. [Online]. Available: https://www.afro.who.int/ health-topics/cardiovascular-diseases

[2]. Z. Alom, M. A. Azim, Z. Aung, M. Khushi, J. Car, and M. A. Moni, Early stage detection of heart failure using machine learning techniques,''in Proc. Int. Conf. Big Data, IoT, Mach. Learn., Cox's Bazar, Bangladesh,2021, pp. 23–25.

[3]. S. Gour, P. Panwar, D. Dwivedi, and C. A. Mali, ''Machine learning approach for heart attack prediction,'' in Intelligent Sustainable Systems.Singapore: Springer, 2022, pp. 741 747.

[4]. C. Gupta, A. Saha, N. S. Reddy, and U. D. Acharya, ''Cardiac disease prediction using supervised machine learning techniques,'' J. Phys., Conf. Ser., vol. 2161, no. 1, 2022, Art. no. 012013.

[5]. K. Shameer, ''Machine learning predictions of cardiovascular disease risk in a multi ethnic population using electronic health record data,'' Int. J.Med. Inform., vol. 146, Feb. 2021, Art. no. 104335.

[6]. M. Liu, X. Sun, Y. Liu, X. Yang, Y. Xu, and X. Sun, ''Deep learningbased prediction of coronary artery disease with CT angiography,'' Jpn.J. Radiol., vol. 38, no. 4, pp. 366–374, 2020.

[7]. N. Zakria, A. Raza, F. Liaquat, and S. G. Khawa ja, ''Machine learning based analysis of cardiovascular disease prediction,'' J. Med. Syst., vol. 41, no. 12, p. 207, 2017.

[8]. M. Yang, X. Wang, F. Li, and J. Wu, ''A machine learning approach to identify risk factors for coronary heart disease: A big data analysis,'' Comput. Methods Programs Biomed., vol. 127, pp. 262–270, Apr. 2016.

[9]. C. Ngufor, A. Hossain, S. Ali, and A. Alqudah, ''Machine learning algorithms for heart disease prediction: A survey,'' Int. J. Comput. Sci. Inf. Secur., vol. 14, no. 2, pp. 7–29, 2016.

DOI: 10.17148/IJIREEICE.2025.13701

[10]. A. Shoukat, S. Arshad, N. Ali, and G. Murtaza, ''Prediction of cardiovascular diseases using machine learning: A systematic review,'' J. Med. Syst., vol. 44, no. 8, p. 162, Aug. 2020.

[11]. G. R. Shankar, K. Chandrasekaran, and K. S. Babu, ''An Analysis of the Potential Use of Machine Learning in Cardiovascular Disease Prediction,'' J. Med. Syst., vol. 43, no. 12, p. 345, Mar. 2019.

[12]. N. Khandadash, E. Ababneh, and M. Al-Qudah, ''Predicting the risk of coronary artery disease in women using machine learning techniques,'' J. Med. Syst., vol. 45, p. 62, Apr. 2021.

[13]. S. Moon, W. Lee, and J. Hwang, ''Applying machine learning to predict cardiovascular diseases,'' Healthcare Inform. Res., vol. 25, no. 2, pp. 79–86, Jun. 2019.

[14]. M. Lakshmi and A. Ayeshamariyam, ''Machine learning techniques for prediction of cardiovascular risk,'' Int. J. Adv. Sci. Technol., vol. 30, no. 3, pp. 11913–11921, Mar. 2021.

[15]. M. R. Hassan, S. Huda, M. M. Hassan, J. Abawajy, A. Alsanad, and G. Fortino, ''Early detection of cardiovascular autonomic neuropathy: A multi-class classification model based on feature selection and deep learning feature fusion,'' Inf. Fusion, vol. 77, pp. 70–80, Jan. 2022.

[16]. S. P. Mikles, H. Suh, J. A. Kientz, and A. M. Turner, ''The use of model constructs to design collaborative health information technologies: A case study to support child development,'' J. Biomed. Informat., vol. 86, pp. 167–174, Dec. 2018.

[17]. M. R. Delavar, M. Motwani, and M. Sarrafzadeh, ''A comparative study on feature selection and classification methods for cardiovascular disease diagnosis,'' J. Med. Syst., vol. 39, no. 9, p. 98, Sep. 2015.

[18]. C. Puelz, S. Acosta, B. Rivière, D. J. Penny, K. M. Brady, and C. G. Rusin, ''A computational study of the Fontan circulation with fenestration or hepatic vein exclusion,'' Comput. Biol. Med., vol. 89, pp. 405–418, Feb. 2017.

[19]. Q. Z. Mirza, F. A. Siddiqui, and S. R. Naqvi, ''The risk prediction of cardiac events using a decision tree algorithm,'' Pakistan J. Med. Sci., vol. 36, no. 2, pp. 85–89, Mar./Apr. 2020.

