AN IMPROVED MLP-RFE MODEL FOR ENHANCED ACCURACY IN HEART DISEASE PREDICTION USING DEEP LEARNING

ARTICLES FOR FACULTY MEMBERS

A Comprehensive Review on Heart Disease Risk Prediction using Machine Learning and Deep Learning Algorithms. / Karna, V. V. R., Karna, V. R., Janamala, V., Devana, V. N. K. R., Ch, V. R. S., & Tummala, A. B.

Archives of Computational Methods in Engineering
Volume 32 (2022) Pages 1763–1795
https://doi.org/10.1016/j.scp.2022.100726
(Database: Springer Science and Business Media B.V)

A comprehensive review of machine learning for heart disease prediction: challenges, trends, ethical considerations, and future directions. / Kumar, R., Garg, S., Kaur, R., Johar, M. G. M., Singh, S., Menon, S. V., Kumar, P., Hadi, A. M., Hasson, S. A., & Lozanović, J.

Frontiers in Artificial Intelligence
Volume 8 1583459 (2025) Pages 1-31
https://doi.org/10.1007/s10126-025-10555-5
(Database: Frontiers Media SA)

An interpretable machine learning model for predicting all-cause in-hospital mortality in patients with ischemic stroke and coronary heart disease: A multicenter study in China. / Huang, J., Peng, S., Xiang, T., Zhang, Y., Dong, Y., Wang, Q., & Liu, X.

Brain Network Disorders
Volume 2 (2026) Pages 49 - 58
https://doi.org/10.1177/22808000231180390
(Database: ScienceDirect)

29th April 2026

AN IMPROVED MLP-RFE MODEL FOR ENHANCED ACCURACY IN HEART DISEASE PREDICTION USING DEEP LEARNING

ARTICLES FOR FACULTY MEMBERS

A Novel Ensemble Deep Learning Model for Coronary Heart Disease Prediction / Riyaz, L., Butt, M. A., & Zaman, M.

Revue d’Intelligence Artificielle
Volume 36 Issue 6 (2022) Pages 825–832
https://doi.org/10.18280/ria.360602
(Database: International Information and Engineering Technology Association)

Classification models combined with Boruta feature selection for heart disease prediction / Manikandan, G., Pragadeesh, B., Manojkumar, V., Karthikeyan, A. L., Manikandan, R., & Gandomi, A. H.

Informatics in Medicine Unlocked
Volume 44 101442 (2024) Pages 1-12
https://doi.org/10.1016/j.imu.2023.101442
(Database: ScienceDirect)

Effective heart disease prediction using novel MLP-EBMDA approach / Deepika, D., & Balaji, N.

Biomedical Signal Processing and Control
Volume 72 (2022) 103318 Pages 1-9
https://doi.org/10.1016/j.identj.2025.109308
(Database: ScienceDirect)

29th April 2026

AN IMPROVED MLP-RFE MODEL FOR ENHANCED ACCURACY IN HEART DISEASE PREDICTION USING DEEP LEARNING

ARTICLES FOR FACULTY MEMBERS

Heart-related Clinical Biomarker Classification through Machine Learning Algorithms / Safdar, S., Rauf, A., Tariq, A. B., Qazi, S., & Bakhtiar, S. M.

Current Trends in OMICS
Volume 5 Issue 1 (2025) Pages 40 - 66
https://doi.org/10.32350/cto.51.03
(Database: Current Trends in OMICS ,CTO)

Machine learning and Mendelian randomization identify key lifestyle factors in coronary heart disease: A NHANES-Based study / Cui, Y. yang, Zhang, Y., Zeng, L., Li, S., Mei, X., Yang, X., Zhou, P., Xiong, L., Huang, Y., Luo, J., Wu, F., & Yue, R.

International Journal of Cardiology: Cardiovascular Risk and Prevention
Volume 27 (2025) 200536 Pages 1–11
https://doi.org/10.1016/j.ijcrp.2025.
(Database: Science Direct)

29th April 2026

ARTICLES FOR FACULTY MEMBERS

AN IMPROVED MLP-RFE MODEL FOR ENHANCED ACCURACY IN HEART DISEASE PREDICTION USING DEEP LEARNING

A Comprehensive Review on Heart Disease Risk Prediction using Machine Learning and Deep Learning Algorithms. / Karna, V. V. R., Karna, V. R., Janamala, V., Devana, V. N. K. R., Ch, V. R. S., & Tummala, A. B.

Archives of Computational Methods in Engineering
Volume 32 (2022) Pages 1763–1795
https://doi.org/10.1016/j.scp.2022.100726
(Database: Springer Science and Business Media B.V)

Archives of Computational Methods in Engineering (2025) 32:1763–1795
https://doi.org/10.1007/s11831-024-10194-4

REVIEW

A Comprehensive Review on Heart Disease Risk Prediction using Machine Learning and Deep Learning Algorithms

Vishnu Vardhana Reddy Karna[1] · Viswavardhan Reddy Karna[2] · Varaprasad Janamala[3] · V. N. Koteswara Rao Devana[1] · V. Ravi Sankar Ch[1] · Aravinda Babu Tummala[4]

Received: 26 August 2024 / Accepted: 1 October 2024 / Published online: 17 October 2024

Abstract

Cardiovascular diseases claim approximately 17.9 million lives annually, with heart attacks and strokes accounting for over 80% of these deaths. Key risk factors, including hypertension, hyperglycemia, dyslipidemia, and obesity, are identifiable, offering opportunities for timely intervention and reduced mortality. Early detection of heart disease enables individuals to adopt lifestyle changes or seek medical treatment. However, conventional diagnostic methods, such as electrocardiograms—commonly used in clinics and hospitals to detect abnormal heart rhythms—are not effective in identifying actual heart attacks. Additionally, angiography, while more precise, is an invasive method, financial strain on patients, and high chances of incorrect diagnosis, highlighting the need for alternative approaches. The main goal of this study was to assess the accuracy of machine learning techniques, including both individual and combined classifiers, in early detection of heart diseases. Furthermore, the study aims to highlight areas where additional research is necessary. Our investigation covers a decade period from 2014 to 2024, including a thorough review of pertinent literature from international conferences and top journals from the databases like Springer, ScienceDirect, IEEEXplore, Web of Science, PubMed, MDPI, Hindawi and so on. The following keywords were used to search the articles: heart disease risk, heart disease prediction, data mining, data preprocessing, machine learning algorithms, ensemble classifiers, deep learning algorithms, feature selection, hyperparameter optimization techniques. We examine the methodologies used and evaluate their effectiveness in predicting cardiovascular conditions. Our findings reveal notable progress in applying machine learning and deep learning in cardiology. The study concludes by proposing a framework that incorporates current machine learning techniques to enhance heart disease prediction.

1 Introduction

Heart disease, also known as cardiovascular disease (CVD) or coronary artery disease (CAD), is the leading cause of death globally and a major concern in the medical field. CVDs have significant health and economic impacts both in the United States and worldwide [1]. World Health Organization (WHO) reported in 2019, approximately 17.9 million people died from CVD, representing 32% of global deaths. Recent figures indicate that around 20.5 million people die annually from CVD, with projections suggesting that this number could rise to 24.2 million by 2030 [2, 3]. In the

---

✉ Vishnu Vardhana Reddy Karna
vishnuvardhanareddyk@adityauniversity.in

Viswavardhan Reddy Karna
viswavardhank@rvce.edu.in

Varaprasad Janamala
varaprasad.janamala@christuniversity.in

V. N. Koteswara Rao Devana
koteswararaodevana@adityauniversity.in

V. Ravi Sankar Ch
venkataravisankarch@adityauniversity.in

Aravinda Babu Tummala
aravindababu_ece@cbit.ac.in

[1] Department of Electronics and Communication Engineering, Aditya University, Surampalem 533437, India

[2] Department of Artificial Intelligence and Machine Learning, R V College of Engineering, Bengaluru 560059, India

[3] Department of Electrical and Electronics Engineering, Christ University, Bengaluru, Karnataka 560074, India

[4] Department of Electronics and Communication Engineering, Chaitanya Bharati Institute of Technology, Hyderabad 500075, India

United States alone, 30.3 million adults were diagnosed with heart disease in 2018, and it is responsible for over 647,000 deaths each year, making it the leading cause of death in the country. From the observations, it is clear that one in four deaths in the U.S. is due to heart diseases. According to the American Heart Association (AHA), 26% of women are at high risk and will die within a year of experiencing heart attack when compared with 19% of the men. Despite of having several advancements, mortality rates from coronary heart disease in the U.S. have not significantly decreased since 2011 [4, 5].

Timely and accurate diagnosis and prognosis of cardiac disease are crucial for improving patient survival rates. Several risk factors contribute to cardiovascular disease (CVD), including obesity, lack of exercise, genetic mutations, high blood pressure, elevated cholesterol levels, alcohol consumption, and smoking [6–9]. Mortality rates can be reduced through early symptom identification and lifestyle changes, such as increasing physical activity, quitting smoking, and undergoing regular medical checkups. Common diagnostic methods for heart disease include electrocardiography (ECG), blood tests, stress tests, echocardiograms, and coronary angiography [10–13]. Currently, coronary angiograms are widely used by clinicians for precise coronary artery disease diagnosis [14].

However, angiograms are expensive and carry significant risks of complications, making heart disease diagnosis more challenging [15, 16]. Cardiovascular specialists face the difficult task of evaluating numerous factors to make an accurate prognosis. This challenge has driven the development of non-invasive methods for detecting cardiac failure. Presently, heart disease diagnosis often relies on doctors analysing a patient’s medical history, symptoms, and test results. With this approach, clinicians can diagnose cardiovascular disease with approximately 67% accuracy by comparing current symptoms with those observed in previously diagnosed cases [17–19].

Machine learning has diverse applications, from diagnosing illness risk factors to enhancing vehicle safety systems. It provides advanced predictive modelling techniques that address current limitations by leveraging large datasets to develop prediction algorithms. This process involves minimizing the error between predicted and actual results, requiring the computer to discern complex, non-linear relationships between features [20]. By analysing patterns from a given dataset, machine learning can make predictions about new, unfamiliar data. Classification is a powerful machine learning technique used for prediction [21]. As a type of supervised learning, classification is particularly effective in accurately diagnosing diseases when trained with appropriate data.

According to the World Health Organization (WHO), heart attacks and strokes account for over 80% of deaths caused by cardiovascular disease (CVD). Additionally, about one-third of CVD-related deaths occur before the age of 70. Low- and middle-income countries are responsible for more than 75% of CVD fatalities [2]. Timely and accurate prediction of heart disease risk is crucial for enhancing patient survival rates and longevity. Consequently, there is a strong motivation among researchers and academics to develop frameworks or decision support systems that facilitate early disease detection and extend affordable healthcare solutions to low-income regions.

1.1 Informatics of Healthcare

A healthcare system consists of organizations, individuals, and resources designed to deliver high-quality medical care that meets people's health needs. According to the WHO, the goals of a healthcare system are to improve health and medical equity in ways that are responsive and cost-effective, while also using resources efficiently. Advances in technology, particularly in computer networks, have enhanced healthcare services in many countries. As individuals use these services, vast amounts of health data are generated. Collecting this data helps build a comprehensive view of patients, enables personalized treatment, supports current care methods, enhances patient-provider interactions, and improves health outcomes. Health records include personal information (such as name, age, gender, and birthdate), medical details (like diagnoses, treatments, and medications), and clinicians' notes on patients' conditions. These records are essential for delivering quality care. Although medical records are primarily intended to facilitate patient care and assist practitioners in managing their patients, handling them on a large scale is challenging, time-consuming, and prone to errors [22].

In today's healthcare industry, patient medical records are stored in increasingly expansive databases. This data often contains significant repetition and lacks balance. To address these issues, pre-processing is essential for extracting important features, reducing the training algorithm's execution time, and improving classification efficiency. Advances in computing power and the ability to adapt machine learning algorithms have greatly enhanced their capabilities, opening up new research opportunities in healthcare [23]. Early disease prediction is particularly vital for increasing survival rates in conditions such as cancer and cardiovascular diseases. Data mining, which involves cleaning raw data, identifying patterns, building models, and evaluating these models, is crucial for extracting valuable insights from large datasets. This process integrates statistics, machine learning, and database systems and has become increasingly important in predicting disease risk within the healthcare sector [24].

1.2 Background of Cardiovascular Disease

Heart disease is widely recognized as one of the most perilous and life-threatening chronic conditions globally [25]. Heart failure primarily arises from the narrowing and blockage of coronary arteries. These arteries, depicted in Fig. 1, supply oxygen and blood to the heart. Coronary artery disease (CAD) is the most prevalent form of heart disease, leading to symptoms such as chest pain, stroke, or heart attack. A heart attack occurs when blood flow to the cardiac muscle is significantly reduced or completely blocked. This blockage is typically due to the gradual buildup and hardening of plaque—a mixture of cholesterol, lipids, and other substances in the coronary arteries. This condition, known as atherosclerosis, becomes critical when the plaque ruptures, leading to the formation of a blood clot that obstructs blood flow [26, 27]. If the plaque rupture triggers clotting within the artery, a process called thrombosis, it further impairs blood flow to the heart muscle. As a result, the affected muscle tissue may start to die, and if the clot completely obstructs the artery, the entire area of muscle tissue below the blockage can be damaged or destroyed.

Figure 2 illustrates both a healthy and an impaired heart. When cells receive adequate nourishment, the body functions normally. However, a weakened heart fails to supply sufficient blood, leading to heart failure. This condition can cause symptoms such as fatigue, difficulty breathing, and coughing. Everyday activities, such as walking, climbing stairs, and carrying groceries, may become extremely challenging for individuals with heart failure.

Atherosclerotic plaque can reduce blood flow within the carotid artery. If this plaque ruptures, microscopic fragments and clotted blood may travel to the brain through the bloodstream. An embolus is a foreign mass that moves through the circulatory system. If it becomes lodged in a small artery in the brain, it can obstruct blood supply to a portion of the brain, leading to a condition known as atherosclerosis, as illustrated in Fig. 3.

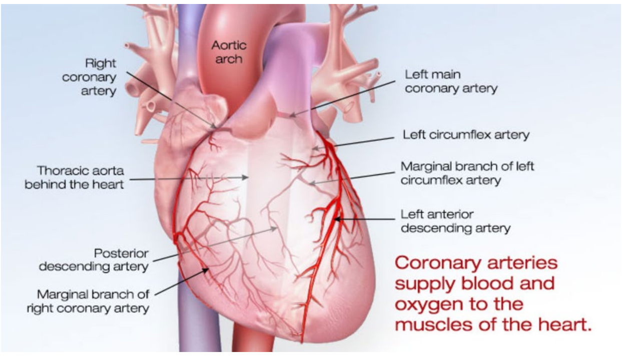


Fig. 1 Heart muscles are supplied with blood and oxygen through the coronary arteries [28]

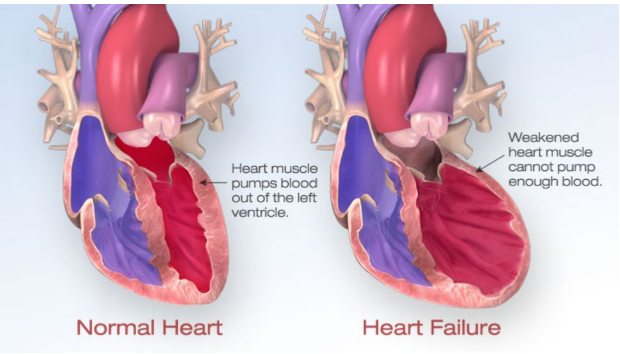


Fig. 2 Normal heart vs abnormal heart [28]

Identifying individuals at risk for cardiovascular diseases (CVDs) and providing them with timely and appropriate treatment can significantly reduce early mortality. To achieve this, it is essential for all public healthcare institutions to have access to medications for noncommunicable diseases and primary healthcare technologies, ensuring that those in need receive the necessary medical care and consultations.

1.3 Cardiovascular Disease Signs and Risk Factors

Pain or discomfort in the center of the chest, arms, left shoulder, elbows, jaw, or back can be indicative of a cardiovascular issue. Additional symptoms may include difficulty breathing or chest tightness, nausea or vomiting, light-headedness or faintness, cold sweat, and paleness. Women are more likely to experience shortness of breath, nausea, vomiting, and discomfort in the back or jaw compared to men. A common sign of a stroke is acute paralysis on one side of the face, arm, or leg. Anyone experiencing these symptoms should seek immediate medical attention.

Key behavioural risk factors for heart disease and stroke include unhealthy eating habits, physical inactivity, smoking, and excessive alcohol consumption. These behaviours can lead to elevated blood pressure, high blood glucose levels, high blood lipids, and obesity—known as "intermediate

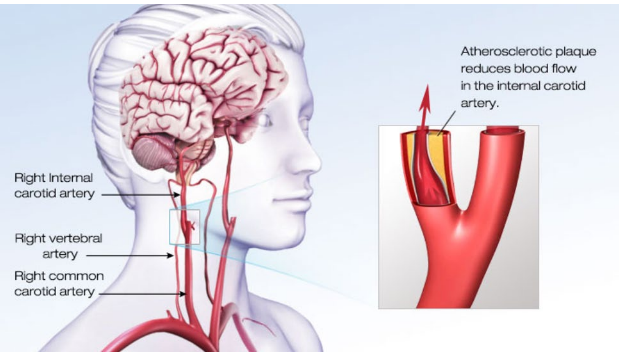


Fig. 3 Heart stroke by atherosclerosis [28]

risk factors." Such factors can be identified in outpatient settings and are associated with a higher risk of stroke, heart attack, heart failure, and other related conditions [29]. Research indicates that measures such as quitting smoking, reducing dietary salt, increasing fruit and vegetable intake, engaging in regular physical activity, and limiting alcohol consumption can effectively lower the risk of heart disease. Promoting health initiatives that provide affordable and accessible healthy options is essential for encouraging and maintaining healthy behaviours. Since many cardiovascular risk factors stem from lifestyle and behaviour, they are often modifiable and preventable.

Traditionally, diagnosing heart disease involved assessing medical history, laboratory test results, and a doctor's evaluation of symptoms. Among conventional methods, angiography is often regarded as one of the most accurate techniques for detecting cardiac issues. However, angiography has several drawbacks, including its high cost, potential side effects, and the requirement for specialized expertise in advanced technology. Moreover, conventional methods can lead to diagnostic errors and are time-consuming due to user-related issues. These techniques are also computationally intensive and take longer to analyze. To review the literature on heart disease risk prediction, databases such as Springer, ScienceDirect, IEEEXplore, Web of Science, PubMed, MDPI, Hindawi were searched using keywords like heart disease risk, heart disease prediction, data mining, data preprocessing, machine learning algorithms, ensemble classifiers, deep learning algorithms, feature selection, hyperparameter optimization techniques. This review encompasses peer-reviewed academic journals, conference papers, and reviews published from 2014 to 2024, written in English, and includes both qualitative and quantitative data. The number and percentage of publications per year included in this review are illustrated in Fig. 4.

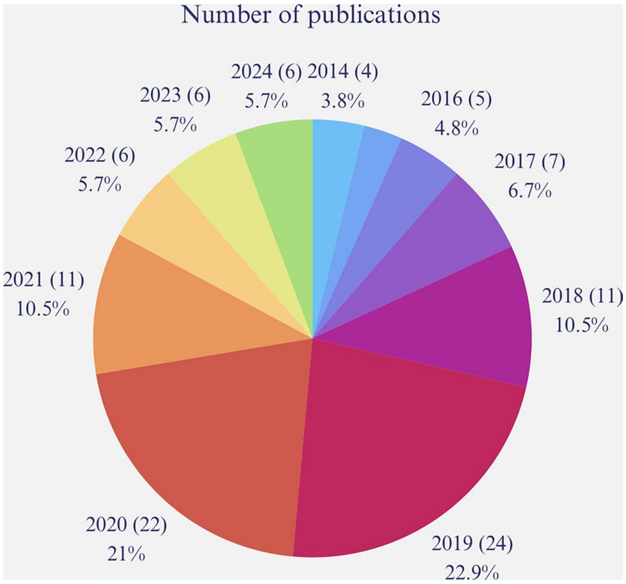


Fig. 4 Year wise number of publications included in this review

2 Related Work

This section provides state-of-the-art works on heart disease risk prediction using machine learning and deep learning algorithms. It includes utilization of machine learning algorithms, utilization of deep learning algorithms, utilization of pre-processing methods, utilization of feature selection techniques, utilization of hyperparameter tuning in HD prediction.

2.1 Utilization of Machine Learning Algorithms

The summary of machine learning algorithms (single and ensemble) utilized in predicting heart disease are provided in Table 1 along with their performance metrics, validation method and feature selection methods and the number of features selected from the heart dataset.

Bashir et al. [30] introduced the majority-vote based ensemble framework (MV5), which employs five classifiers: Naive Bayes (NB), Decision Tree based on Gini Index (DT-GI), Decision Tree based on Information Gain (DT-IG), Memory-Based Learner (MBL), and Support Vector Machine (SVM). They applied this framework to five datasets from the ML repository, including Single Proton Emission Computed Tomography (SPECT), SPECTF, Statlog, Cleveland, and Eric datasets. Missing values were addressed by replacing them with mean and mode values, noise was eliminated using a binning method, and outliers were removed based on ± 1.5 IQR. The MV5 algorithm achieved an average accuracy of 88.52%. Tomar and Agarwal [31] used the Least Square Twin Support Vector Machine (LST-SVM) for diagnosing heart disease. They employed the F-score to compute the weight of each attribute and selected significant attributes from the Statlog dataset, which contains 270 observations and 13 predictors. The dataset was divided into training and test sets in ratios of 50:50, 70:30, and 80:20, and GridSearch was used to enhance performance. By selecting 11 attributes, they achieved the best accuracy of 85.80%. El-Bialy et al. [32] focused on the Cleveland, Hungarian, Long Beach VA, and Statlog heart disease datasets, comprising 303, 294, 200, and 270 observations, respectively. They utilized the Fast Decision Tree (FDT) and pruned C4.5 tree models, training them with ten-fold cross-validation. The classification accuracy achieved on the Hungarian dataset using the C4.5 Decision Tree was 78.57%.

Table 1 Summary of state-of-the-art works on heart disease risk prediction using machine learning algorithms

Using machine learning algorithms											
Author & year	Validation	Feature selection	Classifier	#	Acc. (%)	Sen. (%)	Spe. (%)	Pre. (%)	F-mea. (%)	AUC	Limitations
Bashir et al. (2014) [30]	Hold-out (90:10)	–	Majority Vote	13	82.52	86.96	90.83	–	88.85	–	No feature selection algorithms were employed in this work. Dataset is small. (Statlog dataset)
Tomar and Agarwal (2014) [31]	Hold-out (50:50, 70:30, 80:20)	F-score	LSTSVM	11	85.59	85.71	89.13	–	–	–	No preprocessing steps were discussed in this work. (Statlog dataset)
El-Bialy et al. (2015) [32]	Tenfold CV	Decision Trees	Pruned C4.5 tree	5	78.54	–	–	–	–	–	No pre-processing steps discussed as to avoid/replace the missing values in the datasets
Long et al. (2015) [33]	Hold-out	Chaos firefly and Rough Set	IT2FLS	4	88.30	84.90	93.30	–	–	–	The datasets considered in this wok are smaller in terms of the records. The proposed method took more time to train. (Statlog dataset)
Verma et al. (2016) [34]	Tenfold CV	CFS and PSO	MLR	7	88.40	–	–	–	–	–	Prediction model evaluated by only one metric. The prediction accuracy is not up to the mark
Masetic and Subasi (2016) [35]	Hold-out	Autoregressive Burg	RF	–	100.0	–	–	–	100.0	1.000	Proposed model was leading to overfitting. (PhysioNet database)
Pouriyeh et al. (2017) [36]	Tenfold CV	–	SVM	13	84.15	89.70	–	82.70	86.00	0.836	Not discussed about missing valued instances in the Cleveland dataset
Liu et al. (2017) [37]	Jackknife	RFRS	Boosting with C4.5 DT	13	78.68	–	–	–	–	–	The attribute ‘sex’ is not included in the optimal set that is vital in describing heart disease patient on gender basis
Haq et al. (2018) [38]	Tenfold CV	Relief	LR	6	89.00	77.00	98.00	–	–	0.880	Missing valued instances removed. Non-mutable risk factor ‘age’ of the patient has not been selected by any of the feature selection technique and ignored
		mRMR	NB	6	84.00	77.00	90.00	–	–	0.840	
		LASSO	SVM-RBF	6	88.00	75.00	96.00	–	–	0.890	

Table 1 (continued)

Using machine learning algorithms											
Author & year	Validation	Feature selection	Classifier	#	Acc. (%)	Sen. (%)	Spe. (%)	Pre. (%)	F-mea. (%)	AUC	Limitations
Kale et al. (2018) [39]	Hold-out (70:30)	PCA	Fuzzy Extreme Learning	11	87.65	–	–	–	–	–	The features selected for classifications are not discussed and high computational cost
Saqlain et al. (2018) [40]	Tenfold CV	FFSA	SVM-RBF	7	81.19	72.92	88.68	–	–	–	Only one classifier SVM with RBF kernel is applied in the work. The features selected (6) from the Cleveland dataset are not disclosed
Ma et al. (2018) [41]	Tenfold CV	–	SSRFLE	13	78.68	–	–	–	–	–	No feature selection method employed and computationally expensive
Takci (2018) [42]	Tenfold CV	Backward logit	NB	4	82.96	–	–	–	–	0.881	The attribute ‘thalach’ is excluded by relief method, maximum heart rate very important attribute in determining heart disease
		Forward logit	NB	5	83.70	–	–	–	–	0.880	
		Fisher filtering	NB	9	84.81	–	–	–	–	0.917	
		Relief	SVM Linear	6	84.81	–	–	–	–	0.900	
Dulhare (2018) [43]	Cross-validation	PSO	NB	7	87.91	–	–	–	–	–	The personal features ‘age’, ‘sex’ not considered in the optimal set
Javeed et al. (2019) [44]	Hold-out (70:30)	RSA	RF	7	93.33	95.12	89.79	–	–	0.947	No pre-processing steps discussed in the work
Reddy et al. (2019) [45]	Tenfold CV	Fuzzy rough set theory	AGAFL	–	90.00	91.00	90.00	–	–	–	The pre-processing steps, validation method utilized, and the features selected are not clearly specified
Ed-daoudy et al. (2019) [46]	Hold-out (70:30)	–	RF	13	87.50	86.96	88.37	–	–	–	Considered full features of the Cleveland dataset and no feature selection has employed
Mohan et al. (2019) [47]	Not specified	–	HRFLM	13	88.40	92.80	82.60	90.10	90.00	–	The metrics precision, f-measure, and sensitivity of the proposed method are lesser comparatively with other models

Table 1 (continued)

Using machine learning algorithms											
Author & year	Validation	Feature selection	Classifier	#	Acc. (%)	Sen. (%)	Spe. (%)	Pre. (%)	F-mea. (%)	AUC	Limitations
Alam et al. [48], 2019	Tenfold CV	Relief	RF	12	85.50	85.60	–	85.50	85.50	0.915	Only one random forest classifier is used. Did not benchmark the performance with the existing research works
Raza [49], 2019	Tenfold CV	Chi-square	Majority vote	13	88.88	85.00	92.00	89.00	87.00	0.880	No feature selection technique utilized in the work
Louridi et al. (2019) [50]	Hold-out (80:20)	–	SVM linear	13	86.80	87.14	86.48	87.00	87.00	–	Only three single classifiers are applied in the work. No feature selection techniques employed
Latha and Jeeva (2019) [51]	Not specified	Brute force	Majority voting	9	85.48	–	–	–	–	–	Accuracy is only metric computed to measure the performance of the ensemble classifiers
Amin et al. (2019) [52]	Tenfold CV	Brute force	SVM and Majority vote	9	86.87	–	–	86.86	88.22	-	The attributes ‘age’, and ‘thalach’ (maximum heart rate) not included in the optimal set
Mienye et al. (2020) [53]	Hold-out (70:30)	–	ABWAE	13	93.00	91.00	–	96.00	93.00	–	Feature selection techniques were not employed in the work
Ahmed et al. (2020) [54]	Tenfold CV	Univariate	RF	7	90.00	–	–	–	–	–	The accuracy with the optimal features obtained by univariate, and relief methods is lower than that obtained with the full features
		Relief	DT	7	92.82	–	–	–	–	–	
Budholiya et al. (2020) [55]	Hold-out (80:20)	–	XGBoost	13	91.80	85.71	96.96	–	90.56	0.913	The accuracy of training set is about 6% less than that of test dataset
Gazeloglu et al. (2020) [56]	Not specified	–	SVM PolyKernel	13	85.14	–	–	–	–	0.795	Accuracy of the models is decreased when using feature selection compared to that on full features

Table 1 (continued)

Using machine learning algorithms											
Author & year	Validation	Feature selection	Classifier	#	Acc. (%)	Sen. (%)	Spe. (%)	Pre. (%)	F-mea. (%)	AUC	Limitations
Gupta et al. (2020) [57]	Hold-out (80:20)	FAMD	RF	28	93.44	89.28	96.96	–	92.59	0.931	The one-hot encoding of categorical features during pre-processing increases the number of predictor features from 13 to 28, results in more training time
Ramesh et al. (2020) [58]	Hold-out (80:20)	Information gain	SVM and RF	–	88.98	–	–	–	–	–	Accuracy is the only metric used to measure the performance of the classifiers
Pasha and Mohamed (2020) [59]	Hold-out (70:30)	NFR	LR	9	92.53	–	–	–	–	0.926	The ‘age’ of the patient, essential feature in contributing the heart disease risk, is excluded from the feature set that provided high accuracy
Mohamed et al. (2020) [60]	Tenfold CV	PPA	KNN	4	86.17	86.00	85.00	–	–	–	The selected features for prediction of heart disease are not discussed
Tama et al. (2020) [61]	Tenfold CV	CFS	Stacking ensemble	7	85.71	–	–	–	86.49	0.858	The accuracy of the proposed two-tier ensemble method is lower than the prior works mentioned. The features ‘age’, ‘sex’ not comprised in the optimal set
Muhammad et al. (2020) [62]	Tenfold CV	FCBF	Extra Tree	6	94.14	94.29	93.98	94.47	94.00	0.942	The features ‘age’, ‘sex’ and ‘thalach’ altogether not incorporated in any of the optimal feature set
		mRMR	Extra Tree	6	93.42	93.92	93.88	94.45	94.00	0.932	
		LASSO	Extra Tree	6	89.36	88.21	90.58	88.90	88.00	0.920	
		Relief	Extra Tree	6	94.41	94.93	94.89	95.46	95.00	0.942	
Aslam et al. (2021) [63]	Hold-out (70:30)	–	HRFFNRM	13	90.30	–	–	91.40	–	–	Included only 297 instances after removing the missing valued instances makes the dataset smaller
Rani et al. (2021) [64]	Tenfold CV	GA and RFE	RF	8	86.60	84.14	89.02	88.46	86.25	–	Dataset is smaller and the accuracy is not up to the mark

Table 1 (continued)

Using machine learning algorithms											
Author & year	Validation	Feature selection	Classifier	#	Acc. (%)	Sen. (%)	Spe. (%)	Pre. (%)	F-mea. (%)	AUC	Limitations
Katarya and Meena (2021) [65]	Tenfold CV	–	RF	13	95.60	97.68	–	55.28	–	–	Specificity of the proposed model is very poor
Ansarullah et al. (2022) [66]	Tenfold CV	Filter method	RF	11	84.00	85.00	83.00	85.00	–	0.850	The performance of the proposed model is not so accurate. Deep learning models not investigated. (Kashmir dataset with 5776 records)
Kanagarathinam et al. (2022) [67]	Tenfold CV	Person's Correlation	CatBoost	10	94.34	–	–	–	–	–	Only accuracy metric has been utilized to evaluate the performance of classifiers (Sathvi dataset with 531 records and 12 features)
Trigka et al. (2023) [68]	Tenfold CV	Gain Ratio and RF	Stacking ensemble	15	90.90	87.60	–	96.70	–	0.961	Limited focus on deep learning models. Dataset (with 3655 observations) chosen from Kaggle
Chandrasekhar et al. (2023) [69]	Hold-out (80:20)	–	Soft voting ensemble	12	93.44	90.00	–	90.00	90.00	–	Limited amount heart disease data used. No feature selection algorithm employed (IEEE Dataport dataset with 1190 samples)
Torthi et al. (2024) [70]	Hold-out (75:25)	BAPSO	RF	13	98.71	98.23	–	98.67	98.45	–	Dataset used in this work is small (UCI dataset with 270 records)
Elsedimy et al. (2024) [71]	Hold-out	–	QPSO-SVM	13	96.31	–	93.56	94.23	95.00	–	Random uniform noise (0–0.01) is used for replacing the missing values

Table 1 (continued)

Using machine learning algorithms											
Author & year	Validation	Feature selection	Classifier	#	Acc. (%)	Sen. (%)	Spe. (%)	Pre. (%)	F-mea. (%)	AUC	Limitations
Nissa et al. (2024) [72]	Tenfold CV	–	AdaBosst	26	95.20	95.20	95.30	98.70	96.98	–	Feature importance has been computed using AdaBoost but not provided the details of optimal features selected for the heart disease prediction. (UCI ML library with 8769 records)
Daza et al. (2024) [73]	Hold-out (80:20)	–	Stacking ensemble	12	91.50	91.60	–	91.50	91.49	0.970	Single dataset for broader patient applicability, and exclusion of lifestyle and genetic risk factors. (UCI dataset with 918 observations)
Narayanan et al. (2024) [74]	Hold-out (80:20)	RF ranking	RF	8	96.60	90.00	100.0	100.0	–	–	This work is limited to UCI heart dataset and focused only supervised ML classifiers

#—Number of features

Long et al. [33] developed interval type-2 fuzzy logic system (IT2FLS) for diagnosing heart disease using Statlog and SPECTF datasets. The Statlog dataset consists of 270 instances with 13 attributes and 1 class variable whereas SPECTF consists of 267 images with 44 attributes and 1 class variable. Min–max data normalization has been performed on datasets for translating the data in to [0, 1] range. Chaos Firefly and Rough Set algorithms have been applied to reduce the no. of attributes and trained NB, SVM, ANN, and proposed methods. On Statlog dataset achieved 88.3% of accuracy by reducing the no. of attributes to 4, whereas on SPECTF dataset achieved 87.2% of classification accuracy. Verma et al. [34] collected 335 CAD patient's non-invasive clinical data from Indira Gandhi Medical College (IGMC), India which consists of 25 features and 1 class variable. Dimensionality reduction has been performed using correlation-based feature selection (CFS) and particle swarm optimization (PSO) and selected 5 risk factors. Four ML models multilayer perceptron (MLP), multinomial logistic regression (MLR), fuzzy unordered rule induction algorithm, (FURIA), C4.5 DT, were trained. Achieved an accuracy of 84.17% using MLR. The proposed framework also employed on Cleveland dataset and achieved 88.4% of accuracy using MLR with tenfold CV. The study conducted by Masetic and Subasi [35] collected the ECG signals from the PhysioNet database and extracted the features using autoregressive (AR) Burg method. These features have been applied to 5 ML algorithms, C4.5 DT, KNN, SVM, ANN and RF, to classify heart failure patients. The performance of classifiers has been evaluated using accuracy, sensitivity, specificity, F-measure, and ROC area. Experimental results shown that RF outperformed with an accuracy of 100% and ROC area of 1.0.

Pouriyeh et al. [36] considered Cleveland data and trained both single (DT, NB, MLP. KNN, single conjugative rule learner, radial basis function and SVM) and ensemble (bagging, boosting, and stacking) classifiers with tenfold CV. SVM with boosting outperformed other classifiers with an accuracy of 84.15%, precision of 82.7%, recall of 89.7%, F-measure of 86%, and ROC area of 0.836. The study also performed tuning K value (1, 3, 9, 15) in KNN. Liu et al. [37] deployed hybrid system for heart disease classification. Data discretization has been performed on Statlog dataset using equal interval binning technique. ReliefF algorithm is used to extract the features and Rough Set algorithm is used for feature reduction. Ensemble classifier with C4.5 DT as base classifier has been trained and achieved 92.59% accuracy. Haq et al. [38] checked the performance of various machine learning algorithms; logistic regression, KNN, ANN, SVM, NB, DT, and RF on full features of the Cleveland dataset. Then selected 6 important features based on the feature weights using Relief, Minimal-Redundancy-Maximal-Relevance (mRMR), and least absolute shrinkage and selection operator (LASSO) feature selection (FS) algorithms. Noticed that the logistic regression and SVM are the best classifiers with accuracies 84% and 86% respectively before the feature selection, and 89% and 88% respectively after the feature selection. Kale et al. [39] proposed a PCA-based approach to pick the ideal subset of features for a fuzzy extreme learning machine (PFFELM) to address the weighted classification problem. The entire system is partitioned into four components: Principal Component Analysis (PCA), Feature Selection and Scaling (FSS), fuzzification, and classification. Performed simulations on clinical datasets, including Statlog cardiac data, using a 70% training and 30% testing split. By using the suggested procedure achieved an accuracy of 87.65%.

A method was developed by Saqlain et al. [40] to enhance the prediction of heart disease risk by selecting a subset of features. This was achieved by using two feature selection algorithms: the mean Fisher score-based feature selection algorithm (MFSFSA), the forward feature selection algorithm (FFSA), and the reverse feature selection algorithm (RFSA). The four UCI heart datasets, namely Cleveland, Hungarian, Switzerland, and SPECTF, were used to extract feature subsets. These subsets were then inputted into an SVM classifier based on the RBF kernel. The performance of the classifier was evaluated in terms of accuracy, specificity, and sensitivity. Ma et al. [41] devised a semi-supervised rough fuzzy Laplacian eigenmaps (SSRFLE) technique to reduce the dimensionality of extensive datasets. Constructed a collection of partially guided fuzzy similarity clusters to assess the similarity between the samples. Next, by building a neighbourhood rough fuzzy set model of the granules, the degrees of both samples of similar classes were assessed. Attained an accuracy of 78.68% from the SSRFLE feature reduction technique on Statlog dataset with tenfold cross-validation. Takci [41] proposed support vector machine linear kernel (SVML) with relief feature selection method to predict the heart attack using the Statlog heart dataset. Applied 12 machine learning classifiers and 4 feature selections method on Statlog dataset in TANGARA machine learning tool. The dataset was split into training (90%) and testing (10%) sets and validated the experiments with tenfold cross-validation. The performance of the classifiers has measured based on accuracy, processing time, and ROC curve. An accuracy of 84.81% was achieved using Support Vector Machine with ReliefF feature selection, which utilised 6 features. The K-nearest neighbours (KNN) algorithm achieved the highest Area Under the Curve (AUC) value of 0.951 when applied to the fisher filtering features.

Dulhare [43] proposed NB classifier and particle swam optimization (PSO) feature selection technique to predict the heart disease risk on Statlog dataset of 270 observations. Based on the PSO value selected 7 features (cp, restecg, thalach, oldpeak, ca, thal, target) from the dataset and trained the NB classifier. Concluded that the NB + PSO achieved an accuracy of 87.91% that is accounting to 8.79% of improvement comparatively without feature selection (on full features). Javeed et al. [44] developed a streamlined and less intricate model to enhance the accuracy of heart disease prediction on the Cleveland dataset. They achieved this by employing the random search algorithm (RSA) in conjunction with the random forest (RF) model. 70% of the dataset is used for training, while the remaining 30% is used to test the suggested strategy. Attained a precision rate of 93.33% by utilising a subset of 7 characteristics, resulting in a 3.3% enhancement compared to the normal random forest. The enhanced version of the RSARF model has an AUC of 0.947. Reddy et al. [45] introduced an adaptive genetic algorithm with fuzzy logic (AGAFL) model for early diagnosis of heart disease. Selected significant features using rough set theory and applied adaptive genetic algorithm to optimize the rules generated from fuzzy classifiers for better accuracy on Cleveland, Hungarian, and Switzerland datasets. Acquired 90% accuracy, 91% sensitivity, and 90% specificity with Cleveland dataset.

Ed-daoudy et al. [46] conducted research on a dataset consisting of 303 records obtained from the UCI repository, each including 14 attributes. Assessed the effectiveness of the four primary classifiers, namely SVM, DT, RF, and LR, by utilising Apache Spark together with its MLlib machine learning framework. Evaluated the classifiers by considering their accuracy, specificity, sensitivity, and the time required to build and evaluate the models. With an accuracy rate of 87.5%, the random forest approach outperforms the others, according to the results. Mohan et al. [47] used the Cleveland dataset to create an efficient hybrid random forest with a linear model (HRFLM) that improved the accuracy of heart disease prediction. Utilised machine learning classifiers including Naive Bayes (NB), generalised linear model (GLM), logistic regression (LR), deep learning (DL), decision tree (DT), random forest (RF), gradient boosting tree (GBT), support vector machine (SVM), and voting ensemble (VOTE) to train on all 13 attributes of the dataset. The hybrid model is constructed using only three existing models: DT, RF, and LM, along with their respective classification rules. The analysis determined that the RF and LM methods yielded the most favourable error rate. The proposed HRFLM technique achieved a peak accuracy of 88.4%. Alam et al. [48] developed a feature selection technique for classifying medical data using a ranking-based approach. Next, we utilised an RF classifier to analyse the selected features of 10 different illness datasets, which included the Statlog heart data. We performed a tenfold cross-validation to ensure accurate results. The feature ranking algorithms utilised in the Weka tool include InfoGainAttributeEval, GainRatioAttributeEval, CorrelationAttributeEval, OneRAtrributeEval, and ReliefFAttributeEval. Three models were created based on the rankings acquired from the attribute evaluators. It was determined that the RF algorithm, when combined with the ReliefFAttributeEval method (excluding feature 5), outperformed other strategies with an accuracy rate of approximately 85.5% and an Area Under the Curve (AUC) value of 0.915.

Raza [49] developed an ensemble classifier by merging Logistic Regression, Multilayer Perceptron, and Naïve Bayes algorithms. The classifier was trained using majority voting to enhance the accuracy of predictions. Conducted exploratory data analysis by constructing a decision tree using the chi-squared automated interaction detector (CHAID) on a Statlog dataset consisting of 270 observations. Attained a precision rate of 89% and an AUC (Area Under the Curve) value of 0.88. Louridi et al. [49] suggested a method to determine if cardiac disease is present or absent by preprocessing the data and substituting the mean values for the missing values. Trained 3 machine learning algorithms NB, SVM (linear and radial basis function), and KNN by splitting the UCI heart disease dataset into 50:50, 70:30, 75:25, 80:20 training and testing ratios. SVM-linear provided an accuracy of 86.8% with an 80:30 ratio, SVM-RBF postulated that of 62.77% with 75:25 ratio, KNN provided that of 85.83% with 50:50 ratio, and NB offered that of 86.11% with 80:20 ratio. Latha and Jeeva [51] performed a comparative analysis with NB, Bayes Net (BN), C4.5, Multilayer perception (MLP), and Projective Adaptive Resonance Theory (PART) classifications. C4.5, MLP, and PART are found to be weaker than others. Later, ensemble techniques such as boosting, bagging, stacking, and majority vote are applied to improve the performance compared to individual classifiers. The ensemble technique resulted in a maximum gain in accuracy of 7.26% for a weak classifier. This was achieved by employing a majority vote with the NB, BN, RF, and MLP classifiers, which had an overall accuracy of 85.48%.

Amin et al. [52] aimed to identify the essential attributes to improve the performance of machine learning models in predicting the cardiovascular disease. Applied brute force method to test each possible combination of attributes using seven ML classifiers KNN, DT, NB, LR, SVM, Neural Network, and Vote (NB + LR). Collected the Cleveland heart dataset, removed 6 instances with missing values and converted multiclass ‘target’ attribute into binary during preprocessing. Trained the 7 algorithms by exploiting a total of 8100 combination of the attributes with tenfold cross-validation. Concluded that sex, cp, fbs, restecg, exang, oldpeak, slope, ca, and thal are the 9 most important features based on the number of times occurred and Vote, NB, and SVM are the top three ML classifiers. Finally conducted evaluation test using Statlog dataset. Achieved an accuracy of 86.3%, and 87.41% with 13 and 9 attributes respectively, using Vote. Mienye et al. [53] created an ensemble method from the CART models using an accuracy based weighted aging classifier ensemble (AB-WAE) on Cleveland (303 instances) and Framingham (4238 instances) datasets. Implemented and assessed the performance of various machine learning algorithms including K-Nearest Neighbours (KNN), Logistic Regression (LR), Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), Classification and Regression Tree (CART), Gradient Boosting (GB), and Random Forest (RF). The evaluation was based on metrics such as accuracy, sensitivity, precision, and F1 score. The proposed approach yielded an accuracy of 93% and 91% for the Cleveland and Framingham datasets, respectively. Ahmed et al. [54] performed two feature selection techniques Univariate, and Relief to select best features from Cleveland heart dataset 2016 and trained DT, SVM, RF, and LR classifiers with full features and selected features. Developed an online heart disease predictive system using Twitter by collecting the real-time data from the user tweets. Seven features thalach, oldpeak, ca, cp, exang, age, chol are selected from Univariate, and thal, exang, cp, ca, slope, thalach, oldpeak are selected from Relief methods. Computed accuracy for each case by performing hyperparameter tuning and tenfold cross-validation. Unfortunately, claimed best accuracy of 94.9% using full features with RF method rather feature selection methods.

Budholiya et al. [55] projected XGBoost classifier with Bayesian optimization for heart disease prediction. Utilised one-hot (OH) encoding to transform the categorical attributes of the Cleveland dataset in order to enhance the accuracy of predictions. The expected model was compared to random forest and additional tree classifiers using multiple assessment measures including specificity, accuracy, F1-score, sensitivity, and AUC. Concluded that the experimental results of the proposed method proved to be reliable in heart disease risk prediction with the accuracy of 91.8% and AUC of 0.9134. The one-hot encoding increases the number of features, missing valued instances maintained as a separate feature column. Gazeloglu et al. [56] utilised 18 machine learning models and 3 feature selection strategies (Correlation-based FS, Chi-Square, and Fuzzy Rough Set) to determine the optimal combination for predicting heart disease diagnosis on the Cleveland dataset. The SVM model with a polynomial kernel achieved an accuracy of 85.148% without any feature selection. The Naïve Bayes model achieved an accuracy of 84.818% with feature selection using the CFS algorithm, selecting 6 features. The radial basis function (RBF) network achieved an accuracy of 81.188% with feature selection using the Fuzzy Rough Set and Chi-Square algorithms, selecting 7 features. Gupta et al. [57] assessed the effectiveness of different machine learning algorithms, including LR, KNN, SVM, DT, and RF, using factor analysis of mixed data (FAMD) technique. Obtained 28 characteristics by calculating the Pearson correlation coefficient from the Cleveland dataset. Based on the weight matrix, the RF model has attained the highest accuracy of 93.44%, with sensitivity and specificity values of 89.28% and 96.96% respectively. It is recommended to take into account class unbalanced datasets when addressing real-life situations.

Ramesh et al. [58] proposed information gain feature selection (IGFS) algorithm to improve the machine learning classifiers performance on the Cleveland heart dataset from UCI ML repository. Divided the collected dataset into two part 80% for training and 20% for testing. IGFS has applied on the training dataset to retrieve significant features. Trained various ML algorithms such as KNN, DT, LR, NB, SVM, and RF without and with feature selection. Concluded that SVM and RF achieved 88.9823% and 88.9812% respectively using IGFS, that shows an improvement of about 1.3% compared to without feature selection. Pasha and Mohamed [59] introduced a new model called the feature reduction (NFR) model, which aims to improve the accuracy of heart disease risk prediction on the Cleveland, Hungarian, Statlog, and Switzerland datasets. The performance of LR, RF, BRT, SGB, and SVM algorithms was assessed by lowering the features using different subset combinations. The evaluation was centred on achieving maximum accuracies, AUC values, and minimising the difference between them. The proposed model, utilising logistic regression with 9 attributes, achieved an accuracy of 92.53% and an AUC of 92.68%. Mohamed et al. [60] proposed a meta-heuristic approach known as parasitism-predation algorithms (PPA) to enhance classification accuracy. This strategy combines cat swarm optimisation (CSO), cuckoo search (CS), and crow search algorithm (CSA) to pick a feature subset. Attained an accuracy of 86.17% within a time frame of 49.13 s using the K-Nearest Neighbours (KNN) algorithm. This was accomplished by choosing 4 characteristics from the Statlog heart dataset and employing tenfold cross-validation.

Tama et al. [61] developed a stacked architecture by combining RF, gradient boosting, and extreme gradient boosting with the CFS feature selection approach. They used the particle swarm optimisation (PSO) search method to optimise the architecture on different heart disease datasets. The suggested technique achieved an accuracy of 85.71%, an F1 score of 86.49%, and an AUC of 0.8586 on the Cleveland dataset, which consists of 7 features. These results were obtained using tenfold cross-validation. The accuracy achieved on the Statlog dataset, which consists of 8 features, was 91.18%. Muhammad et al. [62] have developed an advanced computer model that can accurately predict and diagnose cardiac disease at an early stage. Ten different machine learning classifiers, including KNN, DT, Extra Tree, RF, LR, NB, ANN, SVM, Adaboost, and Gradient Boosting (GB), are trained to predict heart disease early. This is done using both the entire and optimal features of the Cleveland and Hungarian datasets, and tenfold cross-validation is used. Subsequently, four feature selection techniques, including fast correlation-based filter (FCBF), mRMR, LASSO, and relief, were employed to acquire the crucial and highly correlated features. Using feature selection strategies such as FCBF, mRMR, LASSO, and relief, we achieved accuracies of 94.14%, 93.42%, 89.36%, and 94.41%, respectively, while using the Extra Tree classifier. Aslam et al. [63] introduced a new model that utilises Hierarchical Random Forest Formation with Nonlinear Regression Model on the Cleveland dataset. The data has been preprocessed and split the dataset into 70:30 for training and testing, respectively. The bottom-up technique was employed with Artificial Neural Networks (ANN) to extract notable characteristics, while a mixture of Support Vector Machines (SVM) and the Apriori algorithm was utilised for classification. The HRFFNRM technique achieved an accuracy of 90.3%, precision of 91.4%, and a misclassification rate of 8.74%.

Rani et al. [64] developed a system for early prediction heart disease based on hybrid feature selection (FS) with the combination of GA and recursive feature elimination (RFE). Considering Cleveland dataset, multivariate imputation by chained equations (MICE) has been performed to replace the missing values in the dataset, then applied SMOTE and standard scaling method were used to pre-process the data. Eight features were selected using hybrid FS method. Then, SVM, NB, LR, RF and AdaBoost classifiers were trained on the data and achieved 86.6% prediction accuracy through RF. Katarya and Meena [65] provided a comparison analysis on machine and deep learning techniques for heart disease. They've trained various ML (LR, KNN, SVM, NB, DT, and RF)and DL (ANN, MLP, and DNN) algorithms on pre-processed UCI dataset and the performance of these algorithms has been evaluated using accuracy, root mean squared error, mean absolute error, precision, and recall. RF has reached best accuracy of 95.60%. Ansarullah et al. [66] utilized various feature selection techniques (filter, wrapper and embedded) for early-stage evaluation of heart disease. The study collected heart dataset from various heterogeneous sources in Kashmir, that consist of 5776 patient records. RF, NB, KNN, SVM and DT classifiers were trained on preprocessed data. RF outperforms the others with an accuracy of 84%.

Kanagarathinam et al. [67] proposed Nb, KNN, SVM, MLP, XGBoost, and CatBoost algorithms for predicting heart disease. This study chosen a dataset named 'Sathvi' with 531 observations and 12 features. Applied Pearson correlation coefficient method to select 10 significant features based on the threshold value 0.5. After training ML classifiers achieved 94.34% accuracy with tenfold CV using CatBoost ensemble. Trigka et al. [68] collected a coronary dataset from Kaggle that consists of 3655 observations, 15 features, and 1 binary target class. SMOTE technique has been employed to balance the dataset with respect to target class. Gain Ratio and Random Forest method were utilized for ranking the features of the dataset. Numerous single (NB, LR, MLP, KNN. RF, J48) and ensemble (RotF, Stacking, Voting, Bagging) ML models were trained on the pre-processed dataset. The stacking ensemble reached best prediction accuracy 90.9%, precision 96.7%, recall 87.6%, AUC 0.961. Chandrasekhar et al. [69] used six ML algorithms (RF, KNN, LR, NB, gradient boosting, and AdaBoost) for heart disease risk prediction. This study has been carried out on Cleveland and IEEE Dataport datasets. To improve the performance of the classifiers, the study also employed GridSearchCV hyperparameter optimization with fivefold CV. AdaBoost outperformed other classifiers with the prediction accuracy of 90.16% and 90% on Cleveland and IEEE Dataport, respectively. The results also shown that significant improvement in classifiers performance by applying soft voting ensemble method with 93.44% accuracy on Cleveland dataset, and 95% on IEEE Dataport dataset.

Torthi et al. [70] proposed hybrid method for early and accurate prediction of heart disease using bat algorithm (BA), particle swarm optimization (PSO) and RF methods. This work utilized UCI heart dataset with 270 observations and 14 attributes. The data has been normalized using min–max method. The study reached utmost accuracy of 98.71% using proposed BAPSO-RF method by employing feature selection and GridSearchCV hyperparameter tuning. Elsedimy et al. [71] implemented a novel model for heart disease prediction using quantum-based particle swarm optimization (QPSO) and SVM classifier. Min–max method for scaling the data. The parameters of the QPSO-SVM model were tuned using self-adaptive threshold techniques. The model has been applied on Cleveland dataset after pre-processing and reached the accuracy of 96.31%. Nissa et al. [72] developed a framework for cardiovascular disease prediction using CatBoost, RF, GradientBoost, Light GBM, and AdaBoost ensemble techniques. This study assessed on dataset (with 8769 records and 26 features) collected from UCI ML library. Different pre-processing steps such as removal of noise, outlier, and normalization, were performed and trained ensemble techniques using tenfold CV. They've reached best accuracy of 95.2% with the AdaBoost.

Daza et al. [73] developed 4 stacking ensemble methods for heart disease diagnosis based on optimal hyperparameters. This study utilized UCI dataset that consists of 918 observations and 12 features. Hyperparameter tuning of 11 ML classifiers has been performed using GridSearchCV with CV = 5. Achieved prediction accuracy of 91.5% using stacking 2 ensemble with oversampling. Naraynan et al. [74] employed oversampling technique, SMOTE, to overcome the data imbalance issues and implemented an efficient ML technique for cardiac disease prediction. This study chosen Cleveland dataset and trained RF, NB, DT, SVM, KNN, LR, and AdaBoost techniques. Eight most significant feature have been selected using RF feature ranking method and reached an accuracy of 96.6% with SMOTE. Concluded that he results have been improved with feature selection and SMOTE.

2.2 Utilization of Deep Learning Algorithms

Artificial neural networks (ANNs), particularly deep learning methods, are extensively employed for heart disease risk prediction due to their ability to effectively capture complex relationships within data. These models excel at analyzing intricate patient information, which can lead to earlier diagnoses and better management of cardiovascular conditions. Deep learning algorithms are highly adaptable, enabling them to learn from large datasets and evaluate multiple risk factors associated with heart disease. Table 2 summarizes cutting-edge research on deep learning algorithms for heart disease risk prediction. Samuel et al. [75] utilized a fuzzy analytic hierarchy process (fuzzy-AHP) to determine the global weights of heart failure attributes, subsequently training an ANN classifier to efficiently identify patients at risk of heart disease. The system's performance was assessed by measuring accuracy, sensitivity, specificity, false positive rate, false negative rate, and the ROC curve. Their method achieved a prediction accuracy of 91.1% by dividing the Cleveland dataset of 297 samples into training, validation, and testing sets in a 65:20:15 ratio. Ali et al. [76] developed a computerized diagnostic system for heart disease using chi-square to select the most relevant features and optimized deep neural networks for classification. They employed the Cleveland heart dataset, splitting it into a 70:30 ratio for training and testing. Their hybrid approach demonstrated an accuracy of 93.33% and an AUC of 0.94.

Javeed et al. [77] proposed floating window with adaptive size for feature elimination (FW AFE) technique for feature refinement and ANN, DNN for classification of heart disease on Cleveland dataset. Appraised the performance of the classifiers using accuracy, sensitivity, specificity, MCC, and ROC. Achieved 93.33% of accuracy with FW AFE + DNN method. Das et al. [78] developed a novel hybrid method called neuro-fuzzy (NF) model with post-feature reduction to extract the essential features from the various biomedical datasets. To overcome the issues of increased model complexity by fuzzy expansion of input features based on class labels, selected appropriate fuzzified transformed features by applying linear discriminant analysis (LDA) feature reduction technique. Finally, the reduced fuzzified features are passed through the neural network model for the classification of disease. Six machine learning models (ANN, ANN-PCA, ANN-LDA, NF, NF-PCA, NF-LDA) were trained on 12 biomedical datasets including Statlog Heart data. With the proposed NF-LDA method achieved an accuracy of 85.83% on Statlog heart dataset.

Tougui et al. [79] conducted a comparison of the performance of six machine learning algorithms (LR, SVM, KNN, ANN, NB, and RF) in classifying heart disease. They used six commonly used data mining tools: Orange, Weka, RapidMiner, Knime, MATLAB, and Scikit-Learn. By choosing UCI Cleveland dataset with 303 instances, 13 features and one target variable, pre-processed to remove the missing values in the dataset. Then, applied the six ML techniques with tenfold cross validation to sample the dataset, and extracted the confusion matrices to calculate the three performance metrices accuracy, sensitivity, and specificity. Obtained the highest accuracy of 85.86%, and sensitivity of 83.94% with MATLAB's ANN technique whereas, RapidMiner's SVM method provided highest specificity of 94.38%. Mehmood et al. [80] developed a framework named CardioHelp to predict the cardiovascular disease using deep convolutional neural networks (CNN). The researchers utilised a dataset from the UCI ML repository and implemented the least absolute shrinkage and selection operator (LASSO) to both regularise the data and choose the most relevant features. The suggested approach achieved 96.7 percent accuracy, 97.06% precision, 96.35% recall, and 96.7% F1-score.

Nancy et al. [81] developed an IoT based smart healthcare system for predicting HD risk using deep learning. The study considered two heart datasets Cleveland and Hungarian with 303 and 294 instances, respectively. The number of instances has been increased to 100000 using Mockaroo tool. Various pre-processing steps were performed to make the dataset effective. The data has been divided into 70:30 for training and testing purposes. A deep learning model, bidirectional long short-term memory (Bi-LSTM) was implanted and reached 98.86% accuracy, 98.9% precision, 98.8% sensitivity, 98.89% specificity and 98.6% F-measure. Vayadande et al. [82] utilized both machine learning and deep learning techniques for heart disease risk prediction on full features of the Cleveland heart dataset. The dataset features have been analyzed using various plots such as histogram, box and heatmap to get good insights about the dataset. An accuracy of 88.52% has been achieved using the LR, RF, and XGBoost classifiers, while the MLP, a deep learning method, provided an accuracy of 86.89%, which is about 2% less than machine learning methods.

Garcıa-Ordas et al. [83] applied deep learning (CNN) on combined heart dataset (918 observations, 11 features) for predicting the heart disease. This work utilized Sparse Autoencoder (SAE) for feature augmentation. Results shown that 90.09% of accuracy with the proposed model. Almazroi et al. [84] established a decision support system for predicting heart disease using Dense Neural Network (DNN). This study collected 4 datasets Cleveland, Switzerland, Hungarian, and Long Beach and performed various pre-processing

Table 2 Summary of state-of-the-art works on heart disease risk prediction deep learning algorithms

Using deep learning algorithms											
Author & year	Cross validation	FS method	Classifier	#	Acc. (%)	Sen. (%)	Spe. (%)	Pre. (%)	F-mea. (%)	AUC	Limitations
Samuel et al. (2017) [75]	Hold-out (65:20:15)	Fuzzy-AHP	ANN	–	91.10	100.0	84.0	–	–	–	Small number of samples used for training, validation, and testing. Prediction time for clinical applications not investigated
Ali et al. (2019) [76]	Hold-out (70:30)	Chi-square	Optimized DL network	11	93.33	85.36	100.0	–	–	0.940	Smaller dataset size of 297 observations after removing the missing valued instances
Javeed et al. (2020) [77]	Hold-out (70:30)	FWAFE	Deep Neural Networks	–	93.33	85.36	100.0	–	–	0.940	The testing set accuracy 83.57% is about 10% lower than that of training set
Das et al. (2020) [78]	Hold-out	NFLDA	Neural Networks	–	85.83	59.00	–	86.40	69.70	–	Number of features are selected are not specified. The model is computationally expensive
Tougui et al. (2020) [79]	tenfold CV	–	ANN	13	85.86	83.94	87.50	–	–	–	No feature selection has employed. Removed the missing values and considered the small dataset with 297 instances
Mehmood et al. (2021) [80]	tenfold CV	LASSO	CNN	13	97.00	96.35	-	97.06	96.70	–	Imbalance in dataset affects variable selection with LASSO regression
Nancy et al. (2022) [81]	tenfold CV	–	Bi-LSTM	13	98.86	98.80	98.89	98.90	98.60	–	No feature selection methods were investigated in this study
Vayadande et al. (2022) [82]	Hold-out	–	Multi-layer Perceptron	13	86.89	–	–	–	–	–	The accuracy of deep learning techniques is lower than that of machine learning techniques. No feature selection method has employed
Garcıa-Ordas et al. (2023) [83]	tenfold CV	Sparse Autoencoder	CNN	11	90.09	–	–	–	–	–	Proposed method is computationally expensive as the more number of neurons were employed (Combined dataset with 918 samples)

Table 2 (continued)

Using deep learning algorithms											
Author & year	Cross validation	FS method	Classifier	#	Acc. (%)	Sen. (%)	Spe. (%)	Pre. (%)	F-mea. (%)	AUC	Limitations
Almazroi et al. (2023) [84]	tenfold CV	–	DNN	13	83.03	90.90	69.27	–	87.37	–	No feature selection methods were implemented, and the obtained accuracy is not up to the mark

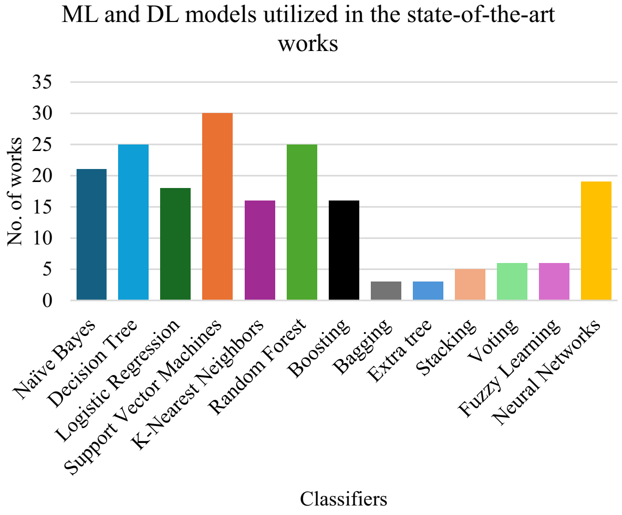


Fig. 5 ML and DL models utilized in the state-of-the-art works

steps such as imputation, outlier detection and removal, standardization. They trained both ML and DL models using tenfold CV on all the datasets. The study reached maximum accuracy of 83.03% using DDN on Hungarian dataset.

Figure 5 shows the statics of the various ML and DL models utilized in the state-of-the-art research works for heart disease risk prediction. The classifiers NB, DT, LR, SVM, KNN, RF, Boosting and NN stood on top 8 classifiers. SVM classifier mounted on top with 30 works, followed by DT and RF with 25 works, NB accounting to 21, LR with 18, KNN with 16 and NN with 19 works.

2.3 Utilization of Pre-processing Methods in Disease Prediction

Data pre-processing is a vital step in data analysis, as it involves cleaning, normalizing, and transforming raw data into a format suitable for analysis. This process is crucial for maintaining the accuracy, consistency, and completeness of the data used in analytical tasks. Effective data pre-processing can significantly improve data quality and enhance the performance of machine learning algorithms in predicting various diseases, such as cardiac conditions, liver diseases, diabetes, and more. Figure 6 illustrates the various data pre-processing techniques employed in cutting-edge research.

Data imputation is the process of filling in estimated values for missing or incomplete data to ensure the dataset is complete for analysis. Data integration involves merging data from different sources that share the same attributes to create a unified dataset, thus increasing the number of observations. Data normalization adjusts the data to a specific range, usually between 0 and 1, while data standardization transforms the dataset to have a mean of 0 and a standard deviation of 1, thereby rescaling the data to match these

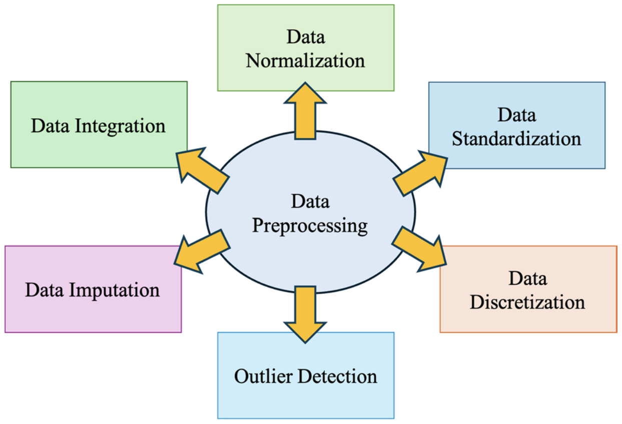


Fig. 6 Data pre-processing methods utilized in state-of-the-art works

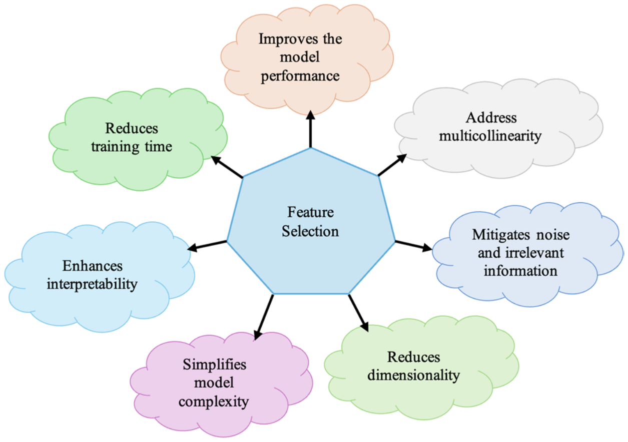


Fig. 7 Significance of feature selection

statistical properties. Data discretization involves converting continuous data into discrete intervals or categories. To identify outliers, the interquartile range (IQR) method is used: calculate the difference between the 75th percentile (Q3) and the 25th percentile (Q1) to determine the IQR. By multiplying the IQR by a factor (e.g., 1.5) and adding or subtracting this value from Q1 and Q3, respectively, you establish the upper and lower limits for detecting outliers.

Reddy et al. [85] combined five heart disease datasets—Cleveland, Hungarian, Long Beach VA, Switzerland, and Statlog—to create a larger dataset with approximately 1190 instances. Yadav et al. [86] and Almustafa [87] aggregated 1025 observations from four datasets: Cleveland, Hungarian, Long Beach VA, and Switzerland. Burse et al. [88] recommended data normalization for predicting heart disease using neural networks. Abdar et al. [89] applied normalization to the Z-Alizadeh Sani dataset to enhance coronary artery disease prediction. Uddin et al. [90] used normalization on a diabetes dataset and employed the Synthetic Minority Over-sampling Technique (SMOTE) to balance the class variable. Long et al. [33] applied Min–Max normalization to heart disease data to scale it within the [0, 1] range. Saboor et al. [91] reported significant improvements in heart disease prediction through data standardization. Md et al. [92] performed standardization on skewed columns of a liver disease dataset to boost machine learning performance and used multivariate imputation for missing values. Paul et al. [93] replaced missing values with the mean or median of the respective attribute during pre-processing. Alaa et al. [94] used data imputation on the UK Biobank cardiovascular disease dataset. Amin et al. [95] addressed missing values and outliers in a liver disease dataset. Fitriyani et al. [96] employed the isolation forest method for outlier detection and removal in a diabetes dataset and used Density-Based Spatial Clustering of Applications with Noise (DBSCAN) for the same purpose in another study [97]. Tomov and Tomov [27] developed Heart Evaluation for Algorithmic Risk-reduction and Optimization Five (HEARO-5) to automatically handle outliers and missing values. Liu et al. [37] applied data discretization for heart disease diagnosis.

2.4 Utilization of Feature Selection Techniques

Figure 7 illustrates the importance of feature selection in disease risk prediction using machine learning algorithms. By focusing on the most relevant features and reducing overfitting, feature selection enhances the performance of machine learning models. It improves efficiency by significantly cutting down training time and simplifies the model by concentrating only on essential features. This approach not only makes the model easier to interpret and understand but also facilitates maintenance and deployment in production environments. Additionally, feature selection reduces data dimensionality, which can lead to better generalization and decreased processing requirements. It also addresses multicollinearity by removing redundant or highly correlated features, resulting in more robust models. By filtering out noisy or irrelevant features, feature selection can further boost model accuracy. Table 3 provides an overview of the features selected in cutting-edge research.

From Table 3, the most selected features for the heart disease risk prediction by the research works are chest pain (cp), the number of major vessels colored by fluoroscopy (ca), followed by thallium scan (thal), exercise-induced angina (exang), the slope of the peak exercise ST segment (slope) and ST depression induced by exercise relative to rest. Though the number of features selected by the works is identical, the features differ from one work to another. El-Bialy et al. [32] selected only five features by implementing decision trees on the Cleveland dataset. Verma et al. [34] collected 335 CAD patient's non-invasive clinical data from Indira Gandhi Medical College (IGMC), India which

Table 3 Features selected by the state-of-the-art works

Author	Dataset used	Feature Selection	#	Features selected
El-Bialy et al. (2015) [32]	Cleveland	Decision Trees	5	cp, age, ca, thal, thalach
Verma et al. (2016) [34]	IGMC	CFS + PSO	5	smoking, diabetes mellitus, High Density Lipoprotein, Duke Tread Mill Score, Duration recovery with ST changes
	Cleveland		7	cp, thalch, exang, old peak, slope, ca, thal
Liu et al. (2017) [37]	Statlog	RFRS	7	age, cp, restecg, thalach, slope, ca, thal
Haq et al. (2018) [38]	Cleveland	Relief	6	thal, exang, cp, slope, ca, thalach
		mRMR	6	cp, chol, slope, ca, sex, thal
		LASSO	6	sex, ca, exang, cp, slope, thal
Takci (2018) [42]	Statlog	Backward logit	4	cp, oldpeak, ca, thal
		Forward logit	5	thal, ca, exang, oldpeak, cp
		Fisher filtering	9	thal, ca, exang, thalach, oldpeak, cp, slope, sex, age
		Relief	6	sex, thal, restecg, ca, cp, exang
Dulhare (2018) [43]	Statlog	PSO	7	cp, restecg, thalach, exang, oldpeak, ca, thal
Alam et al. (2019) [48]	Statlog	Relief	12	All features except ‘chol’
Latha et al. (2019) [51]	Cleveland	Brute Force	9	age, sex, cp, chol, fbs, exang, oldpeak, slope, ca
Amin et al. (2019) [52]	Cleveland	Brute Force	9	age, sex, cp, chol, fbs, exang, oldpeak, slope, ca
	Statlog	Brute Force	9	sex, cp, fbs, restecg, exang, oldpeak, slope, ca, thal
Ahmed et al. (2020) [54]	Cleveland	Univariate	7	thalach, oldpeak, ca, cp, exang, age, chol
		Relief	7	thal, exang, cp, ca, slope, thalach, oldpeak
Gazeloglu et al. (2020) [56]	Cleveland	CFS	6	cp, trestbps, thalach, exang, oldpeak, ca
		FRS and Chi-square	7	age, cp, trestbps, chol, thalach, oldpeak, ca
Pasha et al. (2020) [59]	Cleveland	NFR	9	ca, slope, sex, cp, thalach, thal, exang, oldpeak, restecg
Ali et al. (2019) [76]	Cleveland	Chi-square	11	age, sex, cp, trestbps, restecg, thalach, exang, oldpeak, slope, ca, thal
Tama et al. (2020) [61]	Cleveland	CFS	7	cp, restecg, thalach, exang, oldpeak, ca, thal
	Statlog	CFS	8	sex, cp, restecg, thalach, exang, slope, ca, thal
Muhammad et al. (2020) [62]	Cleveland	FCBF	6	thal, cp, exang, ca, slope, restecg
		mRMR	6	slope, cp, oldpeak, ca, sex, thal
		LASSO	6	sex, ca, exang, cp, slope, thal
		Relief	6	cp, age, fbs, ca, thalach, thal
Rani et al. (2021) [64]	Cleveland	GA and RFE	8	sex, cp, restecg, exang, oldpeak, slope, ca, thal
Kanagarathinam et al. (2022) [67]	Sathvi	Pearson Correlation	10	age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, and oldpeak
Narayanan et al. (2024) [74]	Cleveland	RF ranking	8	cp, ca, thal, thalach, oldpeak, age, trestbps, chol

consists of 25 features and 1 class variable. Dimensionality reduction has been performed using correlation-based feature selection (CFS) and particle swarm optimization (PSO) and selected 5 risk factors. Achieved an accuracy of 84.17% using multinomial logistic regression (MLR). The proposed framework also employed on 7 selected features from Cleveland dataset and achieved 88.4% of accuracy using MLR with tenfold CV. Liu et al. [37] applied ReliefF and rough set (RFRS) method to the Statlog dataset and selected seven features. Takci [42] selected four features using backward logit, five features using Forward logit, nine features using fisher filtering, and six features using relief methods from the Statlog dataset. Haq et al. [38] applied relief, minimum redundancy maximum relevance (mRMR), and least absolute shrinkage and selection operator (LASSO) feature selection methods on the Cleveland dataset and selected six features from each method. Dulhare [43] employed particle swarm optimization (PSO) method on the Statlog dataset to select seven optimal features.

Amin et al. [52] and Latha et al.[51] utilized the brute force method and selected the same 9 nine features of the Cleveland dataset. Amin et al. [52] also employed Brute Force on the Statlog dataset and selected nine different features from Cleveland. Alam et al. [48] chosen all the features except ‘chol’ from the Statlog dataset using relief method. Ali et al. [76] selected 11 features by applying chi-square method. Saqlain et al. [40] utilized forward feature selection algorithm (FFSA), and Javeed et al. [44] applied random search algorithm (RSA) and selected seven features from the Cleveland dataset but did not specify the features they selected. Mohamed et al. [60] used the parasitism predation algorithm (PPA) and selected four features from the Statlog dataset. Gazeloglu et al. [56] selected six features from CFS and seven features from fuzzy rough set (FRS) and chi-square methods on the Cleveland dataset. Ahmed et al. [54] applied univariate and relief methods to select seven significant features from each method on the Cleveland dataset. Pasha et al. [59] chose nine optimal features from the Cleveland dataset using novel feature reduction (NFR) method based on accuracy and area under the ROC curve (AUC) values. Tama et al. [61] implemented the CFS method on the Cleveland dataset and reduced the 46% of feature space by selecting seven significant features in their work. Also employed on Statlog dataset and selected eight features from that. Muhammad et al. [62] selected six significant features through the fast correlation-based Filter (FCBF), mRMR, LASSO, and Relief methods on the Cleveland dataset for the early prediction of heart disease. Figure 8 shows the utilization of feature selection methods in heart disease risk prediction.

However, the inclusion of irrelevant features can lead to model overfitting, and selecting pertinent features becomes challenging due to the vast number of possibilities. As the number of features grows, the size of the dataset increases exponentially, making a comprehensive and efficient search technique crucial for effective feature selection. It is important to conduct a thorough evaluation of features rather than relying solely on those suggested by feature selection algorithms. For instance, many studies overlook static features such as 'age' and 'sex' when determining the optimal set of features. In heart disease datasets, the presence of irrelevant or highly correlated features can cause machine learning classifiers—such as Decision Trees (DT), Naive Bayes (NB), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN)—to overfit. Thus, selecting the most relevant features remains a complex challenge due to the extensive range of potential features to consider.

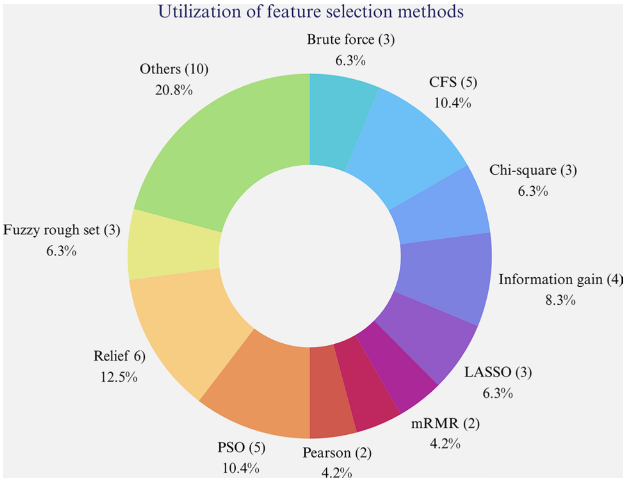


Fig. 8 Utilization of feature selection methods

2.5 Utilization of Hyperparameter Tuning

Hyperparameters are crucial in machine learning algorithms because they directly influence the behavior of training processes and significantly impact model performance. The choice of hyperparameters can greatly affect how well an algorithm performs, making hyperparameter optimization a critical, albeit challenging task [54]. Hyperparameter optimization involves finding the best values for these parameters by minimizing the validation error. The goal is to identify hyperparameters that result in the lowest error on the validation set, with the expectation that these parameters will also perform well on the testing set. However, evaluating the objective function can be expensive, as it requires training the machine learning model with various sets of hyperparameters. Table 4 details the hyperparameter optimization techniques employed in state-of-the-art research.

From Table 4, Haq et al. [38] employed manual tuning in their work for LR, KNN, ANN, SVM, DT, and RF models. Ahmed et al. [54] performed grid search hyperparameter tuning in their work for LR, DT, SVM, and RF classifiers to improve the heart disease prediction accuracy on selected features of the Cleveland dataset obtained through univariate and relief methods. Tama et al. [61] also implemented grid search hyperparameter tuning for the RF algorithm. Muhammad et al. [62] also implemented manual tuning of KNN, SVM, LR, and ANN algorithms. For some application domains, several strategies have been developed and effectively utilized. As a result, developing an effective hyperparameter tuning technique to optimize any given machine learning method can significantly enhance its efficiency. There are primarily two types of hyperparameter tuning approaches manual search and automated search. Manual search is a method of manually testing hyperparameter sets. It is based on the core perception and expertise of professional users who could identify the crucial elements that substantially impact the outcomes and then utilize visualization tools to evaluate the relationship between specific parameters and results. Manual optimization consumes a lot of time instead of focusing on the critical steps such as feature engineering and interpreting the results. Hyperparameter tuning is progressively achieved by automated methods to find optimal hyperparameters in less time by using an intelligent query without requiring manual effort beyond the initial set-up. There are three types of optimization algorithms: grid search, random search, and Bayesian. Grid and random search algorithms need long running times compared to Bayesian because they spend

Table 4 Hyperparameter optimization performed in the state-of-the-art works

Author	Model	Hyperparameters	Optimization technique	Limitations
Tomar and Agarwal (2014) [31]	LSTSVM	Initial C, sigma	GridSearchCV	Only one classifier utilized in this work
Haq et al. (2018) [38]	LR	Regularization strength (C)	Manual tuning	Only one parameter per classifier has been tuned since it takes more time to tune more than one parameter
	KNN	Number of nearest neighbors (K)		
	ANN	Number of hidden neurons		
	SVM	Complexity parameter (C) and gamma (g)		
	DT	Number of trees		
	RF	Number of iterations		
Ahmed et al. (2020) [54]	LR	maxIter, regParam	GridSearchCV	Computational complexity is high. Performed optimization on only four classification models
	DT	impurity, maxDepth, maxBins		
	SVM	regParam, maxIter, Kernel type		
	RF	maxDepth, maxBins, numTrees		
Muhammad et al. (2020) [62]	KNN	No. of nearest neighbors (K)	Manual tuning	Only one parameter per classifier has been tuned since it takes more time to tune more than one parameter
	SVM	Kernel		
	LR	Regularization strength (C)		
	ANN	Number of hidden neurons		
Tama et al. (2020) [61]	RF	ntrees, max depth, min row, nbins, nbins cats, sample rate, histogram type, and distribution	GridSearchCV	Computational complexity is high. Performed optimization on only one classification model
Torthi et al. (2023) [70]	RF	No. of trees, min_samples_leaf, min_samples_split, max_depth, max_features	GridSearchCV	This work is limited small dataset of 270 observations
Trigka et al. (2023) [68]	MLP	learning rate, momentum, training time	Manual tuning	Ten algorithm parameters were tuned. But the manual tuning takes more time and results also not up to the mark
	KNN	K, search algorithm, cross-validate		
	Stacking	Base classifier and meta classifier		
Reddy et al. (2021) [98]	KNN	No. of neighbors, distance metric, distance weight, standardize data	Bayesian	Still the accuracy can be improved
Reddy et al. (2022) [99]	KNN	No. of neighbors, distance weight	Manual tuning	Manual tuning of hyperparameters took more time
	RotF	Base classifier, no. of. iterations, removed percentage		

time analyzing unpromising areas of the search space [100, 101].

3 Dataset Description

Majority of existing works utilized Cleveland dataset which is available in the University of California Irvine (UCI) machine learning repository. This UCI Cleveland heart dataset consists of 303 instances with 76 numeric attributes. However, only 14 attributes (13 predictor variable and 1 class variable) have been considered as significant for the research works to predict the heart disease risk. The description and range of the heart disease dataset attributes is provided in the Table 5.

The ‘age’ attribute has 41 distinct values between 29 and 77 with 1% (4) of unique values, mean of 54.439, and standard deviation of 9.039. The ‘sex’, represents gender, has 2 distinct values 0 = female, 1 = male with mean of 0.68, and standard deviation of 0.467. The ‘cp’, represents chest pain type, has 4 distinct values 1 = typical angina, 2 = atypical angina, 3 = non-angina pain, 4 = asymptomatic with mean of 3.158, and standard deviation of 0.96. The ‘trestbps’, represents resting blood pressure in mm Hg, has 49 distinct values between 94 and 200 with 6% (17) of unique values, mean of 131.69, and standard deviation of 17.6. The ‘chol’, represents serum cholesterol in mg/dl, has 152 distinct values between 126 and 564 with 20% (61) of unique values, mean of 246.693, and standard deviation of 51.777. The ‘fbs’, represents fasting blood

Table 5 Description and range of the heart disease dataset attributes

S. NO	Attribute	Description
1.	age	Age in years between 29 and 77
2.	sex	Gender: 1 represents male, 0 represents female
3.	cp	Type of chest pain 1-typical angina, 2-atypical angina, 3-non-angina pain, 4-asymptomatic
4.	trestbps	Resting blood pressure in mm Hg between 94 and 200
5.	chol	Serum cholesterol in mg/dl between 126 and 564
6.	fbs	Fasting blood sugar > 120 mg/dl: 1 indicates True, 0 indicates False
7.	restecg	Resting electrocardiographic results 0-normal, 1-ST-T wave abnormality, 2-definite left ventricular hypertrophy by Estes' criteria
8.	thalach	Maximum heart rate achieved between 71 and 202
9.	exang	Exercise induce angina: 1 represents Yes, 0 represents No
10.	oldpeak	ST depression induced by exercise relative to rest between 0 and 6.2
11.	slope	The slope of the peak exercise ST segment: 1-upsloping,2-flat, 3-downsloping
12.	ca	Number of major vessels colored by fluoroscopy between 0 and 3
13.	thal	The heart status 3-normal, 6-fixed defect, 7-reversible defect
14.	target	Prediction attribute 0-no risk of heart disease, 1 to 4-risk of heart disease

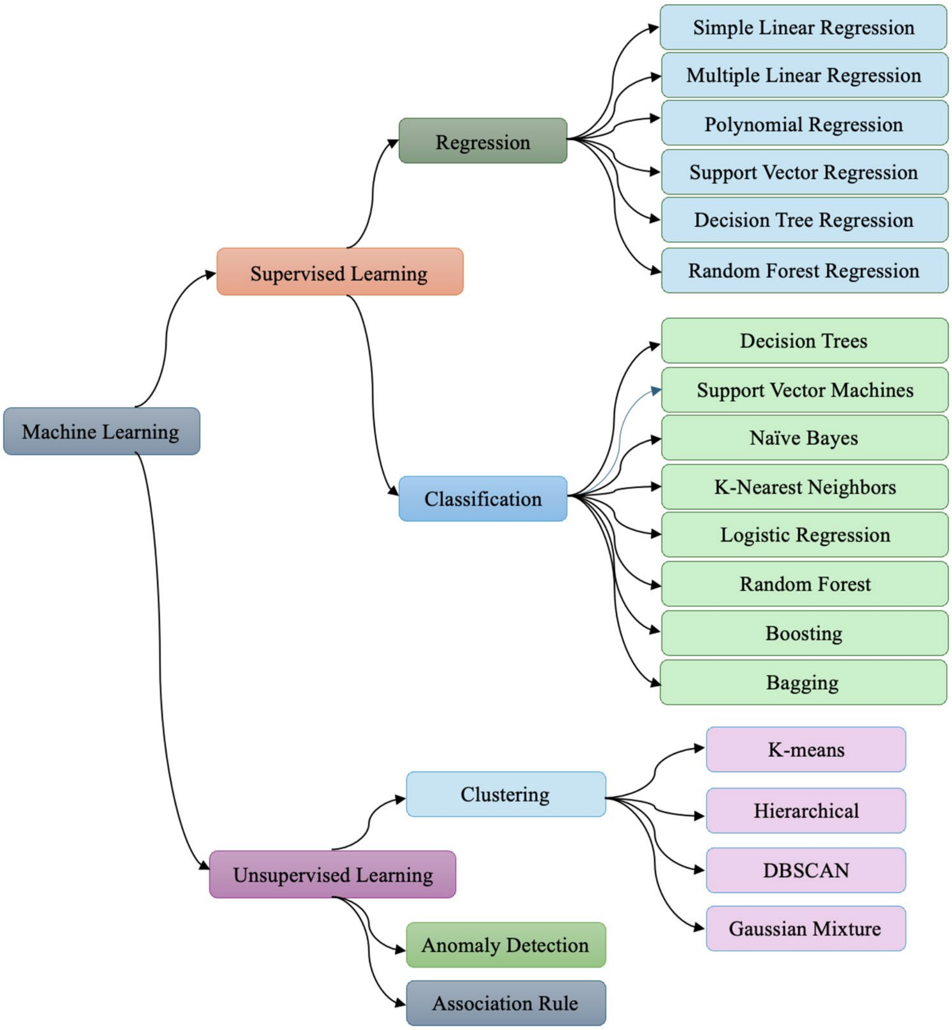
sugar > 120 mg/dl, has 2 distinct values 0 = false, 1 = true with mean of 0.149, and standard deviation of 0.356. The 'restecg', represents resting electrocardiographic results, has 3 distinct values 0 = normal, 1 = having ST-T wave abnormality, 2 = probable or definite left ventricular hypertrophy with mean of 0.99, and standard deviation of 0.995. The 'thalach', represents maximum heart rate, has 91 distinct values between 71 and 202 with 9% (28) of unique values, mean of 149.607, and standard deviation of 22.875. The 'exang', represents exercise induced angina, with 2 distinct values 0 = no, 1 = yes with mean of 0.327, and standard deviation 0.47. The 'oldpeak' attribute, represents ST depression induced by exercise relative to rest, has 40 distinct values between 0 and 6.2 with 3% (10) of unique values, mean of 1.04, and standard deviation of 1.161. The 'slope', represents the slope of the peak exercise ST segment, has 3 distinct values 1 = up-sloping, 2 = flat, 3 = down-sloping with mean of 1.601, and standard deviation of 0.616. The 'ca', represents number of major vessels colored by fluoroscopy, has 4 distinct values between 0 and 3 with mean of 0.672, and standard deviation of 0.937. This attribute as 4 missing values (about 1%). The 'thal', represents the thallium stress result, has 3 distinct values 3-normal, 6-fixed defect, and 7-reversible defect with mean of 4.734, and standard deviation of 1.94. It has 2 missing values (about 1%). The 'target' attribute, represents predicted attribute, has 5 distinct values 0 = no risk of heart disease, 1, 2, 3, 4 = risk of heart disease with the count of 164 and 139, respectively.

4 Machine Learning Algorithms

Machine learning is a data analysis technique that automates the development of analytical models. As a branch of artificial intelligence, it focuses on the concept that computers can learn from data, identify patterns, and make decisions with minimal human intervention. Machine learning is broadly categorized into two types: supervised and unsupervised learning. In supervised learning, the goal is to predict an output variable associated with each input item, or to forecast target values based on labelled data. This approach requires a dataset with both input and corresponding output data to train the model for making predictions on new inputs. Supervised learning includes classification, which assigns discrete classes as target values, and regression/estimation, which deals with continuous target values. Classifiers learn patterns from the features of a dataset and use this knowledge to predict the target class for new data [102]. Figure 9 illustrates the main categories of machine learning algorithms.

In many scenarios, input data lack labels, and in such cases, the challenge is to uncover useful patterns within the data. Unsupervised machine learning is designed to extract meaningful structures or knowledge from data without labelled examples. One common technique in this domain is clustering, which groups similar instances together. Choosing the right algorithm can be challenging due to the variety of both supervised and unsupervised machine learning algorithms available. Identifying the most suitable algorithm often requires experimentation,

Fig. 9 Main categories of machine learning algorithms



as even experienced data scientists cannot ascertain an algorithm's effectiveness without testing it. The selection of algorithms depends on several factors, including the data's characteristics—such as its type, size, and the specific insights sought. The most commonly used machine learning classifiers for heart disease risk prediction, as utilized by researchers, are discussed here.

4.1 Naïve Bayes

Naïve Bayes (NB) is a straightforward probabilistic model grounded in Bayes' theorem. This algorithm estimates the likelihood that a given data point belongs to a specific class by calculating the probability for each possible class. It operates under the assumption that each attribute is independent of the others, a condition that may not always hold true in real-world data. Naïve Bayes is often used in text classification tasks, such as sentiment analysis and spam filtering, due to its effectiveness in handling multiple features and classes. It performs particularly well with small datasets and categorical input variables. In Naïve Bayes, training observations are generated from samples of various statistical distributions, with each response class having its own distribution. The model assigns a probability to each distribution, indicating the likelihood of a new data point falling within that distribution. The normal distribution's parameters are the mean and standard deviation. Naïve Bayes can also handle datasets with substantial amounts of missing data by predicting missing values based on general probability distributions. Bayes' theorem, the mathematical foundation of this algorithm, calculates the probability of an event occurring given prior knowledge of conditions related to the event [87].

Mathematically,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

where A and B are events, $P(A|B)$ is the probability of event A occurring given that B is true, $P(B|A)$ is the probability of event B occurring given that A is true. $P(A)$ and $P(B)$ are the marginal or prior probabilities.

4.2 Decision Tree

Decision Trees (DT) are a crucial algorithm in machine learning for predictive modeling, capable of handling both regression and classification tasks. They operate by recursively splitting the data into smaller, more manageable groups based on the most significant features at each step, ultimately forming a tree-like structure. However, they are prone to overfitting, especially when trained on noisy data or allowed to grow excessively deep. Decision Trees are versatile and can handle both nominal and numerical data. Their sensitivity to slight variations in the training data means that different datasets may result in different tree structures. To mitigate overfitting, pruning techniques can be applied to remove nodes that have minimal predictive value or contribute little to the overall accuracy. In a classification tree, which is used when the target variable has a finite set of distinct values, the decision tree model is represented as a binary tree. Each node corresponds to a single input variable (x) and a specific split point for that variable if it is numeric. The terminal nodes, or leaf nodes, include a dependent variable (y) used for predictions. To make predictions, the model traverses the tree branches until reaching a leaf node, where it outputs the class value associated with that leaf node [103]. Decision Trees are known for their quick learning and rapid prediction capabilities, and they can perform well even with significant amounts of missing data and large datasets.

4.3 Logistic Regression

Logistic Regression (LR) is used for binary classification problem. The algorithm maps the probability of the output to the input features using a logistic function. Because of its ease of use, interpretability, and effectiveness with big datasets, this algorithm is well-liked. It is extensively utilized for activities including predicting illness risk, customer attrition, and credit scoring in a variety of industries, including marketing, finance, healthcare, and more. With the use of strategies like one-versus-rest or multinomial logistic regression, logistic regression may be expanded to deal with multi-class classification issues. Logistic regression is capable of handling both numeric and categorical input features. A linear regression model is a mathematical equation that involves multiplying each predictor by the coefficient and then summing them together [104]. The total is used as the input for the logistic function, which predicts the class to which the observation belongs. For a single observation x with n features the response y is given by

$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n)}} \tag{2}$$

Values that exceed the threshold are classified as positive. Individuals who fall below the specified threshold are categorised as belonging to the category of the negative. Logistic regression can work natively with mixed numerical and categorical variables.

Figure 10 shows target class labelled according to the threshold set which relies upon the output value returned by Sigmoid function. Assuming threshold value of below 0.5 are set to fall under Class 0. If the returned value is above 0.5, then the target class would be labelled as Class 1.

4.4 K-Nearest Neighbors

The K-Nearest Neighbors (KNN) model is a supervised machine learning algorithm commonly used for classification tasks [105]. KNN classifiers are also referred to as instance-based or memory-based learning algorithms. This approach works by storing labeled instances from the training set and recalling them when needed. Later, these stored instances are used to classify new data points. The 'K' in KNN represents the number of nearest neighbors the classifier will retrieve to make a prediction. When faced with a new instance for classification, the KNN algorithm searches its stored training data to find the K examples with the most similar features [106]. The classifier then retrieves the class labels of these K nearest neighbors and aggregates them to predict the label of the new instance. To make predictions, KNN identifies the K training examples closest to the new input in the feature space. KNN is a simple and versatile technique that can be applied to both regression and classification problems. The value of K plays a crucial role in

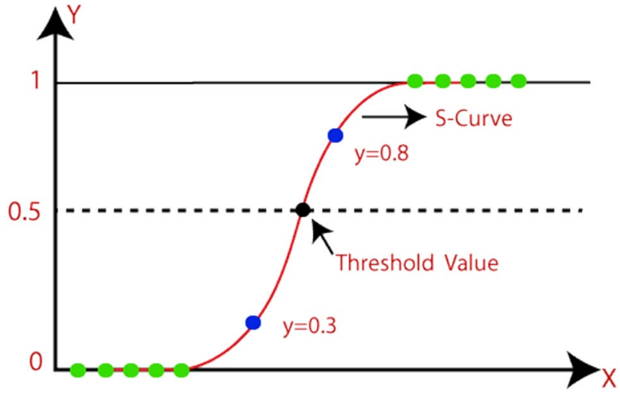


Fig. 10 Sigmoid Function

determining the model's accuracy and generalization. Since KNN is a non-parametric method, it does not assume any specific distribution of the underlying data. However, one drawback of KNN is its computational inefficiency, particularly with large datasets, as it requires storing all the training data and calculating distances for each prediction.

Figure 11 illustrates how the target class is assigned to a new data point based on the majority class of its nearest neighbors. For example, the red color datapoint will be labelled as Category A since there are three nearest neighbors positioned closer to this new data point.

4.5 Support Vector Machines

The Support Vector Machine (SVM) is widely recognized as one of the most commonly used machine learning algorithms for classification tasks. Its primary function is to find the hyperplane that best separates different classes within the feature space. SVM can handle both linear and non-linear data by using various kernel functions, which transform the input data into a higher-dimensional space. Because SVM relies on a subset of the training data points (called support vectors) in its decision-making process, it is efficient in high-dimensional spaces and requires less memory. SVMs are applied across various fields, such as bioinformatics, text classification, and image recognition. However, one limitation of SVM is its sensitivity to the selection of kernel and regularization parameters, which often require careful tuning for optimal performance. The SVM algorithm generates a hyperplane that separates data points into two distinct classes in an n-dimensional feature space. This separation is achieved by employing the maximal-margin technique [107]. A hyperplane can be described as the set of data points x that satisfy the following condition:

$$\mathbf{w}.\mathbf{x} - \mathrm{b} = 0 \tag{3}$$

where b is the offset value from the origin, w is a normal vector, perpendicular to the surface of the hyperplane. The parameter $\frac{b}{\|w\|}$ determines the offset of the hyperplane from the origin along the normal vector **w**. Data should be standardized since the distance measurements are sensitive to scale. Figure 12 shows the optimized hyperplane having large marginal distance from data points (binary classes) to the hyperplane created.

From Fig. 12, the larger the marginal distance, the more confidence it is to classify whether the person is suffering from diabetes or not.

4.6 Random Forest

Random Forest (RF) is a widely used and versatile supervised machine learning algorithm, effective for both regression and classification tasks. It consists of multiple decision trees trained through a bagging technique to improve overall predictive performance [108]. Random Forest builds several decision tree classifiers and combines their predictions, reducing variance while only slightly increasing bias, thus controlling overfitting. This results in a more accurate and reliable prediction model. The RF algorithm introduces randomness by selecting random thresholds for each feature as the trees are built. When splitting a node, RF chooses the best feature from a random subset of available features, enhancing the model's performance. Due to its ability to handle large, high-dimensional datasets without overfitting, Random Forest is widely adopted across many machine learning tasks. Its effectiveness comes from applying the method to both regression and classification problems, while incorporating randomness in feature selection and data sampling, which reduces variance and improves generalization.

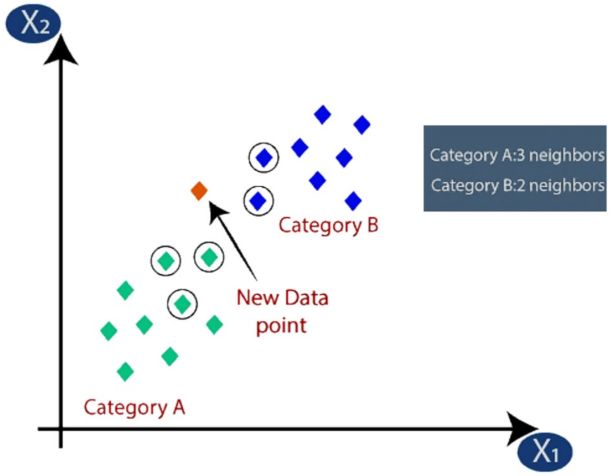


Fig. 11 Target class assigned to the closest neighbor's class

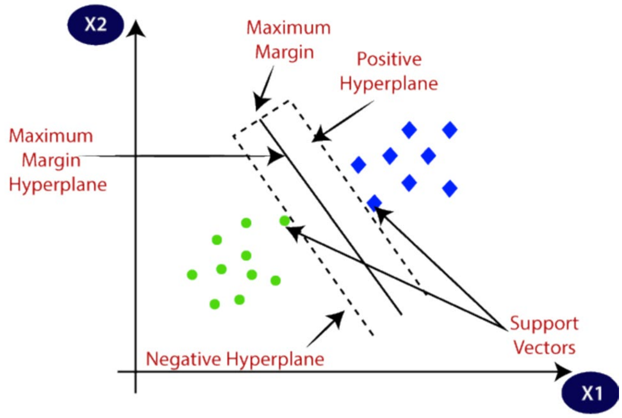


Fig. 12 Marginal distance from data points

Furthermore, Random Forest can handle missing values effectively, maintaining accuracy even when some data is incomplete [109]. Figure 13 illustrates how Random Forest is composed of multiple decision trees using bagging and random feature selection techniques.

For instance, individual models of decision tree have predicted "1" for six times and "0" for three times. By using the majority voting mechanism, the final model is labelled as class "1".

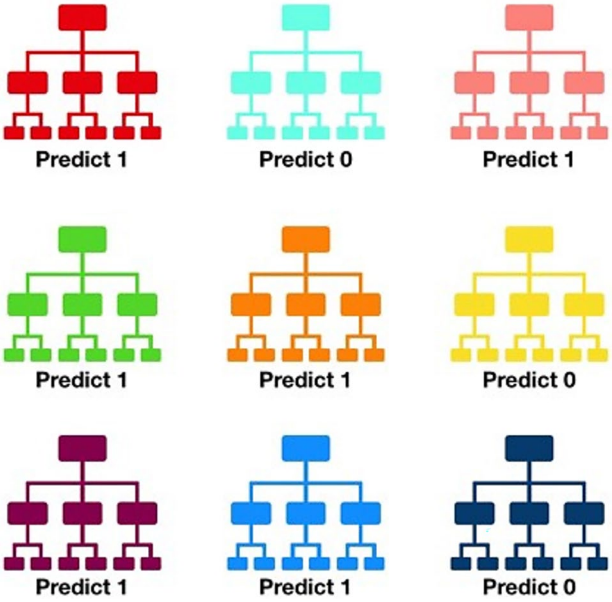


Fig. 13 Random Forest with many decision trees using bagging and random feature selection technique

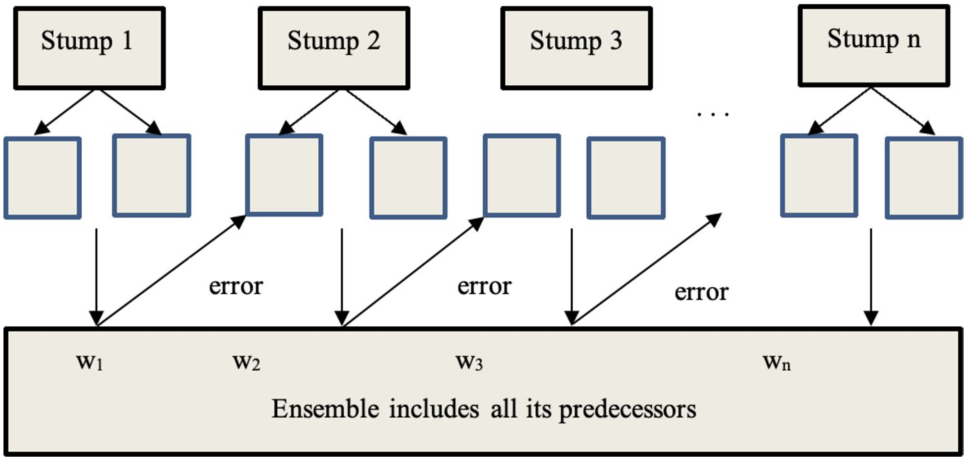
4.7 Boosting

Boosting is a potent machine learning strategy that combines several weak learners to increase a model's predicted accuracy. In order to repair the mistakes made by earlier models, iteratively training new models involves assigning greater weight to previously misclassified cases. XGBoost, Gradient Boosting Machine (GBM), and AdaBoost are a few well-liked boosting techniques. Although it may be applied to other kinds of models as well, decision trees are frequently utilized in conjunction with boosting. For classification and regression applications, boosting can produce strong and highly accurate predictive models when used properly. Adaptive Boosting (AB) is frequently referred to as the best classifier with the decision trees as weak learners [110]. It is analogous to RF in that it uses numerous decision trees to make predictions. Figure 14 illustrates AB which is distinguished by three important characteristics:

- Instead of trees, AB constructs a forest of decision stumps. A tree with only one node and two leaves is known as a decision stump.
- In the final prediction, the decision stumps that are formed are not equally weighted.
- Because each stump tries to minimize the errors generated by the prior stump(s), the sequence in which they are created is crucial.

Pruning is the process of deleting poor-performing weak classifiers to reduce the boost classifier's storage and execution time. Weight or margin-trimming are the most basic strategies, and they can be predominantly effective when used in combination with entirely corrective training: when a weak classifier's coefficient, or contribution to the overall test error, falls below a specified threshold, that classifier will be deleted.

Fig. 14 Adaptive Boosting



4.8 Bagging

Bagging is a commonly used ensemble technique in machine learning that helps reduce overfitting and enhances the model's predictive performance. In bagging, the same learning algorithm is trained on multiple instances, each using different subsets of the training data, and the predictions from these instances are combined to produce the final output. Typically, bootstrap sampling is used to generate the subsets, where samples are drawn from the original dataset with replacement [111]. Bagging is employed by popular algorithms such as Bagged Decision Trees and Random Forest. Since each model in a bagging ensemble is trained independently, the process can be parallelized, which may significantly reduce training time.

Bootstrap sampling is used by bagging to generate m new training sets, denoted as D_i, from the initial training dataset D. Each new training set is of size n' and is created by randomly selecting samples from D with replacement. Each D_i may contain repeated samples obtained through sampling with replacement. For large values of n, if n' is equal to n, it is expected that the set D_i will contain approximately 63.2% $(1 - 1/e)$ of the distinct samples in D, while the other samples will be duplicates. Because sampling with replacement does not rely on previously chosen samples, each bootstrapping is distinct from its predecessors. Next, by averaging the result (for regression) or polling (for classification), m models are trained and merged using these m bootstrap samples. The out-of-bag dataset contains the samples that were left out of the bootstrap sample. The difference between the original training dataset and bootstrap datasets can be used to compute it. Because the bootstrap sample and out-of-bag datasets are used to assess the correctness of a random forest method, they are critical.

Afterwards, the ensemble creates decision trees using bootstrapping samples. It evaluates each feature and determines the number of samples in which the presence or absence of the feature leads to an outcome that is either positive or negative. A classifier C_i is developed by learning algorithm from each bootstrapped sample D_i to accomplish the classification of setD_i. On the original training data D, the final classifier $C*$ is produced by employing the earlier created set of classifiersC_i.The final classification is the one predicted by the sub-classifiers C_i most frequently. The diagram of Bagging is shown in Fig. 15.

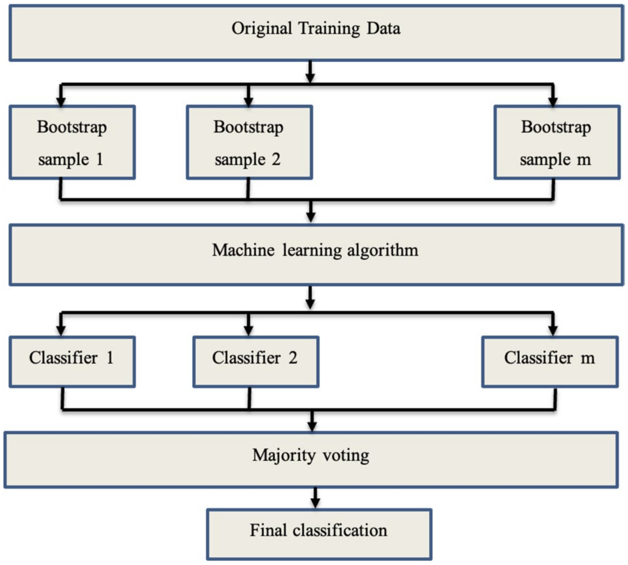


Fig. 15 Bagging

5 Deep Learning Algorithms

Deep learning is a branch of machine learning that use multilayered neural networks, known as deep neural networks, to replicate the intricate decision-making capabilities of the human brain. Neural networks, or artificial neural networks, endeavour to replicate the human brain by utilising data inputs, weights, and biases, all functioning as silicon brain cells. These components collaboratively function to precisely identify, categorise, and delineate items within the data. There are various types of neural networks, including Artificial Neural Networks (ANN), convolutional neural networks (CNNs), recurrent neural networks (RNNs), and specialized architectures like autoencoders and generative adversarial networks (GANs).

5.1 Artificial Neural Networks

Artificial neural networks (ANNs) are made up of interconnected nodes, referred to as artificial neurons, which are inspired by the structure and functioning of the human brain. These neurons process and analyze data through connections, where each neuron receives one or more weighted inputs, sums them, and applies a non-linear function (activation function) to produce an output. The output of one neuron is passed as input to subsequent neurons via these connections, with each connection assigned a weight that determines its importance. A single neuron can have multiple input and output connections, and the input to a neuron is calculated by taking the weighted sum of the outputs from previous neurons and their connections [107]. The network adjusts its weights during training based on a learning rule, using the error (the difference between the network's output and the desired output) to improve its predictions. ANNs are widely used in tasks like natural language processing, image and speech recognition, and decision-making. They are foundational to modern supervised learning algorithms, which train the network to make predictions or classifications from labelled input data.

5.2 Convolutional Neural Networks

Convolutional neural networks (CNNs) are predominantly utilized in computer vision and image classification tasks. They can identify features and patterns in images and videos, facilitating activities such as recognizing objects, recognition of images, recognition of patterns, and recognition of faces. CNNs are a distinct category of neural networks, with node layers that include an input layer, one or more hidden layers, and an output layer. Every node is interconnected with another, possessing a corresponding weight and threshold. When the output of a specific node exceeds the designated threshold value, that node becomes activated and transmits data to the subsequent layer of the network. If not, no data is transferred to the network's subsequent layer. A CNN has at least three primary types of layers: the convolutional layer, the pooling layer, and the fully connected (FC) layer. For intricate applications, a CNN may comprise thousands of layers, with each layer augmenting the preceding ones. Through convolution detailed patterns can be unveiled. As each layer is added, the CNN enhances its complexity, discerning larger segments of the image. The initial layers concentrate on fundamental elements, including colors and borders. As the picture data traverses the layers of the CNN, it begins to discern larger components or shapes of the item until it ultimately identifies the target object [80, 83].

5.3 Recurrent Neural Networks

Recurrent neural networks (RNNs) are a category of neural networks capable of processing sequential input, including time series and natural language. RNNs utilise their memory to incorporate information from previous inputs, hence affecting the current input and result. Traditional deep neural networks presume that inputs and outputs are independent, whereas the output of recurrent neural networks (RNNs) is contingent upon preceding elements in the sequence. RNNs utilise parameter sharing across all layers of the network and maintain identical weight parameters across each layer, with these weights modified by backpropagation and gradient descent to provide reinforcement learning. RNNs employ a backpropagation through time (BPTT) technique to compute gradients, which differs marginally from conventional backpropagation as it is tailored for sequential data. The principles of BPTT mirror those of conventional backpropagation, wherein the model self-trains by computing errors from the output layer to the input layer.

6 Conclusion

Cardiovascular disease, commonly referred to as heart disease, is a major concern in the medical field. According to recent estimates from the World Health Organization (WHO), over 20.5 million people are dying from cardiovascular disease, accounting for 31.5% of all global deaths. It is also projected that by 2030, the annual death toll will rise to 24.2 million. This paper critically reviews and summarizes the research works from 2014 to 2024 on predicting heart disease risk using machine learning and deep learning algorithms. It also highlights the feature selection techniques applied in previous studies and the key features identified to enhance heart disease risk prediction. Additionally, this work explores the hyperparameter tuning methods used in state-of-the-art studies to boost the performance of machine learning models. The findings indicate that SVM and RF techniques are the most commonly used, offering better accuracy in heart disease prediction. Models based on NB, KNN, and ANN also performed well in most cases. However, the accuracy of these models varies depending on factors such as the tool/software used, dataset size, feature set, number of instances, data preprocessing, feature selection methods, and the choice of classifier. There remains significant research potential in addressing issues like missing data, outliers, and overfitting. Current works often lack hyperparameter optimization, which is crucial for enhancing machine learning performance but is complex and time-consuming. It is recommended to explore new feature selection methods for heart disease datasets and perform hyperparameter tuning to improve prediction accuracy heart disease risk.

7 Future Research Directions

7.1 Exploration of Emerging Algorithms

- Hybrid Models: Explore hybrid models that combine the strengths of machine learning (ML) and deep learning (DL) algorithms. For example, using ensemble learning techniques that integrate decision trees (e.g., Random Forest) with deep neural networks to improve prediction accuracy.
- Explainable AI (XAI): Heart disease prediction requires high transparency for medical professionals. Future research could explore interpretability tools (like SHAP, LIME, or Grad-CAM) to explain model decisions clearly.
- Transformer-based Models: With the success of transformers in NLP and computer vision, investigating their use for time-series or sequential health data could be a novel direction.

7.2 Incorporation of Multi-modal Data

- Integration of Diverse Data Sources: Use multi-modal datasets that include medical imaging (e.g., ECG, echocardiograms), clinical notes (NLP techniques), genetic information, and traditional tabular data to improve model performance.
- Wearable Sensor Data: Heart rate and ECG data from wearable devices like smartwatches could provide real-time data streams for early detection.

7.3 Personalized and Precision Medicine Approaches

- Personalized Risk Prediction Models: Incorporate personalized models based on individual genetic markers, lifestyle factors, and medical history to tailor heart disease risk prediction to specific populations.
- Transfer Learning for Personalized Care**:** Transfer learning from one patient cohort to another can help build adaptable models for patients from different demographics, improving model generalizability.

7.4 Data Quality and Bias Mitigation

- Addressing Class Imbalance and Bias: Heart disease datasets often suffer from class imbalance (more healthy patients than those with heart disease) and bias in race, age, and gender. Future works should focus on using techniques like SMOTE, ADASYN, or cost-sensitive learning to handle imbalances and bias.
- Ethical Considerations and Fair AI: Implement fairness-aware algorithms that ensure unbiased predictions across various demographics. Develop frameworks to assess the fairness of predictions in heart disease models.

7.5 Time-Series and Longitudinal Studies

- Longitudinal Health Data Analysis: Using temporal modeling techniques like Long Short-Term Memory (LSTM) and Temporal Convolutional Networks (TCNs) to analyze patient health records over time and predict long-term heart disease risks.
- Survival Analysis: Apply deep learning-based survival analysis to predict not just the risk of heart disease but also the time until a cardiovascular event occurs.

7.6 Real-World Implementation and Clinical Validation

- Clinical Trials and Real-World Validation: Incorporate ML/DL models into clinical trials for real-world validation. Address the gap between lab-developed models and their clinical adoption.
- Mobile and Cloud-based Deployments: Work on deploying these models in real-world applications, such as mobile applications or cloud services that allow for continuous heart health monitoring and early intervention.

7.7 Federated Learning for Health Data Privacy

- Federated Learning (FL): Investigate the use of federated learning to train ML models across multiple healthcare institutions while preserving patient privacy. FL can help overcome data-sharing limitations in heart disease research.

7.8 Integration of Social Determinants of Health (SDOH)

- Incorporating SDOH: Future research could include socioeconomic, environmental, and behavioral factors (e.g., income, education, access to healthcare) to develop a holistic heart disease prediction model.
- Community Health Predictions: Utilize publicly available datasets and social factors to understand how community health influences individual heart disease risk.

7.9 Optimization of Feature Engineering

- Automatic Feature Selection and Extraction: Use feature selection techniques (like recursive feature elimination, PCA, or autoencoders) to identify key variables that most influence heart disease risk.
- Synthetic Data Generation: Explore generative adversarial networks (GANs) for generating synthetic data, especially for rare heart disease cases, to enhance model training.

7.10 Future Dataset Availability and Standardization

- Creation of Standardized Heart Disease Datasets: Collaborate on creating standardized and open-access datasets that can facilitate better benchmarking across different algorithms and institutions.
- Data Augmentation Strategies: Explore advanced data augmentation techniques to expand existing datasets and improve model robustness, especially when working with limited medical data.

These directions offer pathways for improving heart disease risk prediction models and pushing the frontier of research in this domain. They balance technical advancements, practical considerations, and the ethical implications of ML and DL applications in healthcare.

Author Contributions Vishnu Vardhana Reddy Karna: writing – review & editing, writing—original draft, conceptualization, investigation, prepared tables. Viswavardhan Reddy Karna: writing—original draft, writing—review & editing, prepared figures. Varaprasad Janamala: writing – review & editing, prepared tables. V N Koteswara Rao Devana: writing – review & editing, formal analysis. V Ravi Sankar Ch: Writing – review & editing, formal analysis, prepared figures. Aravinda Babu Tummala: formal analysis, prepared tables.

Funding This work received no external funding.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflicts of interest The authors declare no competing interests.

References

1. Aparicio et al. H J (2021) Heart disease and stroke statistics-2021 update: a report from the american heart association. https://doi.org/10.1161/CIR.0000000000000950.
2. WHO [Online]. https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_2
3. Dhar S, Roy K, Dey T, Datta P, Biswas A (2018) A hybrid machine learning approach for prediction of heart diseases, In: 2018 4th International conference on computing communication and automation, ICCCA 2018, pp. 1–6, https://doi.org/10.1109/CCAA.2018.8777531.
4. Thomas J, Heart disease: facts, statistics, and you [Online]. https://www.healthline.com/health/heart-disease/statistics
5. Heidenreich P et al (2011) Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. Circulation 123(8):933–944. https://doi.org/10.1161/CIR.0b013e31820a55f5
6. Obasi T, Omair Shafiq M (2019) Towards comparing and using machine learning techniques for detecting and predicting heart attack and diseases, Proceedings-2019 IEEE international conference on big data, Big data 2019, pp. 2393–2402, https://doi.org/10.1109/BigData47090.2019.9005488.
7. Das Adhikari N C, Alka A, Garg R (2017) HPPS : heart problem prediction system using machine learning, December, pp. 23–37, https://doi.org/10.5121/csit.2017.71803.
8. Miranda E, Irwansyah E, Amelga AY, Maribondang MM, Salim M (2016) Detection of cardiovascular disease risk's level for adults using naive bayes classifier. Healthc Inform Res 22(3):196–205. https://doi.org/10.4258/hir.2016.22.3.196
9. Georga E I et al. (2019) Artificial intelligence and data mining methods for cardiovascular risk prediction, Cardiovascular computing—methodologies and clinical applications, Series in BioEngineering, Springer, Singapore, pp. 279–301. https://doi.org/10.1007/978-981-10-5092-3_14.
10. Shahid AH, Singh MP, Roy B, Aadarsh A, (2020) Coronary artery disease diagnosis using feature selection based hybrid extreme learning machine, Proceedings—3rd international conference on information and computer technologies, ICICT 2020, pp. 341–346, https://doi.org/10.1109/ICICT50521.2020.00060.
11. Radha BK, Harsoor B, Das R (2018) A novel framework for prediction and detection of myocardial ischemia with single lead ECG and PAT, Proceedings of the 2018 3rd IEEE international conference on research in intelligent and computing in engineering, RICE 2018, pp. 1–6, https://doi.org/10.1109/RICE.2018.8509036.
12. Kandala RNVPS et al (2019) Towards real-time heartbeat classification: evaluation of nonlinear morphological features and voting method. Sensors 19(23):5079. https://doi.org/10.3390/s19235079
13. Verma P (2020) Ensemble models for classification of coronary artery disease using decision trees. Int J Recent Technol Eng 8(6):940–944. https://doi.org/10.35940/ijrte.f7250.038620
14. Ghiasi MM, Zendehboudi S, Mohsenipour AA (2020) Decision tree-based diagnosis of coronary artery disease: CART model. Comput Methods Progr Biomed 192:105400. https://doi.org/10.1016/j.cmpb.2020.105400
15. Arabasadi Z, Alizadehsani R, Roshanzamir M, Moosaei H, Yarifard AA (2017) Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. Comput Methods Progr Biomed 141:19–26. https://doi.org/10.1016/j.cmpb.2017.01.004
16. Ali L et al (2019) A feature-driven decision support system for heart failure prediction based on $\chi 2$ statistical model and Gaussian Naive Bayes. Comput Math Methods Med. https://doi.org/10.1155/2019/6314328
17. Ramalingam VV, Dandapath A, Raja MK (2018) Heart disease prediction using machine learning techniques: a survey. Int J Eng Technol (UAE). https://doi.org/10.14419/ijet.v7i2.8.10557
18. Allen LA et al (2012) Decision making in advanced heart failure: a scientific statement from the american heart association. Circulation 125(15):1928–1952. https://doi.org/10.1161/CIR.0b013e31824f2173
19. Nguyen K et al. (2021) Heart disease classification using novel heterogeneous ensemble, In: 2021 IEEE EMBS international conference on biomedical and health informatics (BHI), Athens, Greece, pp. 1–4, https://doi.org/10.1109/BHI50953.2021.9508516.
20. Khan Y, Qamar U, Yousaf N, Khan A (2019) Machine learning techniques for heart disease datasets: a survey, ACM international conference proceeding series, vol. Part F1481, pp. 27–35, https://doi.org/10.1145/3318299.3318343
21. Goel S, Deep A, Srivastava S, Tripathi A (2019) Comparative analysis of various techniques for heart disease prediction, 2019 4th International conference on information systems and computer networks, ISCON 2019, pp. 88–94, https://doi.org/10.1109/ISCON47742.2019.9036290.
22. Yang H, Garibaldi JM (2015) A hybrid model for automatic identification of risk factors for heart disease. J Biomed Inform 58:S171–S182. https://doi.org/10.1016/j.jbi.2015.09.006
23. Swain D, Pani SK, Swain D, (2018) A metaphoric investigation on prediction of heart disease using machine learning, In: 2018 International conference on advanced computation and telecommunication, ICACAT 2018, pp. 1–6, https://doi.org/10.1109/ICACAT.2018.8933603.
24. Ahamed J, Mir RN, Chishti MA (2022) Industry 4.0 oriented predictive analytics of cardiovascular diseases using machine learning, hyperparameter tuning and ensemble techniques. Ind Robot 49(3):544–554. https://doi.org/10.1108/IR-10-2021-0240
25. Jindal H, Agrawal S, Khera R, Jain R, Nagrath P (2021) Heart disease prediction using machine learning algorithms. IOP Conf Ser Mater Sci Eng. https://doi.org/10.1088/1757-899X/1022/1/012072
26. Sowmiya C, Sumitra P (2018) Analytical study of heart disease diagnosis using classification techniques, Proceedings of the 2017 IEEE international conference on intelligent techniques in

control, Optimization and signal processing, INCOS 2017, vol. 2018-Feb, pp. 1–5, https://doi.org/10.1109/ITCOSP.2017.8303115.
27. Tomov NS, Tomov S, (2018) On deep neural networks for detecting heart disease. https://doi.org/10.48550/arXiv.1808.07168
28. Heart failure. [Online]. https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure
29. Bharti R, Khamparia A, Shabaz M, Dhiman G, Pande S, Singh P (2021) Prediction of heart disease using a combination of machine learning and deep learning. Comput Intell Neurosci. https://doi.org/10.1155/2021/8387680
30. Bashir S, Qamar U, Khan FH, Javed MY (2014) MV5: a clinical decision support framework for heart disease prediction using majority vote based classifier ensemble. Arab J Sci Eng 39(11):7771–7783. https://doi.org/10.1007/s13369-014-1315-0
31. Tomar D, Agarwal S (2014) Feature selection based least square twin support vector machine for diagnosis of heart disease. Int J Bio-Sci Bio-Technol 6(2):69–82. https://doi.org/10.14257/ijbsbt.2014.6.2.07
32. El-Bialy R, Salamay MA, Karam OH, Khalifa ME (2015) Feature analysis of coronary artery heart disease data sets. Procedia Comput Sci 65:459–468. https://doi.org/10.1016/j.procs.2015.09.132
33. Long NC, Meesad P, Unger H (2015) A highly accurate firefly based algorithm for heart disease prediction. Expert Syst Appl 42(21):8221–8231. https://doi.org/10.1016/j.eswa.2015.06.024
34. Verma L, Srivastava S, Negi PC (2016) A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data. J Med Syst. https://doi.org/10.1007/s10916-016-0536-z
35. Masetic Z, Subasi A (2016) Congestive heart failure detection using random forest classifier. Comput Methods Programs Biomed 130:54–64. https://doi.org/10.1016/j.cmpb.2016.03.020
36. Pouriyeh S, Vahid S, Sannino G, De Pietro G, Arabnia H, Gutierrez J (2017) A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease. Proc IEEE Symp Comput Commun. https://doi.org/10.1109/ISCC.2017.8024530
37. Liu X et al (2017) A hybrid classification system for heart disease diagnosis based on the RFRS method. Comput Math Methods Med. https://doi.org/10.1155/2017/8272091
38. Haq AU, Li JP, Memon MH, Nazir S, Sun R, Garciá-Magarinõ I (2018) A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Inform Syst. https://doi.org/10.1155/2018/3860146
39. Kale AP, Sonavane S (2018) PF-FELM: a robust PCA Feature selection for fuzzy extreme learning machine. IEEE J Sel Top Sign Proces 12(6):1303–1312. https://doi.org/10.1109/JSTSP.2018.2873988
40. Saqlain SM et al (2019) Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines. Knowl Inf Syst 58(1):139–167. https://doi.org/10.1007/s10115-018-1185-y
41. Ma M, Deng T, Wang N, Chen Y (2019) Semi-supervised rough fuzzy Laplacian Eigenmaps for dimensionality reduction. Int J Mach Learn Cybern 10(2):397–411. https://doi.org/10.1007/s13042-018-0784-y
42. Takci H (2018) Improvement of heart attack prediction by the feature selection methods. Turk J Electr Eng Comput Sci 26(1):1–10. https://doi.org/10.3906/elk-1611-235
43. Dulhare UN (2018) Prediction system for heart disease using Naive Bayes and particle swarm optimization. Biomed Res (India) 29(12):2646–2649. https://doi.org/10.4066/biomedicalresearch.29-18-620
44. Javeed A, Zhou S, Yongjian L, Qasim I, Noor A, Nour R (2019) An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection. IEEE Access 7:180235–180243. https://doi.org/10.1109/ACCESS.2019.2952107
45. Reddy GT, Reddy MPK, Lakshmanna K, Rajput DS, Kaluri R, Srivastava G (2020) Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. Evol Intell 13(2):185–196. https://doi.org/10.1007/s12065-019-00327-1
46. Ed-daoudy A, Maalmi K, (2019) Performance evaluation of machine learning based big data processing framework for prediction of heart disease, In: 2019 International conference on intelligent systems and advanced computing sciences (ISACS), Taza, Morocco, pp. 1–5. https://doi.org/10.1109/ISACS48493.2019.9068901.
47. Mohan S, Thirumalai C, Srivastava G (2019) Effective heart disease prediction using hybrid machine learning techniques. IEEE Access 7:81542–81554. https://doi.org/10.1109/ACCESS.2019.2923707
48. Alam MZ, Rahman MS, Rahman MS (2019) A Random Forest based predictor for medical data classification using feature ranking. Inform Med Unlocked. https://doi.org/10.1016/j.imu.2019.100180
49. Raza K, (2019) Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule, In: Advances in ubiquitous sensing applications for healthcare, U-healthcare monitoring systems, Academic Press, pp. 179–196, https://doi.org/10.1016/b978-0-12-815370-3.00008-6.
50. Louridi N, Amar M, El Ouahidi B, (2019) Identification of cardiovascular diseases using machine learning, In: 7th Mediterranean congress of telecommunications 2019, CMT 2019, pp. 1–6, https://doi.org/10.1109/CMT.2019.8931411
51. Latha CBC, Jeeva SC (2019) Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Inform Med Unlocked 16:100203. https://doi.org/10.1016/j.imu.2019.100203
52. Amin MS, Chiam YK, Varathan KD (2018) Identification of significant features and data mining techniques in predicting heart disease. Telemat Inform 36:82–93. https://doi.org/10.1016/j.tele.2018.11.007
53. Mienye ID, Sun Y, Wang Z (2020) An improved ensemble learning approach for the prediction of heart disease risk. Inform Med Unlocked 20:100402. https://doi.org/10.1016/j.imu.2020.100402
54. Ahmed H, Younis EMG, Hendawi A, Ali AA (2020) Heart disease identification from patients' social posts, machine learning solution on Spark. Futur Gener Comput Syst 111:714–722. https://doi.org/10.1016/j.future.2019.09.056
55. Budholiya K, Shrivastava SK, Sharma V (2020) An optimized XGBoost based diagnostic system for effective prediction of heart disease. J King Saud Univ Comput Inform Sci. https://doi.org/10.1016/j.jksuci.2020.10.013
56. Gazeloğlu C (2020) Prediction of heart disease by classifying with feature selection and machine learning methods. Prog Nutr 22(2):660–670. https://doi.org/10.23751/pn.v22i2.9830
57. Gupta A, Kumar R, Singh Arora H, Raman B (2020) MIFH: a machine intelligence framework for heart disease diagnosis. IEEE Access 8:14659–14674. https://doi.org/10.1109/ACCESS.2019.2962755
58. Ramesh G, Madhavi K, Reddy PDK, Somasekar J, Tan J (2021) Improving the accuracy of heart attack risk prediction based on information gain feature selection technique. Mater Today Proc. https://doi.org/10.1016/j.matpr.2020.12.079
59. Pasha SJ, Mohamed ES (2020) Novel feature reduction (NFR) model with machine learning and data mining algorithms for

effective disease risk prediction. IEEE Access 8:184087–184108. https://doi.org/10.1109/ACCESS.2020.3028714

60. Mohamed AAA, Hassan SA, Hemeida AM, Alkhalaf S, Mahmoud MMM, Baha Eldin AM (2020) Parasitism—predation algorithm (PPA): a novel approach for feature selection. Ain Shams Eng J 11(2):293–308. https://doi.org/10.1016/j.asej.2019.10.004
61. Tama BA, Im S, Lee S (2020) Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble. Biomed Res Int. https://doi.org/10.1155/2020/9816142
62. Muhammad Y, Tahir M, Hayat M, Chong KT (2020) Early and accurate detection and diagnosis of heart disease using intelligent computational model. Sci Rep 10(1):1–17. https://doi.org/10.1038/s41598-020-76635-9
63. Aslam M and Jaisharma K, (2021) Hierarchical random forest formation with nonlinear regression model for cardiovascular diseases prediction, In: 2021 International conference on computer communication and informatics (ICCCI), Coimbatore, India, pp. 1–5, https://doi.org/10.1109/ICCCI50826.2021.9402571.
64. Rani P, Kumar R, Ahmed NMOS, Jain A (2021) A decision support system for heart disease prediction based upon machine learning. J Reliab Intell Environ 7(3):263–275. https://doi.org/10.1007/s40860-021-00133-6
65. Katarya R, Meena SK (2021) Machine learning techniques for heart disease prediction: a comparative study and analysis. Health Technol (Berl) 11(1):87–97. https://doi.org/10.1007/s12553-020-00505-7
66. Ansarullah SI, Saif SM, Kumar P, Kirmani MM (2022) Significance of visible non-invasive risk attributes for the initial prediction of heart disease using different machine learning techniques. Comput Intell Neurosci. https://doi.org/10.1155/2022/9580896
67. Kanagarathinam K, Sankaran D, Manikandan R (2022) Machine learning-based risk prediction model for cardiovascular disease using a hybrid dataset. Data Knowl Eng. https://doi.org/10.1016/j.datak.2022.102042
68. Trigka M, Dritsas E (2023) Long-term coronary artery disease risk prediction with machine learning models. Sensors. https://doi.org/10.3390/s23031193
69. Chandrasekhar N, Peddakrishna S (2023) Enhancing heart disease prediction accuracy through machine learning techniques and optimization. Processes. https://doi.org/10.3390/pr11041210
70. Torthi R, Marapatla ADK, Mande S, Gadiraju HKV, Kanumuri C (2024) Heart disease prediction using random forest based hybrid optimization algorithms. Int J Intell Eng Syst 17(2):134–144. https://doi.org/10.22266/ijies2024.0430.12
71. Elsedimy EI, AboHashish SMM, Algarni F (2024) New cardiovascular disease prediction approach using support vector machine and quantum-behaved particle swarm optimization. Multimed Tools Appl 83(8):23901–23928. https://doi.org/10.1007/s11042-023-16194-z
72. Nissa N, Jamwal S, Neshat M (2024) A technical comparative heart disease prediction framework using boosting ensemble techniques. Computation. https://doi.org/10.3390/computation12010015
73. Daza A et al (2024) Stacking ensemble based hyperparameters to diagnosing of heart disease: future works. Results Eng. https://doi.org/10.1016/j.rineng.2024.101894
74. Narayanan J (2024) Implementation of efficient machine learning techniques for prediction of cardiac disease using SMOTE. Procedia Comput Sci. https://doi.org/10.1016/j.procs.2024.03.245
75. Samuel OW, Asogbon GM, Sangaiah AK, Fang P, Li G (2017) An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. Expert Syst Appl 68:163–172. https://doi.org/10.1016/j.eswa.2016.10.020
76. Ali L, Rahman A, Khan A, Zhou M, Javeed A, Khan JA (2019) An automated diagnostic system for heart disease prediction based on $\chi 2$ statistical model and optimally configured deep neural network. IEEE Access 7:34938–34945. https://doi.org/10.1109/ACCESS.2019.2904800
77. Javeed A, Rizvi SS, Zhou S, Riaz R, Khan SU, Kwon SJ (2020) Heart risk failure prediction using a novel feature selection method for feature refinement and neural network for classification. Mobile Inform Syst. https://doi.org/10.1155/2020/8843115
78. Das H, Naik B, Behera HS, Jaiswal S, Mahato P, Rout M (2020) Biomedical data analysis using neuro-fuzzy model with post-feature reduction. J King Saud Univ Comput Inform Sci. https://doi.org/10.1016/j.jksuci.2020.01.007
79. Tougui I, Jilbab A, El Mhamdi J (2020) Heart disease classification using data mining tools and machine learning techniques. Health Technol (Berl) 10(5):1137–1144. https://doi.org/10.1007/s12553-020-00438-1
80. Mehmood A et al (2021) Prediction of heart disease using deep convolutional neural networks. Arab J Sci Eng 46(4):3409–3422. https://doi.org/10.1007/s13369-020-05105-1
81. Nancy AA, Ravindran D, Raj Vincent PMD, Srinivasan K, Gutierrez Reina D (2022) IoT-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning. Electronics (Switzerland). https://doi.org/10.3390/electronics11152292
82. Vayadande K et al (2022) Heart disease prediction using machine learning and deep learning algorithms, In: 2022 International conference on computational intelligence and sustainable engineering solutions (CISES), IEEE, pp. 393–401. https://doi.org/10.1109/CISES54857.2022.9844406.
83. García-Ordás MT, Bayón-Gutiérrez M, Benavides C, Aveleira-Mata J, Benítez-Andrades JA (2023) Heart disease risk prediction using deep learning techniques with feature augmentation. Multimed Tools Appl 82(20):31759–31773. https://doi.org/10.1007/s11042-023-14817-z
84. Almazroi AA, Aldhahri EA, Bashir S, Ashfaq S (2023) A clinical decision support system for heart disease prediction using deep learning. IEEE Access 11:61646–61659. https://doi.org/10.1109/ACCESS.2023.3285247
85. Reddy KVV, Elamvazuthi I, Aziz AA, Paramasivam S, Chua HN, and Pranavanand S, (2021) Prediction of heart disease risk using machine learning with correlation-based feature selection and optimization techniques, In: 2021 7th International conference on signal processing and communication, ICSC 2021, Institute of Electrical and Electronics Engineers Inc., pp. 228–233. https://doi.org/10.1109/ICSC53193.2021.9673490.
86. Yadav DC, Pal S (2020) Prediction of heart disease using feature selection and random forest ensemble method. Int J Pharmaceut Res 12(4):56–66. https://doi.org/10.31838/ijpr/2020.12.04.013
87. Almustafa KM (2020) Prediction of heart disease and classifiers' sensitivity analysis. BMC Bioinform 21(1):1–18. https://doi.org/10.1186/s12859-020-03626-y
88. Burse K, Kirar VPS, Burse A, Burse R (2019) Various preprocessing methods for neural network based heart disease prediction, smart innovations in communication and computational sciences. Adv Intell Syst Comput. https://doi.org/10.1007/978-981-13-2414-7_6
89. Abdar M, Książek W, Acharya UR, Tan RS, Makarenkov V, Pławiak P (2019) A new machine learning technique for an accurate diagnosis of coronary artery disease. Comput Methods Programs Biomed. https://doi.org/10.1016/j.cmpb.2019.104992

90. Uddin MdA et al (2024) Machine learning based diabetes detection model for false negative reduction. Biomed Mater Dev 2(1):427–443. https://doi.org/10.1007/s44174-023-00104-w
91. Saboor A, Usman M, Ali S, Samad A, Abrar MF, Ullah N (2022) A method for improving prediction of human heart disease using machine learning algorithms. Mobile Inform Syst. https://doi.org/10.1155/2022/1410169
92. Md AQ, Kulkarni S, Joshua CJ, Vaichole T, Mohan S, Iwendi C (2023) Enhanced preprocessing approach using ensemble machine learning algorithms for detecting liver disease. Biomedicines. https://doi.org/10.3390/biomedicines11020581
93. Paul AK, Shill PC, Rabin MRI, Akhand MAH, (2016) Genetic algorithm based fuzzy decision support system for the diagnosis of heart disease, In: 2016 5th International conference on informatics, electronics and vision, ICIEV 2016, pp. 145–150, https://doi.org/10.1109/ICIEV.2016.7759984.
94. Alaa AM, Bolton T, Di Angelantonio E, Rudd JHF, van der Schaar M (2019) Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK Biobank participants. PLoS ONE 14(5):1–17. https://doi.org/10.1371/journal.pone.0213653
95. Amin R, Yasmin R, Ruhi S, Rahman MH, Reza MS (2023) Prediction of chronic liver disease patients using integrated projection based statistical feature extraction with machine learning algorithms. Inform Med Unlocked. https://doi.org/10.1016/j.imu.2022.101155
96. Fitriyani NL, Syafrudin M, Alfian G, Rhee J (2019) Development of disease prediction model based on ensemble learning approach for diabetes and hypertension. IEEE Access 7:144777–144789. https://doi.org/10.1109/ACCESS.2019.2945129
97. Fitriyani NL, Syafrudin M, Alfian G, Rhee J (2020) HDPM: an effective heart disease prediction model for a clinical decision support system. IEEE Access 8:133034–133050. https://doi.org/10.1109/ACCESS.2020.3010511
98. Reddy KVV, Elamvazuthi I, Aziz AA, Paramasivam S, Chua HN, (2021) Heart disease risk prediction using machine learning with principal component analysis, In: International conference on intelligent and advanced systems: enhance the present for a sustainable future, ICIAS 2021, Institute of Electrical and Electronics Engineers Inc., https://doi.org/10.1109/ICIAS49414.2021.9642676.
99. Karna VVR, Paramasivam S, Elamvazuthi I, Chua HN, Aziz AA, Satyamurthy P, (2022) Novel feature engineering for heart disease risk prediction using optimized machine learning, In: 2022 International conference on future trends in smart communities, ICFTSC 2022, Institute of Electrical and Electronics Engineers Inc., pp. 158–163. https://doi.org/10.1109/ICFTSC57269.2022.10040063.
100. James B, Yoshua B (2012) Random search for hyper-parameter optimization, J Mach Learn Res, 13: 281–305, [Online]. http://dl.acm.org/citation.cfm?id=2188395
101. Wu J, Chen XY, Zhang H, Xiong LD, Lei H, Deng SH (2019) Hyperparameter optimization for machine learning models based on Bayesian optimization. J Electron Sci Technol 17(1):26–40. https://doi.org/10.11989/JEST.1674-862X.80904120
102. Aljanabi M, Qutqut M, Hijjawi M (2018) Machine learning classification techniques for heart disease prediction: a review. Int J Eng Technol 7:5373–5379. https://doi.org/10.14419/ijet.v7i4.28646
103. Ahamed J, Koli AM, Ahmad K, Jamal MA, Gupta BB (2022) CDPS-IoT: cardiovascular disease prediction system based on IoT using machine learning. Int J Interactive Multimed Artif Intell 7(4):78–86. https://doi.org/10.9781/ijimai.2021.09.002
104. Ananey-Obiri D, Sarku E (2020) Predicting the presence of heart diseases using comparative data mining and machine learning algorithms. Int J Comput Appl 176(11):17–21. https://doi.org/10.5120/ijca2020920034
105. Sneha N, Gangil T (2019) Analysis of diabetes mellitus for early prediction using optimal features selection. J Big Data. https://doi.org/10.1186/s40537-019-0175-6
106. Zhang S (2020) Cost-sensitive KNN classification. Neurocomputing 391:234–242. https://doi.org/10.1016/j.neucom.2018.11.101
107. Uddin S, Khan A, Hossain ME, Moni MA (2019) Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 19(1):1–16. https://doi.org/10.1186/s12911-019-1004-8
108. Wang X et al (2021) Exploratory study on classification of diabetes mellitus through a combined random forest classifier. BMC Med Inform Decis Mak 21:105. https://doi.org/10.1186/s12911-021-01471-4
109. Perumal R (2020) Early prediction of coronary heart disease from cleveland dataset using machine learning techniques. Int J Adv Sci Technol 29(06):4225–4234
110. Kégl B, (2014) The return of ADABOOST.MH: multi-class hamming trees, In: 2nd International conference on learning representations, ICLR 2014—conference track proceedings, https://doi.org/10.48550/arXiv.1312.6086
111. Perales-González C, Carbonero-Ruz M, Becerra-Alonso D, Pérez-Rodríguez J, Fernández-Navarro F (2019) Regularized ensemble neural networks models in the extreme learning machine framework. Neurocomputing 361:196–211. https://doi.org/10.1016/j.neucom.2019.06.040

ARTICLES FOR FACULTY MEMBERS

AN IMPROVED MLP-RFE MODEL FOR ENHANCED ACCURACY IN HEART DISEASE PREDICTION USING DEEP LEARNING

A comprehensive review of machine learning for heart disease prediction: challenges, trends, ethical considerations, and future directions. / Kumar, R., Garg, S., Kaur, R., Johar, M. G. M., Singh, S., Menon, S. V., Kumar, P., Hadi, A. M., Hasson, S. A., & Lozanović, J.

Frontiers in Artificial Intelligence
Volume 8 1583459 (2025) Pages 1-31
https://doi.org/10.1007/s10126-025-10555-5
(Database: Frontiers Media SA)

Frontiers in Artificial Intelligence

TYPE Review
PUBLISHED 13 May 2025
DOI 10.3389/frai.2025.1583459

OPEN ACCESS

EDITED BY
Tim Hulsen,
Rotterdam University of Applied Sciences,
Netherlands

REVIEWED BY
Minjae Yoon,
Seoul National University Bundang Hospital,
Republic of Korea
Annalisa Santucci,
University of Siena, Italy

*CORRESPONDENCE
Jasmina Lozanović
✉ jasmina.lozanovic@gmail.com
Rupinder Kaur
✉ rupinderkaur1588@gmail.com

RECEIVED 28 February 2025
ACCEPTED 24 April 2025
PUBLISHED 13 May 2025

CITATION
Kumar R, Garg S, Kaur R, Johar MGM, Singh S, Menon SV, Kumar P, Hadi AM, Hasson SA and Lozanović J (2025) A comprehensive review of machine learning for heart disease prediction: challenges, trends, ethical considerations, and future directions. *Front. Artif. Intell.* 8:1583459. doi: 10.3389/frai.2025.1583459

A comprehensive review of machine learning for heart disease prediction: challenges, trends, ethical considerations, and future directions

Raman Kumar[1,2], Sarvesh Garg[3], Rupinder Kaur[4]*, M. G. M. Johar[5], Sehijpal Singh[1,6], Soumya V. Menon[7], Pulkit Kumar[8,9], Ali Mohammed Hadi[10], Shams Abbass Hasson[11] and Jasmina Lozanović[12]*

[1]Department of Mechanical and Production Engineering, Guru Nanak Dev Engineering College, Ludhiana, India, [2]Jadara Research Center, Jadara University, Irbid, Jordan, [3]Department of Computer Science and Engineering, Guru Nanak Dev Engineering College, Ludhiana, India, [4]Department of Information Technology, Guru Nanak Dev Engineering College, Ludhiana, India, [5]Management and Science University, Shah Alam, Malaysia, [6]Department of Mechanical Engineering, Graphic Era (Deemed to be University), Dehradun, India, [7]Department of Chemistry and Biochemistry, School of Sciences, JAIN (Deemed to be University), Bangalore, India, [8]Department of Electrical Engineering, Chandigarh University, Mohali, India, [9]Chitkara University Institute of Engineering and Technology, Centre for Research Impact & Outcome, Chitkara University, Rajpura, India, [10]Department of Pharmacy, Mazaya University College, Dhiqar, Iraq, [11]Laboratories Techniques Department, College of Health and Medical Techniques, Al-Mustaqbal University, Babylon, Iraq, [12]Department of Engineering, FH Campus Wien - University of Applied Sciences, Vienna, Austria

This review provides a thorough and organized overview of machine learning (ML) applications in predicting heart disease, covering technological advancements, challenges, and future prospects. As cardiovascular diseases (CVDs) are the leading cause of global mortality, there is an urgent demand for early and precise diagnostic tools. ML models hold considerable potential by utilizing large-scale healthcare data to enhance predictive diagnostics. To systematically investigate this field, the literature is organized into five thematic categories such as "Heart Disease Detection and Diagnostics," "Machine Learning Models and Algorithms for Healthcare," "Feature Engineering and Optimization Techniques," "Emerging Technologies in Healthcare," and "Applications of AI Across Diseases and Conditions." The review incorporates performance benchmarking of various ML models, highlighting that hybrid deep learning (DL) frameworks, e.g., convolutional neural network-long short-term memory (CNN-LSTM) consistently outperform traditional models in terms of sensitivity, specificity, and area under the curve (AUC). Several real-world case studies are presented to demonstrate the successful deployment of ML models in clinical and wearable settings. This review showcases the progression of ML approaches from traditional classifiers to hybrid DL structures and federated learning (FL) frameworks. It also discusses ethical issues, dataset limitations, and model transparency. The conclusions provide important insights for the development of artificial intelligence (AI) powered, clinically applicable heart disease prediction systems.

KEYWORDS

heart disease prediction, machine learning (ML), deep learning models, federated learning, explainable artificial intelligence (XAI)

1 Introduction

1.1 Background of the study

Cardiovascular diseases (CVDs) cause around 17.9 million deaths each year, accounting for 32% of deaths worldwide. Heart disease continues to be one of the most significant health problems globally and nationally, as it is the leading cause of death around the globe and in the US. In 2021 alone, coronary heart disease was accountable for approximately 9 million deaths. In the US, coronary heart disease caused 1 out of 5 deaths in 2022, affecting all genders and races. The magnitude of this issue is enormous; in the United States alone, heart disease caused approximately two hundred and 52.2 billion dollars in direct and indirect costs from 2019 to 2020 (Khan Minhas et al., 2024). The prevalence of CVDs in the US is anticipated to increase sharply, as 61% of adults are expected to be hypertensive by 2050. The worldwide burden of CVDs is expected to rise by 90% from 2025 to 2050, increasing the number of deaths from 20.5 million in 2025 to 35.6 million by 2050. Therefore, immediate attention needs to be put towards effective heart disease preventive measures, greater detection capabilities, and fairness in healthcare access (Roth et al., 2020; Al-Ajlouni et al., 2024).

Many patients can be kept alive through effective healthcare interventions. This, however, requires early detection (Ferdous Azam and Sarwar, 2023). By taking proactive measures, one can help alleviate the bad consequences of the disease, improve the possible prognosis, and save money to be spent on treating the problem. Unfortunately, most diagnostic methods, such as Electrocardiogram (ECG), echocardiograms, and stress testing, need considerable time and skill to administer, and even then, accurate diagnosis may still not be achieved (Dala Ali et al., 2023; Faraji et al., 2023). Such limitations are even more pronounced in underdeveloped areas where such facilities are hard to come by. Machine Learning (ML), a subfield of artificial intelligence (AI), provides solutions to such problems (Pathirana et al., 2018; Pathirana et al., 2019). Complex ML algorithms can recognize intricate structures and correlations existing within a vast data set that are not readily available using traditional techniques. Such an attribute enables chronic diseases of the heart to be diagnosed at intervals much earlier than is possible when patients start showing symptoms. Thus, ML can facilitate the adoption of preventive measures and strides towards a patient-centered approach. Further, ML gives global health a powerful tool for applying affordable and efficient diagnostic technology to populations that need it most (Naruka et al., 2022). Figure 1 depicts the worldwide prevalence (in millions) of major cardiovascular conditions as of 2021. Coronary heart disease remains the most prevalent, impacting roughly 250 million people, followed by peripheral arterial disease (110 million), stroke (94 million), and atrial fibrillation (53 million) (Jagannathan et al., 2019). These figures highlight the significant global challenge posed by CVDs and emphasize the urgent need for effective predictive models driven by ML and AI to facilitate early diagnosis and prompt intervention. Incorporating these technologies into healthcare systems can significantly reduce mortality and enhance patient outcomes.

1.2 Role of machine learning in healthcare

The adoption of electronic health records (EHRs) and wearable devices, and sophisticated imaging technologies is aiding the healthcare industry in data management (Bai and Mardini, 2024). The ability of ML to use such data to improve clinical processes and patient interaction is astounding. In predicting heart disease, diverse data sources are harnessed by ML models such as (Pathirana et al., 2019; Dissanayake and Johar, 2023):

- Clinical Data: Data about the patient that includes demographics, medical history, lab results, and medications.

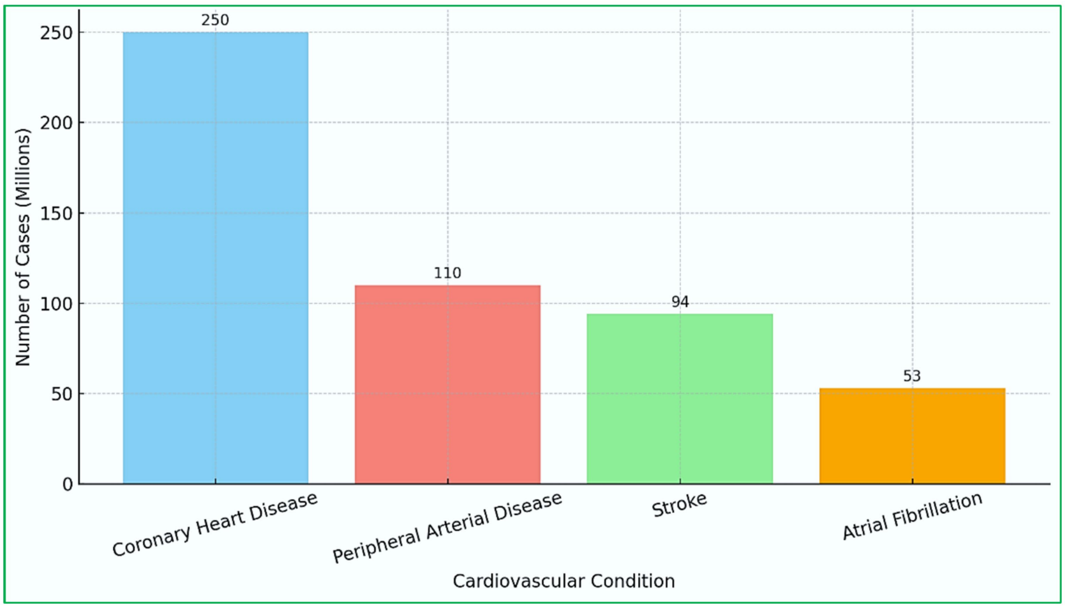


FIGURE 1
Global prevalence of major cardiovascular diseases.

- Imaging Data: Echocardiograms, angiograms, and computed tomography (CT) scans.
- Biometric Signals include ECG, heart rate variability, and blood pressure.
- Data from wearable devices includes daily physical activities, sleep patterns, and vital signs.

Devices that classify people as high risk and low risk of heart disease are based on supervised learning techniques like random forest (RF), support vector machine (SVM), and neural network (NN). Slicers of the ML set are DL, which enhances a machine's learning capabilities. For instance, convolutional neural network (CNN) has been used to identify arrhythmias from ECG signals with great precision. Furthermore, clustering techniques consider the unsupervised learning approach in which algorithms detect groups within a patient population that can be linked to specific risk levels or responses to treatment, thus paving the way for targeted medicine (Dissanayake and Johar, 2021; Dissanayake et al., 2023). Another innovative area is the application of reinforcement learning (RL) to improve treatment plans and the allocation of resources in healthcare settings. Figure 2 depicts ML in heart disease prediction, from data collection to model deployment (Nadeem et al., 2021; Kwon and Dong, 2022). Figure 3 presents an integrated heart disease prediction (Siramshetty et al., 2018; Benhar et al., 2020).

1.3 Objectives and scope of the review

The review aims to incorporate findings from previous research studies on heart diseases while creating, developing, and applying ML technologies that predict heart diseases. Grouping these studies into thematic clusters may help understand the advancements in the field, highlight strengths and challenges, and lay out the following objectives.

- *Categorization of Research:* Formulating a primary information scheme by grouping the studies into five main clusters.
- *Analysis of ML Models:* Performed in-depth analysis of the models in terms of algorithms, techniques, components of the systems, and their merits and demerits alongside real-world applications.
- *Feature Engineering and Optimization:* Striving to improve model performance through feature selection, dimensionality reduction, and hyperparameter tuning.
- *Emerging Technologies:* Exploiting the effects of innovations such as FL, quantum computing (QC), and Internet of Things (IoT) devices on the diagnostics of heart diseases.
- *Other Applications of AI:* AI's involvement in treating diseases like cancer, diabetes, and other neurological disorders should be emphasized to better understand how heart disease prediction can be approached.

This review proposes to find patterns, gaps, and trends by incorporating these clusters in literature patterns to provide better insights for evolving studies and their implementations.

1.4 Comparison with existing literature and novel contributions

There is a substantial body of literature on ML applications in the healthcare sector, with review articles examining ML usage in healthcare diagnoses, predictions, and treatments. The literature is extensive. While many studies have been conducted, a gap remains in the focus on heart disease prediction through ML systems. We directly compare what this review achieves to existing works to address this gap.

1.4.1 Comparative analysis of related reviews

To illustrate how this review differentiates from prior works, Table 1 compares key literature, focusing on scope, methodology, datasets, and technological advancements.

1.4.2 Novel contributions of this review

Though other reviews provide essential insights into ML-based heart disease prediction, they tend to suffer from some drawbacks. Unlike other works, our work is unique in the following aspects.

1.4.2.1 Newer dataset analysis

o This review covers newer datasets such as IoT-based healthcare data, wearable sensor datasets, and FL-enabled datasets, unlike

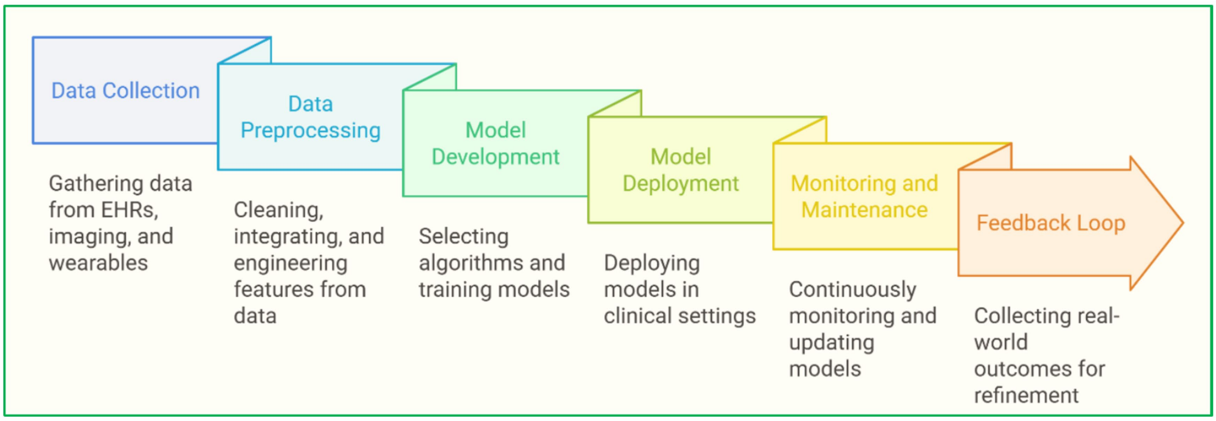


FIGURE 2
ML in heart disease prediction, from data collection to model deployment.

other reviews focusing on classical datasets like UCI heart disease data and Framingham data.
- Covers real regulatory challenges such as imbalance and noise.

1.4.2.2 Emerging technology integration

This review deeply analyzes new emerging AI tools like:

- FL—privacy-preserving ML.
- QC—high-speed disease modeling.
- Explainable AI (XAI)—model understanding and trust.
- Other reviews hardly address the confluence of AI with regulatory frameworks like General Data Protection Regulation (GDPR), Food and Drug Administration (FDA), and Health Insurance Portability and Accountability Act (HIPAA).

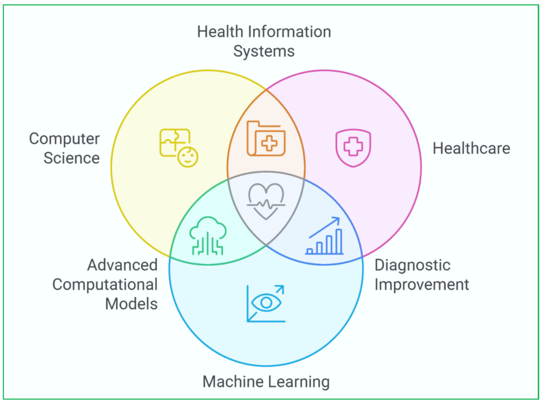


FIGURE 3
Integrated heart disease prediction.

1.4.2.3 Ethical and regulatory considerations

- Covers biases and patient privacy ML in healthcare ethics issues and explainability.
- Provides an in-depth discussion of the legal aspects of AI implementation in healthcare.

1.4.2.4 Structured thematic classification

- Unlike other systematic reviews, this piece of work classifies and indexes research into 5 different thematic clusters.
- Methods and Algorithms of Heart Disease Detection and Diagnostics.
- ML Models and Algorithms.
- Emerging Feature Molding and Engineering.
- Advanced Emerging Technology.
- Multi-Disease AI Technology Applications.

1.4.2.5 Bridging the gap between research and clinical application

- Prior works focus primarily on ML model accuracy, while this review focuses on actual clinical application.
- Explains how different hospitals, medical practitioners, and policymakers can use ML-based systems for diagnosis in real-life settings.

1.4.3 Heatmap provides feature correlations

A heatmap is used in a dataset with heart disease features to represent the relationships of different attributes. Redundancy may exist for the features with high correlations, such as cholesterol and blood pressure, but weakly correlated features

TABLE 1 Comparison of existing literature on ML for heart disease prediction.

References	Scope	Methodology	Datasets used	Technological focus	Challenges discussed	Novel contributions
Hajiarbabi (2024)	Overview of ML for CVDs	A systematic review of ML models	UCI Heart Disease, Framingham, PhysioNet	DL and supervised learning	Model performance and dataset limitations	Lacks discussion on ethical challenges and explainability
Ahsan and Siddique (2022)	ML applications in healthcare	Meta-analysis of diagnostic accuracy	Multiple EHR datasets	Traditional ML SVM, RF, K-nearest neighbor (KNN)	Data imbalance and interpretability	It does not explore FL or QC
Jafari et al. (2023)	DL models for ECG-based heart disease detection	Experimental comparison	Physikalisch-Technische Bundesanstalt Extended ECG Dataset (PTB-XL), MIT-BIH, ECG datasets	CNN, LSTM, Transfer Learning	Bias, data augmentation	Lacks coverage on FL and real-world integration
Present review	Comprehensive review of ML for heart disease prediction	Thematic classification with clustering	EHR, ECG, Cleveland, Framingham, Emerging IoT-based datasets	Supervised and unsupervised learning, DL, FL, QC	Ethical concerns, dataset quality, integration into clinical practice	Integrates AI trends (IoT, FL, QC) and highlights regulatory, transparency, and privacy challenges in ML adoption

suggest that they would be independent in their ability to predict. Figure 4 presents a feature correlation heatmap that depicts the relationships between essential clinical variables like cholesterol, blood pressure, and age, which are vital for enhancing feature selection in heart disease prediction models.

- Proposition: 'Blood Pressure and Cholesterol,' a high correlation suggests potential multicollinearity and multivariate relationships.
- Counter Proposition: 'Age and Resting ECG,' the low correlation suggests a weak direct association.
- Supporting Proposition: 'Smoking and Heart disease' have a moderate correlation, which supports known medical findings.

2 Methodology

2.1 Literature search and inclusion criteria

The review methodology entailed a comprehensive literature search. The search involved a comprehensive query in Scopus using the search string:

((TITLE-ABS-KEY ('heart disease prediction') AND ('ML') OR ('machine learning')) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English"))).

It filtered relevant articles on heart disease prediction using ML, limited to English-language articles and journal articles. The curated results form the foundation for the insights and analyses presented in this review. Studies were excluded if they lacked adequate methodological detail, were not peer-reviewed, or were focused on unrelated topics. Quantitative and qualitative insights were extracted to identify trends, assess proposed methods' effectiveness, and uncover existing research gaps.

2.2 Classification of clusters

The identified studies have been classified into five major clusters based on their primary focus using keywords. Table 2 depicts keywords considered and cluster names based on keywords and the occurrence of keywords. The first cluster, '*Heart Disease Detection and Diagnostics*', includes advanced research on diagnostics and predictions concerning heart-related conditions. The second cluster, '*Machine Learning Models and Algorithms for Healthcare*,' studies the

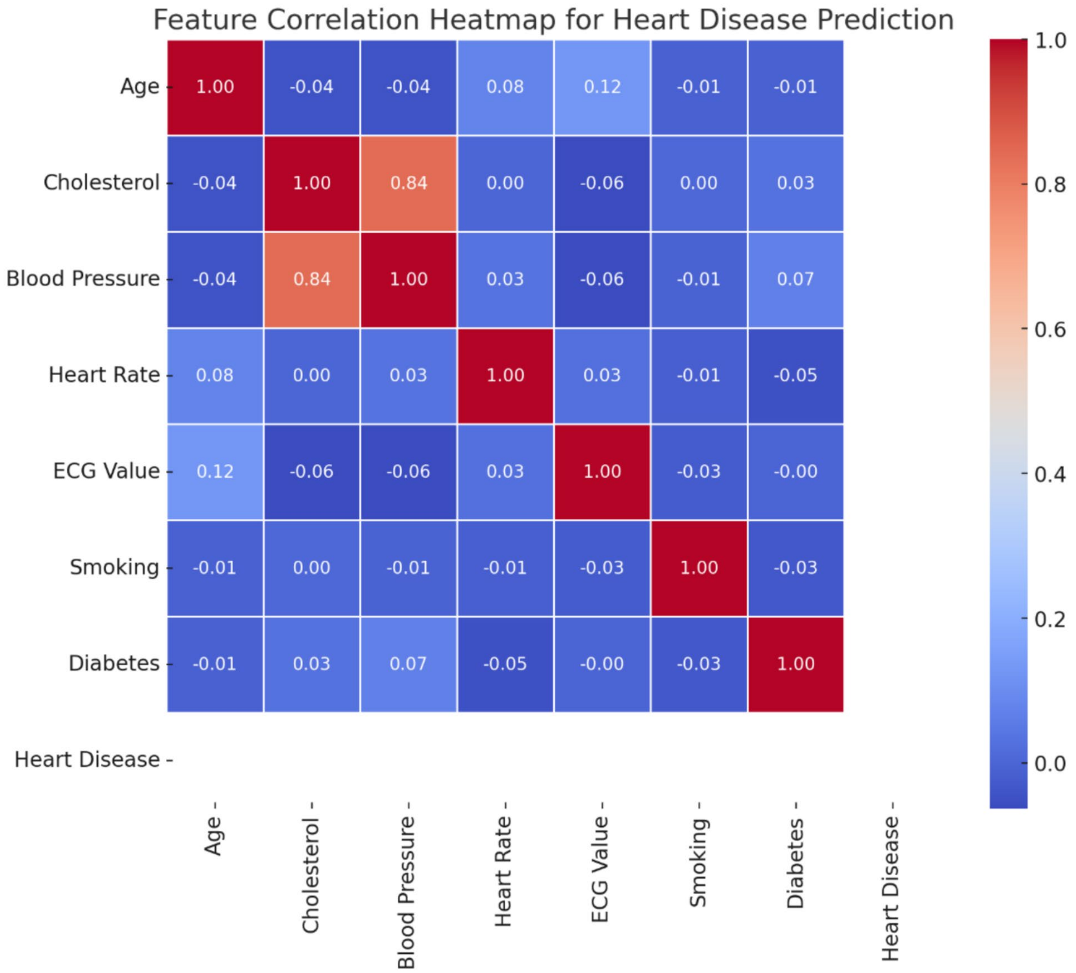


FIGURE 4
Feature correlation heatmap for heart disease prediction.

TABLE 2 Clusters of ML for heart disease prediction.

Cluster number	Cluster name	Keywords	Occurrence
1	Heart Disease Detection and Diagnostics	Heart disease classification, heart disease detection, Heart failure, Arrhythmia, Valvular heart disease, Electrocardiogram (ECG), PCG signal analysis, Cleveland and Framingham Datasets, Acute Myocardial Infarction, Cardiac disease	12
2	Machine Learning Models and Algorithms for Healthcare	Machine learning, Artificial intelligence, Deep learning, Logistic Regression, Decision tree, Support vector machine (SVM), Random Forest classifier, Gradient descent, Sparse autoencoder, Attention mechanisms, Transformer-based models, Recurrent neural network (RNN)	13
3	Feature Engineering and Optimization Techniques	Feature extraction, Feature selection, Principal component analysis (PCA), SHAP, Genetic Algorithm (GA), Grey wolf algorithm, Particle Swarm Optimization (PSO), Sand Cat Swarm Optimization, Coati optimization algorithm, Kepler optimization, Canonical Correlation, Lasso regression	12
4	Emerging Technologies in Healthcare	Healthcare 4.0, Quantum computing, Federated learning, 5G, Wearable devices, Internet of Things (IoT), Cloud platform, Automated Sequential Cryptography, Cloud security, Decryption, Firefly Algorithm, Modified Blowfish	10
5	Applications of AI Across Diseases and Conditions	Diabetes, Diabetic retinopathy, Breast cancer, Parkinson's disease, Stroke prediction, cardiovascular diseases (CVD), Disease Prediction, Early Diagnosis, Scalability in machine learning, Multimodal feature fusion, Cross-modal transfer learning	11

different types of ML model frameworks, including supervised and DL models. The third cluster, '*Feature Engineering and Optimization Techniques*,' discusses papers on feature selection, dimensionality reduction, and other optimization methods in model performance improvement. The fourth cluster, '*Emerging Technologies in Healthcare*', reviews articles on modern innovations like QC, FL, and IoT in heart disease prediction and other applications. The last or fifth cluster, '*Applications of AI Across Diseases and Conditions*,' gives more context by including AI's applications in managing and predicting other diseases such as cancer, diabetes, and neurological disorders. This categorization is shown in Figure 5, facilitating a more profound literature review and allowing a better understanding of each cluster's contributions and limitations.

3 Cluster wise insights: literature review

3.1 Cluster 1: heart disease detection and diagnostics

Technological advancements and methodologies provide innovative approaches and tools for the early detection and prediction of cardiovascular ailments. This group focuses on new imaging and ML model development, which improves diagnostic precision and appropriate timing of interventions to improve the prognosis for patients. Heart disease displays itself as more than just coronary artery disease (CAD); it includes arrhythmias and even heart failure. Therefore, it is essential to stress the multi-dimensional approach to conditions related to heart diseases. As outlined in Figure 6, critical steps are monitoring ECG for electrical activity, echocardiograms for the anatomical view, stress tests for function, cardiac catheterization for the coronary details, and blood tests for the troponin marker, along with the cholesterol marker, which are the most important for attention.

Table 3 shows key insights and performance metrics in heart disease detection and diagnostics. The classification of heart disease proved accurate with DL algorithms and Sand Cat Swarm Optimization (SCSO) for feature selection. Important features were identified using patient pathology data, and models, including CNN, PCA, Restricted Boltzmann Machine (RBM), and deep convolutional generative adversarial networks (DCGAN), analyzed intricate correlations that improved the accuracy of predictions. The method enhanced the reliability of heart disease prognosis through metrics such as accuracy and F1-score (Baviskar et al., 2023). The detection of CVD was improved using both RF and eXtreme Gradient Boosting (XGB) on ECG datasets, particularly Physionet 2016, PASCAL, and MIT-BIH. Pre-processing feature extraction with empirical wavelet transform (EWT), discrete wavelet transform (DWT), and SHapley Additive exPlanations (SHAP) will improve a prediction model's accuracy, and a significant peak can reach up to 98.25 AUC for XGB-based proposed models (Majhi and Kashyap, 2024). A novel Wolf-based Generative Adversarial System (WbGAS), was developed to classify heart diseases, using ECG data to identify normal sinus rhythm, arrhythmia, and congestive heart failure (Goud et al., 2024). A hyperparameter-tuned CNN-based Inception Network model was created to diagnose heart disorders with heart sound data from standard repositories. The model achieves 99.65% accuracy, 98.8% sensitivity, and 98.2% specificity, surpassing the other classifiers (Roy et al., 2024). An electronic stethoscope has been developed, integrated with Raspberry Pi 4B and a CNN-based EfficientNet-B3 model for diagnosing valvular heart diseases. The system reported an accuracy of 99.35%, sensitivity of 98.84%, and specificity of 98.23%, with real-time Phonocardiogram (PCG) signal analysis and cloud-based data storage (Roy et al., 2023). Fatima and Siddiqi (2024) evaluated ML techniques for predicting Myocardial Infarction (MI) and analyzing risk factors using a dataset of 350 individuals, including MI and non-MI patients of both genders.

Maran (2023) developed advanced CNN-based models to analyze multimodal data, including BMI, ECG, and PTB, achieving 98%

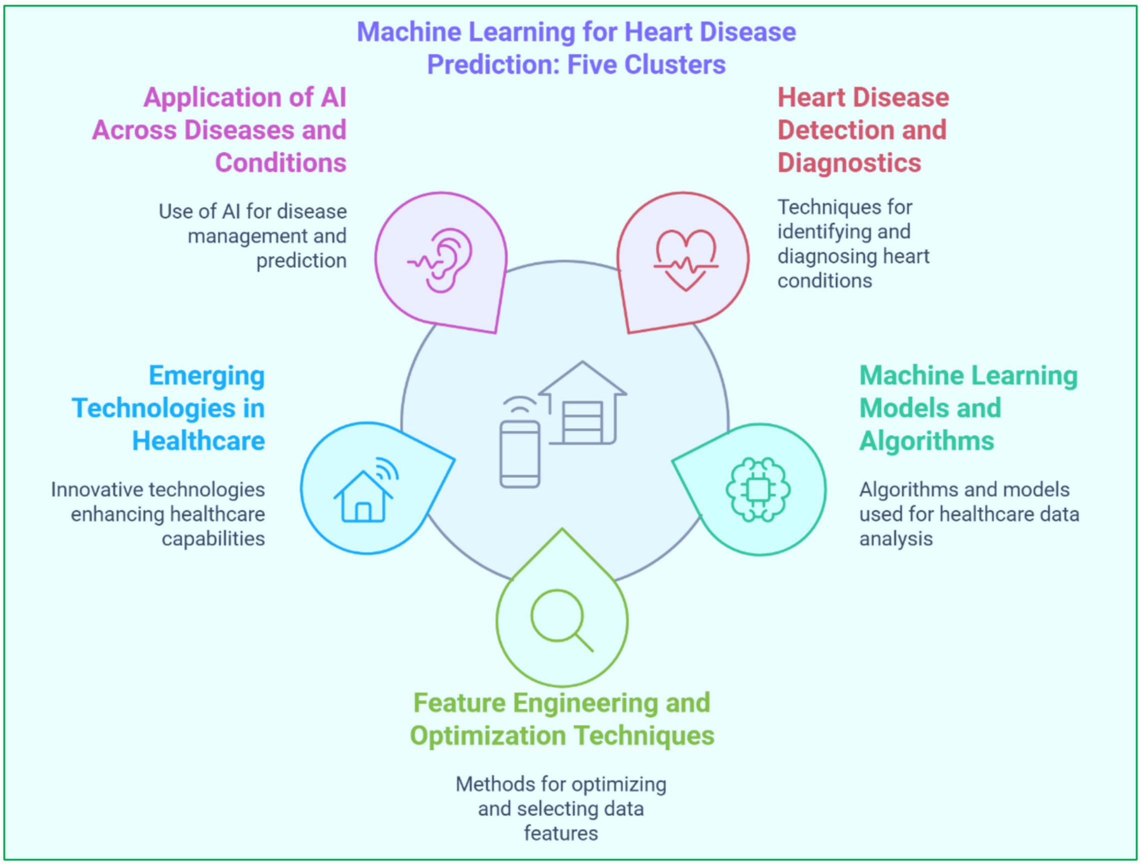


FIGURE 5
Machine learning for heart disease prediction: five clusters.

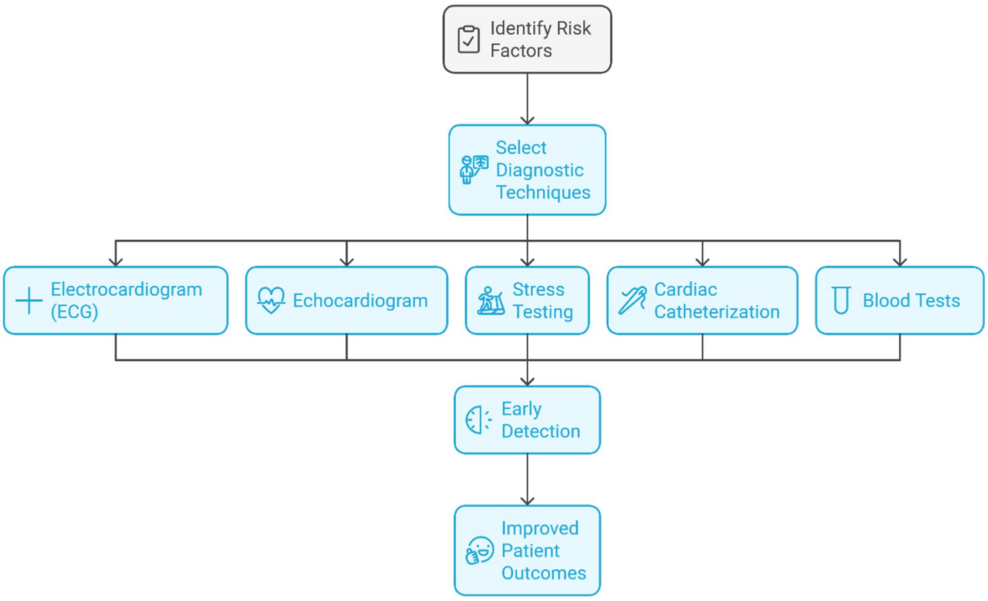


FIGURE 6
Diagnostic techniques and multidimensional approach to heart disease management.

TABLE 3 Key insights and performance metrics: cluster 1 'heart disease detection and diagnostics'.

Research focus	Key methods/ models	Dataset(s) used	Performance metrics	Key insights	Ref.
Heart Disease Detection with SCSO	DL + SCSO, CNN, PCA, RBM, DCGAN	Patient Pathology Data	High Accuracy, Improved F1-score	Enhanced feature selection and reliable prognosis	Baviskar et al. (2023)
ECG Analysis with RF and XGB	RF, XGB, EWT, DWT, SHAP	Physionet 2016, PASCAL, MIT-BIH	Up to 98.25 AUC	Effective feature extraction and prediction	Majhi and Kashyap (2024)
WbGAS for Heart Disease Classification	Wolf-based Generative Adversarial System (WbGAS), Wolf Fitness Function	ECG Data	High Specificity, Precision, Recall, and Accuracy	Outperforms traditional ML methods	Goud et al. (2024)
Hyperparameter-tuned CNN Inception Network	CNN-based Inception Network	Heart Sound Data	99.65% Accuracy, 98.8% Sensitivity, 98.2% Specificity	Efficient handling of high-dimensional data	Roy et al. (2024)
Myocardial Infarction Prediction	Ensemble Classifiers	Dataset of 350 Individuals (MI & Non-MI)	Improved Gender-Specific Precision, High Accuracy	Optimized early MI detection	Fatima and Siddiqi (2024)
Multimodal Data Analysis with CNNs	CNN, Python	BMI, ECG, PTB	98% Accuracy	Reliable prediction from multimodal data	Maran (2023)
Autoencoder + DenseNet on UCI Cleveland Dataset	Autoencoder, DenseNet	UCI Cleveland	99.67% Mean Accuracy, 99.99% Test Accuracy	Outstanding performance requires more	Saleh Alghamdi et al. (2024)
CNN-Bi-LSTM with Attention Mechanisms	CNN-Bi-LSTM, Newton–Raphson Optimizer	Cleveland, Framingham	95.3% Accuracy (Cleveland), 98.1% Accuracy (Framingham)	Optimized cardiac disease prediction	Kayalvizhi et al. (2024)

accuracy in predicting heart diseases. The models were trained on 80% of the data, validated on 20%. Saleh Alghamdi et al. (2024) proposed a method by integrating autoencoder and DenseNet architectures to predict heart disease based on the UCI Cleveland dataset, obtaining a mean accuracy of 99.67% and test accuracy of 99.99%.

3.2 Cluster 2: machine learning models and algorithms for healthcare

Table 4 shows Advances in ML Models and Algorithms for Healthcare. A three-stage wireless body area network (WBAN)-based heart disease prediction model was developed. The three stages include data aggregation, channel selection, and prediction. Channel selection was optimized using the Tunicate Swarm-Sail Fish Optimization algorithm, and statistical features were extracted from the data using a weighted entropy-based method. An enhanced RNN tuned with Tunicate Swarm-Sail Fish Optimization achieved high prediction accuracy (Muthu Ganesh and Nithiyanantham, 2022). This research developed ML models to predict cancer, diabetes, diabetic retinopathy, and heart-related outcomes using EHRs. SVMs achieved 97.08 and 79.75% accuracy for cancer and Pima diabetes datasets, respectively, while Decision Tree (DT) reached 86.42% for heart-related predictions. The study showed that ML could enhance disease prediction and patient outcomes (Sheik Abdullah et al., 2024).

The real-time data will be utilized to predict heart disease with the support of the ML algorithm, as well as attributes like BP, sugar, and heartbeat. The attributes are applied in the dataset on 300 instances with 14. It is used to train and test an R model. Accuracy Evaluation: Accuracy is measured as the basis of several classifiers (Jayakiran et al., 2019). Napa et al. (2024) evaluated recursive feature elimination (RFE) for classifying chronic heart disease based on the Cleveland Hungarian CHD dataset. Different methods of supervised learning have been tested. The KNN model and the DT attained 89.91%. Seeli and Thanammal (2024) explored ML algorithms' performance, specifically artificial neural network (ANN) and logistic regression (LR) for disease prediction. Without scaling, ANN obtained 86.13% accuracy in the heart disease prediction, and when ensemble normalization and standardization were applied, the improvement in ANN accuracy was 98.81%. Patil et al. (2023) presented a prediction model for heart disease incorporating IoT and ML techniques using data from several sensors: BP monitors, blood oxygen sensors, and EEGs. Hybrid feature extraction methods were combined with ML algorithms, where RNN yields higher accuracy than the traditional approach using methods like SVM, Naive Bayes (NB), and RF.

3.3 Cluster 3: feature engineering and optimization techniques

Table 5 shows Feature Engineering and Optimization Techniques in Heart Disease Prediction. The study proposed a method that combines PCA and feature selection to reduce the dimensionality of data and improve the prediction of coronary heart disease. The model using classifiers such as PCA, RF, DT, and AdaBoost achieved 96% accuracy and outperformed traditional precision, recall, and AUC

TABLE 4 Advances in ML models and algorithms for healthcare.

Research focus	Key methods	Dataset(s) used	Performance metrics	Key insights	Ref.
WBAN-based Heart Disease Prediction	Tunicate Swarm-Sail Fish Optimization, RNN, Weighted Entropy-based Features	WBAN Data	High Prediction Accuracy	Optimized channel selection improves prediction performance	Muthu Ganesh and Nithiyanantham (2022)
ML for Cancer, Diabetes, and Heart Disease Prediction	SVM, DT	Cancer, Pima Diabetes, and Heart Disease Datasets	97.08% (Cancer), 79.75% (Diabetes), 86.42% (heart disease)	ML enhances disease prediction and patient outcomes	Sheik Abdullah et al. (2024)
Real-Time Data Prediction Using ML	DT, NB, RF, KNN, NN	Custom Dataset (300 Instances, 14 Attributes)	Performance Varies Across Classifiers	Useful in predicting heart disease using real-time attributes	Jayakiran et al. (2019)
Chronic Heart Disease Prediction with RFE	RFE, KNN, DT	Cleveland Hungarian CHD Dataset	89.91% Accuracy	RFE improves early CHD prediction	Napa et al. (2024)
ANN and LR for Disease Prediction	ANN, LR, Scaling Methods	Heart Disease Dataset	86.13% (Without Scaling), 98.81% (With Scaling)	Scaling methods enhance ANN performance	Seeli and Thanammal (2024)
IoT and ML for Heart Disease Prediction	IoT Sensors, Hybrid Feature Extraction, RNN, SVM, NB, RF	Sensor Data (BP, Oxygen, EEG)	Higher Accuracy with RNN	Reliable for heart disease detection and classification	Patil et al. (2023)

TABLE 5 Cluster 3 feature engineering and optimization techniques in heart disease prediction.

Research focus	Key methods/ models	Dataset(s) used	Performance metrics	Key insights	Ref.
PCA and Feature Selection for CHD Prediction	PCA, RF, DT, AdaBoost	Coronary Heart Disease Data	96% Accuracy	PCA and feature selection improved precision, recall, and AUC for CHD	Cheekati et al. (2024)
C-CADZ System for CAD Diagnosis	Feature Extraction, SMOTE, RF, Extra Trees	Z-Alizadeh Sani CAD Dataset	97.37% Accuracy	Outperformed prior methods by 5.17%, robust performance for heart disease prediction	Gupta et al. (2022)
Diabetes Detection Using Optimized Classifiers	SVM, KNN, RF, PSO Algorithm for Optimization	Indian Pima Diabetes Dataset	94.27% Detection Rate	Outperformed single classifiers for diabetes prediction	Shimpi et al. (2024)
SMOTE-based Hybrid DL Network for CVD Prediction	SMOTE, Adaptive Coati Optimization, Kepler-Optimized Deep Stacked Recurrent Network	Heat-failure-clinical-records Dataset	95.52% Accuracy	SMOTE-HDL network outperformed existing classifiers	Barfungpa et al. (2024)
DL for Cardiac Disorder Detection	ST-CNN-GAP-5, SHAP Analysis	PTB-XL ECG, Arrhythmia Dataset	93.41% AUC, 95.8% Accuracy, 99.46% AUC	Interpretability with SHAP and better performance than existing models	Anand et al. (2022)
Hybrid CCRF Model for Heart Disease Prediction	Canonical Correlation Analysis, RF, Polynomial Features	–	99.45% Accuracy, Improved Sensitivity, Specificity, Precision, and F1 Score	Maximized feature correlations for heart disease prediction	Vetrithangam et al. (2024)
ALAN Method for Heart Disease Prediction	ANOVA, Lasso Regression, ET-ABDF Model	–	88.0% Accuracy, 89.81% Precision, 96.21% AUC	Superior performance compared to other algorithms	Mandula and Vijaya Kumar (2024)
SCSO and DL for Heart Disease Classification	SCSO, CNN, PCA, GANs	Patient Pathology Data	High Accuracy, Precision, Recall, F1-Score	Enhanced prognosis and reliability in heart disease prediction	Lenin and Venkatasalam (2024)

models. The approach effectively enhanced CHD prediction and patient outcomes (Cheekati et al., 2024). Gupta et al. (2022) proposed a computational intelligence system, C-CADZ, for diagnosing CAD using the Z-Alizadeh Sani CAD dataset. The applied feature extraction, Synthetic Minority Over-sampling Technique (SMOTE) for dealing with class imbalance, and ML classifiers such as RF and Extra Trees resulted in an accuracy of 97.37%. C-CADZ outperformed prior methods by 5.17% and exhibited robust performance, thus making it applicable for heart disease predictions.

Shimpi et al. (2024) proposed a model optimized for diabetes detection via SVM, KNN, and RF classifiers with decision-level fusion. The classifiers were optimized utilizing a PSO algorithm, considering clinical data such as age, BMI, blood pressure, and glucose. The model performed at a diabetes detection rate of 94.27%. It outperformed single classifiers and previous methods on the Indian Pima diabetes dataset. Barfungpa et al. (2024) anticipated a new SMOTE-based hybrid DL network for predicting patient survival in CVD. The SMOTE-HDL network, tested on the Heat-failure-clinical-records dataset, obtained a predictive accuracy of 95.52% that outperformed any existing classifiers. Anand et al. (2022) utilized deep NNs on the PTB-XL ECG dataset for cardiac disorder detection, presenting the ST-CNN-GAP-5 model that resulted in an AUC of 93.41%. The model was tested on an arrhythmia dataset and yielded 95.8% accuracy with an AUC of 99.46%, better than other existing approaches. The SHAP analysis showed that the model is interpretable and reveals critical ECG wave changes that can help make diagnoses in resource-constrained environments. Vetrithangam et al. (2024) proposed the Hybrid CCRF model of heart disease prediction, which applied Canonical Correlation Analysis and RF together. The model represented the non-linear relationships and maximized the feature correlations because it generated polynomial features and synthesized canonical variables. Mandula and Vijaya Kumar (2024) proposed the ALAN method, which combines ANOVA and Lasso regression to identify the most essential features for heart disease prediction. The Extra Trees Adaptive Boosted Decision Forest model reached 88.0% accuracy, 89.81% precision, and 96.21% AUC, which is superior to other algorithms. Lenin and Venkatasalam (2024) utilized DL and SCSO for accurate heart disease classification. SCSO selected key features from patient pathology data, enabling CNNs combined with advanced models like PCA and GANs to predict disease severity.

3.4 Cluster 4: emerging technologies in healthcare

Table 6 shows Emerging Technologies in Healthcare and evaluated Quantum Support Vector Classifier (QSVC) and variational quantum classifier (VQC) for chronic heart disease prediction in healthcare 4.0. QSVC outperformed VQC with an accuracy of 82%, showing the potential of quantum ML in healthcare. Several metrics, such as precision, recall, and F1 score, supported the findings (Munshi et al., 2024a). Bhatt et al. (2024) proposed an AI-enabled stroke prediction architecture using FL based on an ANN model, which uses real stroke cases. The architecture, implemented on healthcare wearable devices, aggregates optimizer weights via a 5G communication channel to enhance performance. It outperformed traditional approaches, achieving 5 to 10% higher accuracy.

The study developed a smart healthcare system for heart disease prediction using Bi-LSTM, which integrated data from IoT devices and electronic clinical records. The system achieved an accuracy of 98.86%, along with high precision, sensitivity, specificity, and F-measure, outperforming existing prediction models (Nancy et al., 2022). Vellore Pichandi et al. (2024) introduced a safe e-healthcare system with accurate HD prediction and enhanced cloud storage security. This classifier achieves a high degree of accuracy as 99.36% through HybBPF-ELM, and Intelligent Encryption Framework enhances cloud data security. The system minimizes encryption and decryption time to process the file by 2.2 GB, at 127.55 s and 452.01 s, respectively. Natarajan et al. (2024) proposed a secure e-healthcare system combining accurate heart disease (HD) prediction and enhanced cloud security. A Hybrid Binary Particle Firefly Optimized Extreme Learning Machine classifier achieved 99.36% accuracy, while an intelligent encryption framework improved data security. The system reduced the processing time for encryption and decryption of a 2.2 GB file to 127.55 s and 452.01 s, respectively. FL has allowed it to develop ML models on distributed datasets, like those in hospitals and mobile devices while maintaining data privacy. This survey reviews prior research on its healthcare applications, including key challenges, methods, and use cases. It outlines existing studies and explores its potential for the healthcare industry (Joshi et al., 2022). CVDs were analyzed using hybrid classical-quantum (CQ) transfer learning models to detect cardiomegaly in chest X-rays. The pre-trained DenseNet-121 integrated with quantum circuits through Qiskit and PennyLane obtained Receiver Operating Characteristic (ROC) and AUC scores of up to 0.93 and accuracies of up to 0.87 on a balanced dataset. Grad-CAM++ heatmaps with QC models showed more trustworthiness, thus supporting possible clinical adoption (Decoodt et al., 2023). AI was used in Healthcare 4.0 for early and accurate disease prediction supported by IoT sensors capturing patient data for ML analysis. The seven-classifier ML model predicted nine fatal diseases, where RF obtained the highest accuracy of 97.62% and AUC of 99.32%. This model is intended to help doctors with early diagnosis and better patient outcomes (Kishor and Chakraborty, 2022).

3.5 Cluster 5: applications of AI across diseases and conditions

Table 7 shows the Application of AI Across Diseases and Conditions. This work presented an intelligent heart disease diagnosis method using an integrated filter-evolutionary search-based feature selection (iFES-FS) and an optimized ensemble classifier. The feature selection combined adaptive threshold information gain (aTIG-FS) and evolutionary gravity-search, and a firefly-driven Firefly-Driven Multi-Objective Multi-Verse Optimizer algorithm optimized the classifier's hyperparameters. This model outperformed the existing methods regarding accuracy, precision, sensitivity, specificity, and ROC curve evaluation (Venkata MahaLakshmi and Rout, 2024). The work presented an optimized dual-directional temporal convolution and attention-based density clustering for predicting and classifying diabetic risk levels. The proposed approach also outperformed the previous methods and obtained 98.21% accuracy, 94.46% recall, and 99.01% F1-score on the five datasets considered (Jenefer et al., 2024).

The wearable sensor-based prototype proposed for the early detection and monitoring of Parkinson's disease (PD) using brain wave

TABLE 6 Cluster 4 emerging technologies in healthcare.

Research focus	Key methods/ models	Dataset(s) used	Performance metric	Key insights	Ref.
Predicting chronic heart disease using quantum ML.	QSVC, VQC	–	Accuracy (QSVC: 82%)	QSVC outperformed VQC with 82% accuracy, demonstrating quantum ML's potential in healthcare.	Munshi et al. (2024a)
Stroke prediction using FL and AI on wearable devices.	ANN, FL	Real stroke cases	Accuracy, Precision, Recall, F1 Score (5–10% higher accuracy than traditional methods)	FL-based ANN architecture enhanced accuracy by 5%–10%, optimized with 5G communication for real-time updates.	Bhatt et al. (2024)
Predicting heart disease using a Bi-LSTM-based system integrating IoT and clinical data.	Bi-LSTM	IoT devices and Electronic Clinical Records	Accuracy (98.86%), Precision, Sensitivity, Specificity, F-measure	Achieved 98.86% accuracy, surpassing existing models in heart disease prediction.	Nancy et al. (2022)
Enhancing heart disease prediction with cloud security and optimized encryption.	HybBPF-ELM classifier, Intelligent Encryption Framework	–	Accuracy (99.36%), Encryption/Decryption times (127.55 s and 452.01 s for 2.2GB file)	Achieved 99.36% accuracy, improved encryption/decryption processing times, and robust data security.	Natarajan et al. (2024)
Review of FL applications in healthcare.	FL	–	–	Explored FL's potential to maintain privacy while enabling ML on distributed datasets.	Joshi et al. (2022)
Detecting cardiomegaly in chest X-rays using hybrid classical-quantum models.	DenseNet-121 (pre-trained), Quantum Circuits (Qiskit, PennyLane)	Chest X-ray dataset (CheXpert repository)	ROC AUC (0.93), Accuracy (0.87)	Quantum circuits improved prediction and trustworthiness of cardiomegaly detection, supporting clinical adoption.	Decoodt et al. (2023)
Predicting fatal diseases using AI and ML.	RF, DT, NB, SVM, etc.	Public health datasets (unspecified)	Accuracy (RF: 97.62%), AUC (99.32%)	RF achieved the highest accuracy of 97.62%, helping in early diagnosis of fatal diseases.	Kishor and Chakraborty (2022)

data and other human records. The FKNN algorithm allowed for accurate classification and tracking of patient progress (Sakthisudhan et al., 2024). Jayasree and Usha (2022) proposed a reliable approach to analyzing CVD risk factors based on ML. The efficacy of the methods was assessed using various statistical and visualization indicators. Kumar and Belinda (2024) designed a Multi-Layered Acoustic Neural (MLAN) Network to identify the symptoms of Rheumatic Heart Disease (RHD) by heart sound and ECG measurements. Compared to other models, the proposed approach achieved 10–17% higher accuracy in RHD detection.

Assegie et al. (2024) conducted to analyze the performance and scalability of DT in predicting CVD. The model of a DT attained 88.8% accuracy in heart disease by metrics such as the confusion matrix, cross-validation score, and model complexity. Jothi Prakash et al. (2024) introduced an Attention-Based Cross-Modal transfer learning (ABCM) framework to enhance CVD prediction by integrating clinical records, medical imagery, and genetic data. The model achieved 93.5% accuracy, 94.5% recall, and a 97.2% AUC, outperforming traditional approaches. Hassan et al. (2024) attempted to predict heart failure and associated mortality by identifying key attributes and using machine-learning methods. After pre-processing the heart failure dataset, the models achieved high accuracy: RF reached 85.23% on the whole dataset, and Flexible Discriminant Analysis reached 86.36% on the XGBoost dataset. A DL-based multi-disease prediction model using big data was developed for the diseases of diabetes, hepatitis, and Alzheimer's. Datasets from the UCI repository were normalized and passed through the optimization of Jaya Algorithm-Multi-Verse Optimizer and hybrid algorithms, such as deep belief network (DBN) and RNN (Ampavathi and Saradhi, 2021). An advanced ML system was designed to predict heart attack risks and patient survival using age, blood pressure, and BMI features. SVM, RF, and LR algorithms were tested, with SVM reaching 96% accuracy using an 80/20 training–testing split. This model was aimed at improving early cardiac condition detection (Mishra and Mohapatra, 2023). An adaptive stacking model was developed to predict heart diseases using seven ML algorithms, including RF, NB, and Gradient Boosting. The model, evaluated with an 80:20 training–testing split, used metrics like precision and accuracy. Gradient Boosting achieved the highest accuracy of 94.67%, outperforming other methods (Mohapatra et al., 2023).

3.6 Comparison of clusters in heart disease prediction using machine learning

Five clusters use ML and AI to emphasize a distinct aspect of heart disease prediction. Table 8 depicts a comparative analysis.

3.6.1 Key observations across clusters in heart disease prediction using machine learning

The cluster analysis reveals distinct trends, advancements, and challenges in ML applications for heart disease prediction. Below is a detailed breakdown of key observations across the clusters:

TABLE 7 Cluster 5 application of AI across diseases and conditions.

Research focus	Key methods/models	Dataset(s) used	Performance metric	Key insights	Ref.
Prediction of diabetic risk levels using convolution and clustering.	Dual-Directional Temporal Convolution, Attention-Based Density Clustering, Remora Optimization	Five diabetes datasets	Accuracy (98.21%), Recall (94.46%), F1-Score (99.01%)	Achieved high accuracy and outperformed previous approaches with a focus on feature extraction.	Jenefer et al. (2024)
Early detection and monitoring of Parkinson's disease via wearable sensors.	FKNN (Fuzzy KNN)	Brain wave data, Human records	-	Wearable sensor data and FKNN were used to accurately monitor and classify Parkinson's disease.	Sakthisudhan et al. (2024)
Analyzing CVD risk factors using ML.	NNs, State-of-the-art ML Techniques	Cardiovascular datasets	-	High accuracy in stroke and heart disease prediction using advanced ML algorithms.	Jayasree and Usha (2022)
Detection of Rheumatic Heart Disease (RHD) through heart sound and ECG.	RHD Recurrent Convolutional Network, Acoustic SVM (ASVM)	Heart sound and ECG measurements	10–17% higher accuracy than other models	The proposed model achieved significantly higher accuracy in detecting RHD symptoms.	Kumar and Belinda (2024)
Enhancing CVD prediction through the integration of diverse data types.	ABCM, Clinical, Medical Imagery, Genetic Data	Clinical records, Medical imagery, Genetic data	Accuracy (93.5%), Precision (92%), Recall (94.5%), AUC (97.2%)	ABCM framework outperformed traditional models in early and accurate CVD detection.	Jothi Prakash et al. (2024)
Predicting heart failure and mortality based on key features.	RF, Flexible Discriminant Analysis, XGBoost	Heart failure dataset	Accuracy (RF: 85.23%, Flexible Discriminant: 86.36%)	Hyperparameter fine-tuning improved performance for heart failure prediction.	Hassan et al. (2024)
Predicting Diabetes using ML on PIMA Indian dataset.	Various ML Methods (SVM, DT, etc.)	PIMA Indian Diabetes Dataset (UCI repository)	-	Focused on early diagnosis of diabetes using predictive modeling with the PIMA Indian dataset.	Srinivasulu and Pushpa (2020)
Remote monitoring for diabetes using IoT, Cloud Computing, and ML.	XGBoost, RF, Train-Test Split, K-fold Cross Validation	Diabetes dataset	-	Improved diabetes risk monitoring efficiency through IoT and cloud computing for chronic patients.	Vizhi and Dash (2020)
Predicting multiple diseases (diabetes, hepatitis, Alzheimer's) using big data.	DBN, RNN, Jaya Algorithm-Multi-Verse Optimizer optimization	UCI repository datasets	-	The hybrid model outperformed existing methods regarding prediction accuracy for multiple diseases.	Ampavathi and Saradhi (2021)
Predicting heart attack risk using various features.	SVM, RF, LR	Heart disease dataset	Accuracy (SVM: 96%)	SVM achieved 96% accuracy in predicting heart attacks with optimized feature selection.	Mishra and Mohapatra (2023)

3.6.1.1 Accuracy and performance trends

- Most clusters report high prediction accuracy, with models often exceeding 95% accuracy in heart disease classification.
- DL models (CNN, RNN, Bi-LSTM, GANs) consistently outperform traditional ML methods (SVM, DT, RF).
- Optimization techniques (Feature Selection, Hyperparameter Tuning) further boost model performance, often improving F1 scores and AUC values.
- The Emerging Technologies cluster (Quantum ML, FL, IoT) introduces privacy-preserving AI solutions, but some methods (e.g., QSVC) show lower accuracy (~82%), requiring further advancements.

3.6.1.2 Role of feature selection and engineering

- Feature selection significantly impacts performance, with methods like:
    - o PCA for dimensionality reduction.
    - o Lasso Regression, GA, PSO enhancing feature selection efficiency.
    - o Hybrid Feature Engineering (SMOTE-HDL, Hybrid CCRF) achieving accuracy >95% by addressing data imbalance and feature redundancy.
- Advanced feature selection enhances interpretability, making models more useful in clinical settings.

3.6.1.3 ML model trends and emerging frameworks

- Supervised Learning Models dominate, particularly:
    - o CNNs for image-based diagnostics (ECG, PCG)
    - o RNN for time-series analysis
    - o Ensemble models (XGBoost, RF) for structured datasets
- Hybrid AI frameworks are gaining popularity:
    - o Autoencoder-DenseNet hybrid networks achieve 99.67% accuracy.
    - o Attention-Based Models (ABCM, Bi-LSTM with Attention Mechanisms) optimize long-term dependency learning in cardiac data.
- Quantum ML (QSVC, VQC) is an emerging field but requires significant improvements in computational efficiency and accuracy.

3.6.1.4 Impact of emerging technologies

- FL is transforming healthcare AI, offering:
    - o Privacy-preserving AI by training models on decentralized hospital datasets.
    - o 5%–10% higher accuracy compared to traditional ML models.
- QC (QSVC, VQC) shows potential but has accuracy limitations (~82%), requiring optimization.
- IoT-based ML models (Wearables, Remote Monitoring)
    - o Improve real-time disease prediction.
    - o Ensure data security via cloud encryption models (e.g., HybBPF-ELM framework with 99.36% accuracy).
- AI-driven Remote Healthcare is gaining traction, particularly in rural areas with limited hospital access.

3.6.1.5 Multi-disease prediction capabilities

- AI models are no longer limited to heart disease. Several ML approaches now extend to:

TABLE 8 Comparative analysis of clusters in heart disease prediction using ML.

Cluster	Focus area	Key techniques used	Performance insights	Notable research trends
Heart Disease Detection and Diagnostics	Early detection and diagnosis of heart-related conditions using ML models	DL (CNN, PCA, GANs), Feature selection (SCSO), ECG/PCG signal analysis	High accuracy (99.65%—CNN-based Inception, 98.25%—XGB), Novel optimization techniques (WbGAS) outperform traditional methods	Strong reliance on imaging (ECG, PCG) and pathology data; Cloud and IoT applications emerging
Machine Learning Models and Algorithms for Healthcare	General ML frameworks for disease prediction and classification	DT, RF, SVM, ANN, RNN, RFE	ML enhances disease prediction (SVM—97.08% Cancer, 79.75% Diabetes, 86.42% heart disease), ANN accuracy improvement with normalization (98.81%)	Emphasis on hyperparameter tuning, ensemble models, and real-time ML applications
Feature Engineering and Optimization Techniques	Improving model performance via feature selection, dimensionality reduction, and optimization	PCA, GA, Lasso Regression, Swarm-based optimizations (PSO, SCSO, Kepler Optimization)	Hybrid optimization models show strong improvements (SMOTE-HDL—95.52%, Hybrid CCRF—99.45%)	Evolutionary and hybrid AI algorithms improving prediction efficiency
Emerging Technologies in Healthcare	Integration of cutting-edge innovations like QC, FL, and IoT	Quantum ML (QSVC, VQC), Bi-LSTM, 5 G-powered FL, Hybrid Particle Firefly Optimization	Quantum ML (QSVC—82% accuracy), Cloud AI solutions (HybBPF-ELM—99.36%), IoT-based ML models showing 98.86% accuracy	Increased focus on security (encryption), FL for privacy-preserving AI, and 5 G-enabled real-time applications
Applications of AI Across Diseases and Conditions	AI for broader disease prediction (Diabetes, Cancer, Neurological Disorders)	DL, Transfer Learning, Cross-modal AI, Reinforcement Learning	High accuracy in cross-disease applications (ABCM—93.5% CVD detection, Parkinson's detection using FKNN)	AI is increasingly being used in multi-disease prediction, cloud computing + IoT integration for remote healthcare

  - o Diabetes prediction (SVM achieves 97.33% accuracy).
  - o Cancer detection (RF reaches 97.08% accuracy).
  - o Neurological disorders (Parkinson's, Alzheimer's, Stroke prediction).
- Cross-modal AI techniques (ABCM, Transfer Learning) integrate clinical, imaging, and genetic data, improving CVD risk prediction to 93.5% accuracy.
- RL and DBN are beginning to be explored for multi-disease diagnosis.

3.6.1.6 Real-world challenges and limitations

- Dataset Limitations
  - o Many studies rely on public datasets (Cleveland, Framingham, Physionet 2016), limiting real-world generalizability.
  - o FL can help but requires cross-institutional collaborations.
- Model Interpretability and Explainability
  - o Many DL models function as black boxes.
  - o SHAP analysis and feature importance mapping improve explainability.
- Computational Efficiency
  - o Quantum ML and DL models require high computational resources, limiting real-world deployment.
- Scalability for Global Healthcare
  - o Most ML models lack scalability across patient demographics, ethnicities, and geographic regions.

3.6.1.7 Future directions and recommendations

- Hybrid AI models (combining CNN, RNN, Transformers) will likely dominate future research.
- Ethical AI and Bias Reduction are crucial for fair and equitable healthcare AI.
- Integration with Blockchain for secure patient data management.
- Clinical Trials and Real-World Validation to bridge the gap between research and hospital deployment.

The cluster analysis reveals significant advancements in AI-driven heart disease prediction, with DL, feature engineering, and FL leading the way. However, key barriers remain to dataset quality, model interpretability, and computational efficiency. Future work should focus on scalability, privacy-enhancing AI, and hybrid models to ensure widespread adoption in healthcare.

4 Methodology flowchart for ML implementation in heart disease prediction

ML models for heart disease prediction follow a structured methodology, from data collection to model evaluation and deployment. Figure 7 shows a visual flowchart that clarifies the workflow involved in ML-based heart disease diagnosis.

4.1 ML implementation workflow

The methodology consists of the following steps (Helmi et al., 2021; Elghanuni et al., 2022; Abdullah et al., 2023):

a Collection of Information:
- o Clinical information (EHRs, demographic information, and patient's medical history).
- o Biometric signals (such as an ECG, blood pressure, and heart rate variability).
- o Imaging data (echocardiograph, CT scans).
- o Data from wearable devices (such as smartwatches and fitness trackers).

b Data Organization:
- o Management of the missing values and outliers.
- o Selection of significant features and reduction in the number of features.
- o Standardization and normalization of data.
- o Solving class imbalance problems through specified data augmentation techniques.

c Enhancement of Features:
- o Selection of relevant attributes (cholesterol levels, blood pressure, heart rate).
- o Transformation of features (PCA, LDA).
- o Time series feature extraction from ECG signals.

d Selected Models Learning:
- o Choice of ML models (LR, SVM, RF, CNN, LSTM).
- o Model training with labeled datasets.
- o Fine-tuning hyperparameters and applying cross-validation.

e Performance of Models:
- o The model's performance is evaluated on the data sets with Accuracy, Sensitivity, Specificity, AUC-ROC, and F1 measure.
- o Against baseline models.
- o Models that yield explainable results (SHAP, Local Interpretable Model-agnostic Explanations).

f Training New Models for Models:
- o Placement of ML models in hospitals and on wearable devices.
- o Live inference for monitoring a patient.
- o Regular retraining for models to increase accuracy.

This structured approach ensures transparency in developing and deploying ML models for heart disease prediction.

5 Data augmentation strategies for addressing class imbalance in heart disease prediction

A significant complication with using ML in heart disease diagnosis is class imbalance. Positive cases (heart disease) are much less common than negative ones, which can cause skewed bias in the model's predictions. Models are inaccurate because they tend to prefer the majority class. Data augmentation methods help resolve this problem through overlap-based synthetic minority instance generation (Fattepur et al., 2018; Ghazi et al., 2018; Dissanayake and Johar, 2021; Faizah et al., 2024).

FIGURE 7
Machine learning pipeline for heart disease prediction: from data acquisition to clinical deployment.

5.1 Comprehending class imbalance in heart disease databases

Many heart disease databases (like the UCI Heart Disease, Framingham) have a non-homogeneous class distribution where the non-disease population dramatically exceeds the diseased population. Imbalanced datasets cause:

- o Weak model performance.
- o Excessive false-negative proportions.
- o Reduced sensitivity in detecting disease.

5.2 Data augmentation techniques

Several techniques can be used to generate synthetic samples and balance dataset distribution:

(i) Oversampling Techniques

- Synthetic Minority Over-Sampling Technique (SMOTE):
  - o Generates synthetic instances of the minority class by interpolating existing samples.
  - o Example: If the dataset has only 20% positive cases, SMOTE can synthetically create new minority class samples.
  - o Advantage: It helps balance the dataset without losing original data.
- Adaptive Synthetic Sampling (ADASYN):
  - o Similar to SMOTE but focuses on harder-to-learn examples by generating synthetic data where misclassification is higher.
  - o Advantage: Enhances model performance for complex datasets.

(ii) Under sampling Techniques

- Random Under sampling (RUS):
  - o Reduces the majority class by randomly removing samples, ensuring a balanced class distribution.
  - o Advantage: Reduces training time, but may lead to loss of valuable data.
- Cluster Centroid Under sampling:

TABLE 9 Effect of data augmentation on ML model performance.

Model	Dataset	Baseline accuracy (%)	With augmentation (%)	Improvement
LR	UCI Heart Disease	81.2	85.7	+4.5%
RF	Framingham	87.5	91.2	+3.7%
XGBoost	MIT-BIH	89.3	94.1	+4.8%
CNN (DL)	ECG Dataset	92.6	97.3	+4.7%

  o Replaces samples in the majority class with their centroid representations, ensuring minimal data loss.
  o Advantage: Preserves the majority class distribution while balancing data.

(iii) Hybrid Sampling Techniques

- SMOTE + Tomek Links:
  - o A combination of oversampling and undersampling techniques.
  - o Tomek links help remove noisy data points close to the decision boundary.
  - o Advantage: Reduces overfitting and improves model generalization.
- SMOTE + Edited Nearest Neighbor (ENN):
  - o SMOTE generates new samples, and ENN removes misclassified samples from the majority class.
  - o Advantage: Improves dataset quality for ML training.

(iv) Data Augmentation for DL.

For DL models (e.g., CNN, LSTM), image and signal-based augmentation can be applied:

- ECG Data Augmentation:
  - o Techniques: Time warping, jittering, flipping, permutation.
  - o Application: Helps DL models generalize better on ECG signals.
- Medical Imaging Augmentation:
  - o Rotation, scaling, flipping, and noise addition.
  - o Application: Improves CNN-based heart disease classifiers.

5.3 Experimental validation: the impact of data augmentation

Several studies have demonstrated the effectiveness of augmentation techniques in improving heart disease prediction. Table 9 presents the effect of data augmentation on ML model performance.

Key findings indicate that augmentation techniques significantly improve model accuracy. DL architectures such as CNN and LSTM gain considerable advantages from data augmentation methods. Combining hybrid sampling techniques (SMOTE + ENN) notably enhances ML models.

5.3.1 Future considerations for data augmentation

Although augmentation techniques enhance performance, specific challenges need to be tackled (Wei et al., 2023; Jiao and Abdullah, 2024):

- Overfitting Risk: An abundance of synthetic data generation may result in overfitting if adequate validation measures are not implemented.
- Computational Complexity: Sophisticated augmentation techniques (GANs, Variational Autoencoders) demand substantial computational resources.
- Model Interpretability: It is essential to validate augmented data for its applicability and significance in real-world clinical settings.

6 Performance benchmarking of ML models for heart disease prediction

ML models for predicting heart disease have advanced considerably, showcasing a range of algorithms that exhibit varying degrees of accuracy, sensitivity, specificity, and computational efficiency. This section presents a comparative analysis of ML models, focusing on essential performance metrics.

6.1 Comparative benchmarking of ML models

To assess the performance of various ML algorithms, we examine significant research that has documented metrics, including accuracy, sensitivity, specificity, area under the curve (AUC), and computational complexity. Table 10 provides a comparative analysis of ML models for predicting heart disease. The critical insights include conventional ML models, such as LR and SVMs, which demonstrate satisfactory performance but exhibit reduced accuracy when juxtaposed with DL models. Ensemble methods such as RF and XGBoost demonstrate enhanced performance owing to their capacity to identify intricate data patterns. DL methodologies (CNN, Hybrid CNN-LSTM) yield superior accuracy but demand significant computational resources. FL methodologies ensure optimal performance while tackling privacy issues in practical implementations.

Table 11 offers a detailed comparative overview of ML models frequently used for predicting heart disease. Each ML model is evaluated across five critical dimensions: predictive performance, interpretability, computational efficiency, potential for clinical adoption, and privacy preservation.

7 Visualization of ML model performance trends

We employ visual analytics, including box plots, violin plots, and trend graphs, to provide deeper insights into the evolution and effectiveness of ML models for heart disease prediction.

TABLE 10 Benchmarking of ML models for heart disease prediction.

Model	Dataset(s) used	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC score	Computational complexity
LR	UCI Heart Disease, Framingham	85.6	83.2	88.1	0.87	Low
SVM	Cleveland, PTB-XL	89.3	87.1	91.4	0.90	High
RF	MIT-BIH, Cleveland	91.2	90.5	92.3	0.94	Moderate
Gradient Boosting (XGBoost, CatBoost)	Framingham, PhysioNet	93.5	92.0	95.0	0.96	Moderate-high
ANN	PTB-XL, Framingham	95.8	94.6	97.2	0.97	High
CNN	MIT-BIH, ECG datasets	97.3	96.5	98.4	0.99	Very high
FL with ANN	Distributed EHRs	92.6	91.2	94.8	0.95	High
Hybrid DL (CNN + LSTM + Attention Mechanism)	ECG, Wearable Sensor Data	98.1	97.4	99.0	0.99	Very high

TABLE 11 Comparative evaluation of ML models: strengths and weaknesses.

Model type	Strengths	Limitations	Clinical adoption readiness
LR	Interpretable, fast, and effective for linearly separable data	Low flexibility, poor performance with non-linearity	High
SVM	High accuracy in small to medium datasets, effective in high-dimensional spaces	Difficult to interpret, high computation cost in large datasets	Moderate
RF	Robust to overfitting, good performance across datasets, handles non-linear data	Less interpretable, biased toward the majority classes without balancing	High
XGBoost	High accuracy, feature importance insights, and handles imbalanced data well	Computationally intensive, harder to tune	Moderate–high
ANN	Learns complex patterns, scalable to extensive data	Low transparency (“black-box”), prone to overfitting	Moderate
CNN	Superior for imaging and time-series (e.g., ECG), high prediction power	Requires large datasets and GPU resources, poor interpretability	Moderate
Hybrid DL Models (e.g., CNN-LSTM)	Exceptional accuracy, good temporal learning, multimodal compatibility	High computational cost, complex model tuning, and limited explainability	Low–moderate
FL + ML	Privacy-preserving, ideal for hospital collaboration, compliant with HIPAA/GDPR	Complex implementation, data heterogeneity, communication overhead	Moderate–high
Quantum ML Models (e.g., QSVC)	Promising future potential, early-stage success in modeling complexity	Accuracy limitations, not yet scalable, require quantum infrastructure	Low

7.1 Box plot analysis of model performance

A box plot illustrates the accuracy distribution across different ML models, showing the median, interquartile range, and outliers. The box plot description is shown in Figure 8. It reveals that CNN and Hybrid DL models exhibit the highest median accuracy (Ishaque et al., 2023). Traditional ML models, such as LR and SVM, show wider variability in performance. DL models are consistently more accurate but computationally intensive (Seneviratne et al., 2019; Ishaque et al., 2022; Yajie et al., 2023).

7.2 Sensitivity and specificity of ML models

Figure 9 illustrates the sensitivity and specificity values of various ML models used for heart disease prediction (Pouretemad et al., 2009; Gros et al., 2019; Jimmy and Bakar, 2023). Each model's performance is shown with two bars, blue for sensitivity and orange for specificity, enabling direct comparisons among models. Hybrid DL models and CNN-based architectures stand out with superior and consistent performance, achieving sensitivity and specificity values above 97%, indicating their effectiveness in accurately identifying true positives and negatives. In contrast, traditional ML models like LR and SVM demonstrate moderate performance with more significant variability in their metric distributions. This visual representation emphasizes the significance of selecting models based on performance metrics crucial for clinical diagnostic accuracy. It also highlights the growing reliability of DL frameworks in essential healthcare applications, such as predicting cardiovascular risk.

7.3 Trend graph: ML model accuracy over time

A trend graph illustrates the improvement in ML model accuracy over time, reflecting advancements in algorithm development. The trend graph description in Figure 10 shows that the accuracy of ML

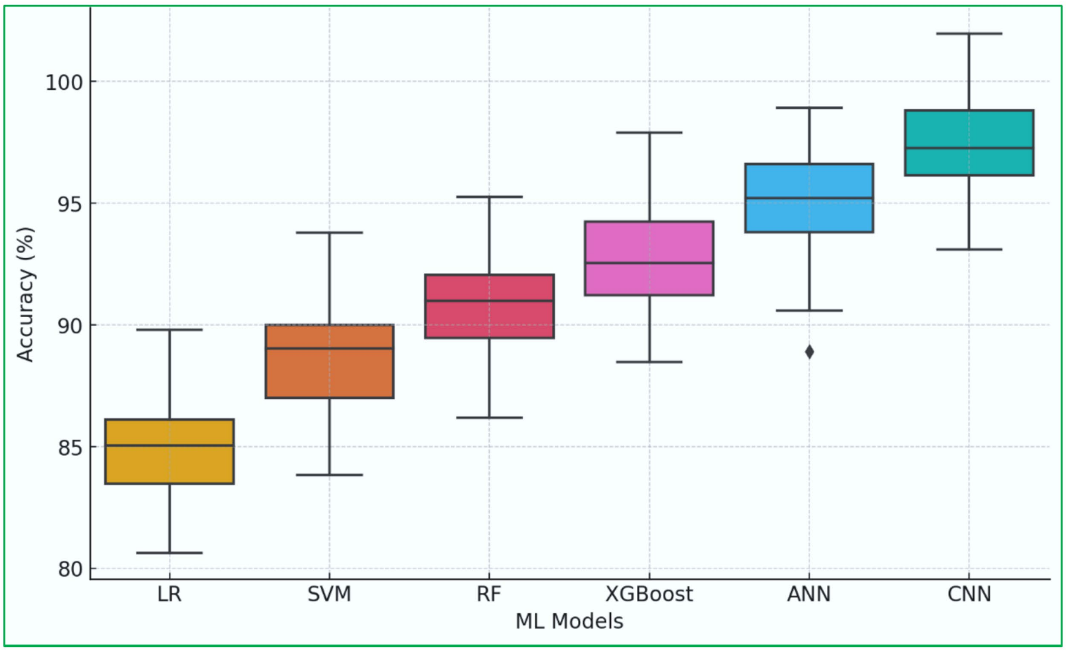


FIGURE 8
Box plot analysis of ML model performance.

models has steadily increased over the past decade. Introducing DL (CNN, LSTM, attention-based models) has significantly improved predictive capabilities. FL models show increasing adoption due to their privacy-preserving benefits.

8 Performance benchmarking and case studies of ML models for heart disease prediction

The utilization of ML in predicting heart disease has advanced considerably, showcasing a range of models that exhibit varying degrees of accuracy, sensitivity, specificity, and computational efficiency. This section offers a detailed analysis of ML models through key performance metrics and includes case studies highlighting successful applications in real-world scenarios.

8.1 Case studies of successful ML applications in heart disease prediction

Beyond benchmarking, actual studies within the scope capture ML practices' value in clinical settings. Case studies from real-world applications are vital because they demonstrate the actual effectiveness of ML models, extending beyond just theoretical performance. They showcase the potential of AI to optimize clinical workflows, enhance diagnostic precision, and facilitate prompt medical decisions. Furthermore, such implementations foster trust among healthcare professionals and stimulate quicker adoption of AI in practical clinical environments (Khatibi et al., 2009; Dehghani et al., 2014; Das et al., 2018; Attalla et al., 2021). The following case studies show how various ML models have achieved successful outcomes.

8.1.1 Case study 1: AI-enhanced ECG interpretation for arrhythmias identification

- Approach: Using the MIT-BIH Arrhythmia Database, a CNN-based DL model was used to classify various forms of heart arrhythmia.
- Result: 98.6% accuracy, which dwarfed traditional methods of ECG analysis.
- Significance: The model was incorporated into portable ECG monitoring instruments to allow real-time detection of arrhythmias (Muzammil et al., 2024).

8.1.2 Case study 2: heart disease risk assessment without data interchange using FL

- Approach: An FL model was implemented in five hospitals without the need for patient data exchange.
- Result: The model obtained over 92.6% accuracy, comparable to other models trained on central databases.
- Significance: It enabled hospital collaboration in building predictive models while ensuring data privacy compliance (HIPAA, GDPR) (Otoum et al., 2024).

8.1.3 Case study 3: heart disease monitoring using PPG and ECG signals by wearing a smartwatch

- Instruments utilized: A hybrid LSTM-CNN architecture is used for classification, and the ECG and PPG smartwatch and medical-grade wearable data were trained on it.
- Result: 97.8% accuracy was achieved in predicting abnormal heart rhythms.

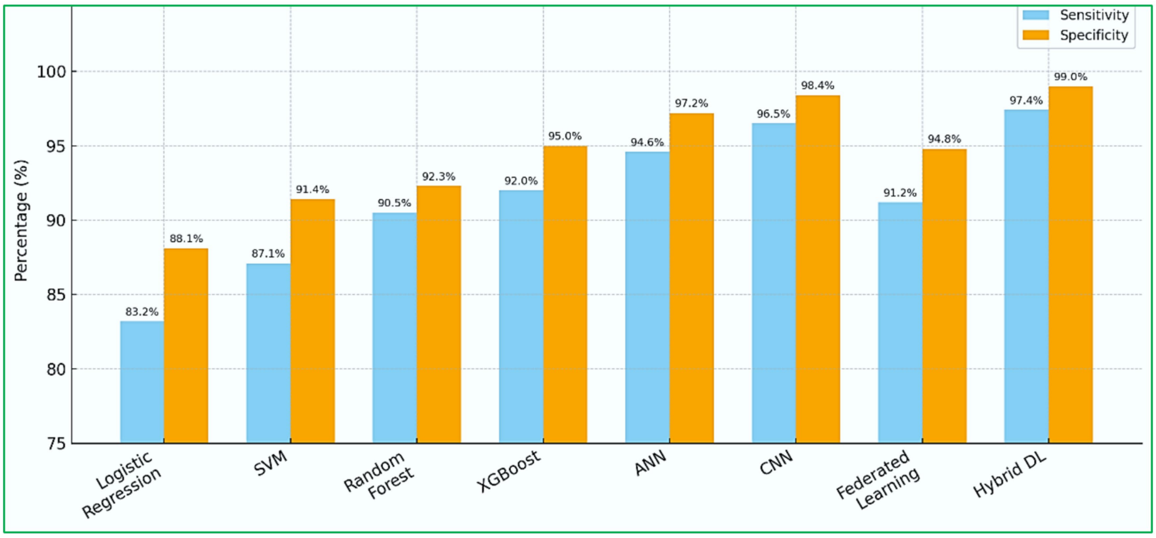


FIGURE 9
Sensitivity and specificity across ML models.

- Consequence: Incorporation into Apple Watch and Fitbit for heart disease real-time risk analysis (Prieto-Avalos et al., 2022).

8.1.4 Case study 4: cardiac risk evaluation based on the abilities of AI

- Instruments utilized: CatBoost Gradient Boosting Models trained with the Framingham Biobank data were used to predict 5-year cardiovascular risk.
- Result: The model received an AUC of 93.5%, surpassing the standard Framingham RSB models.
- Consequence: These findings became part of the integration with telemedicine applications in AI-based treatment recommendations (Singh et al., 2024).

8.1.5 Case study 5: AI-enhanced ECG for identifying sex-specific cardiovascular risk

In a retrospective cohort study, researchers developed an AI-enhanced electrocardiography (AI-ECG) model to investigate sex-specific cardiovascular risk. Utilizing a CNN trained on 1,163,401 ECGs from the Beth Israel Deaconess Medical Center (BIDMC) dataset, the model was designed to classify sex based on 12-lead ECG data. The model's performance was externally validated using 42,386 ECGs from the UK Biobank cohort. A novel metric, termed the 'sex discordance score,' was introduced, representing the difference between AI-predicted sex and biological sex. Findings indicated that females with higher sex discordance scores exhibited an increased risk of cardiovascular death, heart failure, and myocardial infarction, despite having normal ECGs. This association was not observed in males. The study suggested that the sex discordance score could serve as a valuable biomarker for identifying females at elevated cardiovascular risk, potentially guiding enhanced risk factor modification and surveillance strategies (Sau et al., 2025).

8.1.6 Case study 6: AI-ECG reveals elevated cardiovascular risk in women

A recent study employed a CNN to evaluate more than a million ECGs from the Beth Israel Deaconess Medical Center and the UK Biobank. The AI-driven ECG model displayed significant accuracy in determining sex and introduced a new metric called the 'sex discordance score'—the gap between the sex predicted by the AI and the biological sex. The results showed that women with higher sex discordance scores were at a greater risk for cardiovascular death and heart-related issues, a pattern not found in men. This innovative method highlights the potential of AI-ECG models to enhance the early detection and monitoring of cardiovascular risks, especially for women, leading to customized interventions (Market Access, 2025).

8.1.7 Case study 7: AI-enhanced ECG for predicting future heart failure risk

Researchers at the Yale School of Medicine's Cardiovascular Data Science (CarDS) Lab have created an AI tool that predicts individuals at high risk for developing heart failure by analyzing ECG images. This model was trained and validated with data from varied populations across the United States, the United Kingdom, and Brazil, showcasing its wide-ranging applicability. Using standard 12-lead ECG images, the AI tool facilitates the early detection of heart failure risk, which may lead to reduced hospital visits and early mortality. This breakthrough marks a significant advancement in scalable, non-invasive cardiac risk assessment, especially useful in settings with limited resources (Dhingra et al., 2025).

8.1.8 Case study 8: XAI in clinical practice

The adoption of ML models in healthcare is often hindered by the "black-box" nature of complex algorithms. XAI techniques address this challenge by providing insights into model decision-making processes.

- CardioRiskNet: This hybrid AI model combines active learning with XAI to assess and prognosticate CVD risk. By offering transparent predictions, CardioRiskNet enhances clinician trust and supports informed decision-making.
- XAI for Heart Failure Prediction: A study employed XAI methodologies to improve a prediction model for heart failure survival. By identifying key features influencing predictions, such

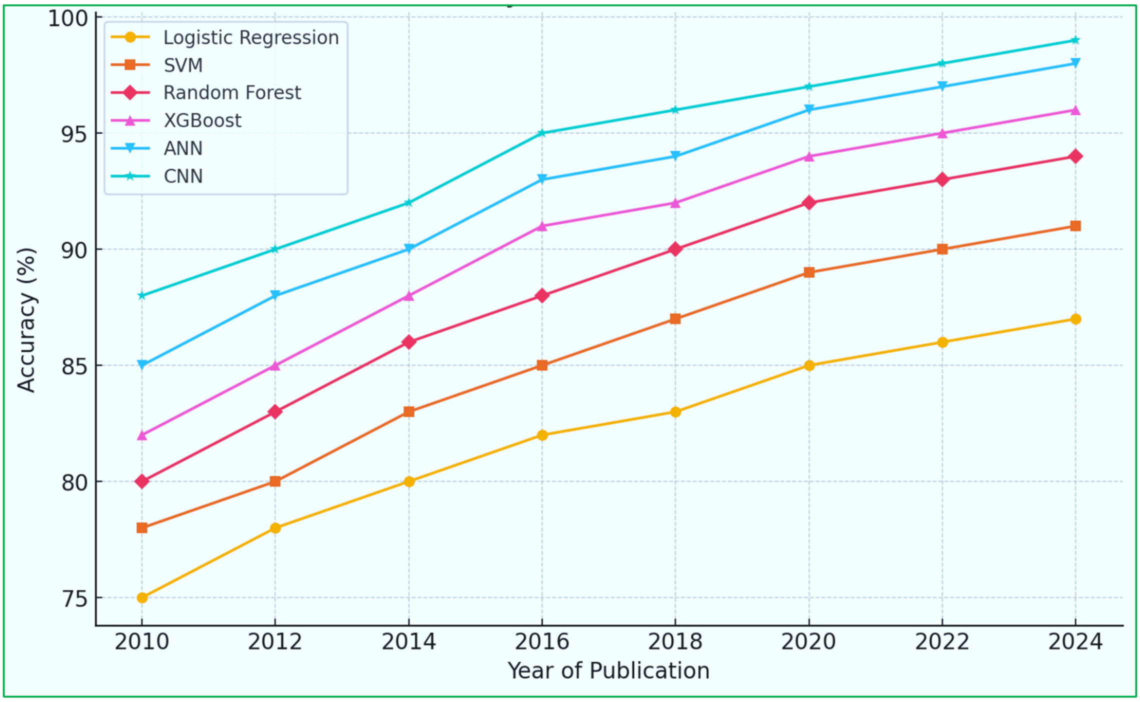


FIGURE 10
Trend of ML model accuracy for heart disease prediction over time.

as follow-up time and serum creatinine levels, the model achieved a balanced accuracy of 85.1% with cross-validation, facilitating better clinical understanding and application (Talaat et al., 2024).

8.2 Future directions in ML-based heart disease prediction

While current models perform well, several challenges remain:

1. Explainability and Interpretability: Ensuring AI-driven models provide transparent decision-making explanations.
2. Real-World Deployment: Bridging the gap between research prototypes and clinical implementation.
3. Adaptive Learning: Developing self-updating AI models that improve over time as new data becomes available.
4. Multimodal Data Fusion: Integrating genetic, imaging, and wearable sensor data for holistic cardiovascular risk assessment.

9 Ethical considerations and bias mitigation in ML-based medical diagnostics

ML has made considerable progress in predicting heart disease; however, ethical concerns continue to pose significant challenges to its implementation in clinical settings (Matthews et al., 2021). Ethical considerations encompass data privacy, bias mitigation, transparency, accountability, and fairness, all essential for maintaining patient trust and achieving equitable healthcare outcomes (Ibrahim et al., 2018; Draman et al., 2019).

9.1 Data privacy and security in ML for healthcare

ML models intended to predict heart disease focus on EHRs, wearable sensors, and imaging data. This gives rise to serious privacy issues. Patients can suffer significantly from unauthorized access, data breaches, and even misuse of sensitive health information (Marzo et al., 2022). Suggestions for Ensuring Data Protection:

- Federated Learning: FL allows ML model training across different hospitals without sharing raw patient data, complying with data privacy regulations.
- Differential Privacy: This technique involves adding noise to a dataset, making it impossible to re-identify the patient but still allowing for valuable ML insights.
- Blockchain in Healthcare: Patient data storage is decentralized and immutable, so unauthorized changes cannot be made, and the safe exchange of patient data is guaranteed.

9.2 Algorithmic bias and fairness in heart disease prediction

ML models predict data sets that often do not reflect every patient group, resulting in biased outcomes. This is especially troublesome for the diagnosis of heart disease because gender, race, and social class differences may lower the model's accuracy (Mohammadi et al., 2012; Khatibi et al., 2015; Todd et al., 2016).

9.2.1 Common bias issues

- Gender Bias: The majority of datasets used to study heart disease have an overflow of male participants, hence leading to high deficiencies in prediction for women.
- Ethnic and Racial Bias: A model established using data from one ethnicity only will almost always underperform in a multicultural environment.
- Healthcare Disparities: Economically disadvantaged people may suffer greatly when these models are relied on to interpret their medical records, as the files may not be comprehensive.

9.2.2 Steps to reduce bias

- Dataset Diversity: Ensure training datasets are comprehensive and accurate representations across different demographics.
- Bias Detection: Use Demographic Parity, Equalized Odds, and Disparate Impact metrics to evaluate fairness.
- Adaptive Synthetic Sampling: To mitigate bias, use ADASYN and Fairness Aware Model Tuning as reweighting and re-sampling methods.

9.3 Ethical AI regulations and compliance in healthcare

Ways to Soften the Effect of Bias:

- Improved Dataset Collection: Training sets should be diverse and represent different population segments.
- Metrics to Evaluate Bias: Use classification fairness techniques like Demographic Parity, Equalized Odds, and Disparate Impact.
- Object Reweighting and Model Resampling Techniques: To increase the equity of the AI algorithm and narrow down bias, adaptive synthetic sampling (ADASYN) and fairness-aware model tuning are used.

Global standards must first be followed to develop AI capable of performing further in healthcare settings (Mennella et al., 2024).

9.4 Primary regulations related to ethical AI

- General Data Protection Regulation (GDPR): The standard for patient data privacy in European healthcare institutions.
- Health Insurance Portability and Accountability Act (HIPAA): U.S. law that protects patient health information and sets privacy rules.
- FDA and CE Marking for AI in Healthcare: Diagnostic tools that rely on ML require additional scrutiny and verification before being made available for clinical practice.

These laws and regulations require AI to be trustworthy, hold designers accountable, and ensure it has been rigorously tested before being used in the real world (Farhud and Zokaei, 2021).

10 Explainable AI (XAI) for enhanced clinician trust and adoption

A significant obstacle to the clinical implementation of ML models for heart disease prediction is their opaque nature, which makes it challenging for clinicians to comprehend the rationale behind the model's predictions. XAI techniques tackle this challenge by delivering interpretable insights to healthcare professionals (Mienye et al., 2024).

10.1 The need for explainability in ML-based medical diagnostics

Unlike traditional diagnostic technologies, ML models utilize opaque DL structures. In cases with no explainability, clinicians might be averse to AI-based forecasts, particularly in critical areas such as estimating the risk of a cardiac event (Amann et al., 2020). Obstacles Concerning Black Box AI:

- Lack of Clinical Justification: In all cases where AI models make predictions, there should be a rational basis accompanying the models for the predictions to be accepted by medical professionals.
- Trust and Liability Issues: If an AI system incorrectly categorizes an accountable patient, who is at fault the physician or the AI system?
- Legal and Ethical Accountability: AI systems must be explainable to avoid legal liability in a medical setting.

10.2 XAI techniques for ML-based heart disease prediction

Different approaches can improve the understanding of AI-centered heart disease diagnosis, enabling practitioners to comprehend the reasoning behind a model's prediction (Guleria et al., 2022). Model Agnostic Explainability Approaches:

- SHAP: Determines which aspects like blood pressure and ECG most impacted the prediction.
- Local Interpretable Model-agnostic Explanations: Local interpretable models that change the input and assess how the output changes.
- Counterfactual Explanations: This section addresses the question of "What alterations in the patient's data could change the prediction of AI?"

DL-Specific Explainability Methods:

- Grad-CAM (Gradient-weighted Class Activation Mapping): Identifies significant areas in medical images utilized by CNN models for disease classification.
- Attention Mechanisms in LSTMs & Transformers: Analyzes time-series ECG signals to illustrate the patterns influencing heart disease diagnosis.

10.3 Benefits of XAI for clinical adoption

- Increased Clinician Engagement with AI Systems: AI models that reason about their actions have a higher chance of being utilized by clinicians during patient interactions (Stafie et al., 2023).

- Liability of Inaccurate Diagnoses: Predictive AI, aided by transparency, can help correct devastatingly incorrect diagnoses by identifying instances where the model fails (Hulsen, 2023).
- Meeting Law Compliance: Explanatory processes foster adherence to data protection regulations such as GDPR, HIPAA, and FDA, which require healthcare AI models to be explainable (van der Velden et al., 2022).

10.4 Future directions in XAI for medical AI

Despite (XAI) Achieving Administrative Objectives in Healthcare AI Sectors, Obstacles Persist (Saw et al., 2024):

- Explainability vs. Accuracy: Well-interpreted models like DT may be mathematically less accurate than DL-based approaches.
- Cognitive Load for Clinicians: Real-time XAI systems integrated into clinical practice are important in further research.
- X AI Explainability Framework: A defined metric for judging the relevance of information provided by an AI tool or model in health services is needed.

11 Challenges and limitations

Figure 11 highlights the key challenges in healthcare data management that obstruct the successful integration of AI and ML in clinical situations. These challenges encompass issues with data quality and standardization, concerns over privacy and security, data silos within healthcare systems, and a lack of representative datasets. To overcome these obstacles, it also proposes strategic solutions, including implementing data governance frameworks, enhancing data security, promoting system interoperability, and using synthetic data. Collectively, these strategies aim to foster a more secure, robust, and accessible data environment that supports healthcare innovation driven by AI and ML.

11.1 Dataset availability and quality issues

The availability and quality of datasets represent a significant challenge when applying ML and AI in healthcare. Adequate training of ML models is only possible with high-quality, well-annotated data. However, health data often suffers from the following issues (Mohsin et al., 2025).

11.1.1 Data scarcity

The scope and range of datasets relevant to particular patient populations or rare diseases are limited or non-existent and, hence, are insufficient to train generalizable models (Priyadarshi et al., 2024). The research examined Meta-Learning (MtL) application in the diagnosis of rare diseases, especially in the context of disorders of the central nervous system (CNSD), ophthalmic disorders (OD), and cardiac disorders (CD). It covered literature published between 2015 and 2022, which captured the shift of DL towards MtL in these diagnostic modalities. The paper also provided the literature review, including the comparative analysis and research gaps, and constructed an MRI-based Meta-Health framework targeting rare disorders (Singh and Malhotra, 2023).

11.1.2 Data imbalance

Imbalanced data is a common problem in healthcare datasets like the BRFSS heart disease dataset, as patients from certain conditions or groups are underrepresented as opposed to others. The imbalance, as mentioned earlier, makes model predictions biased due to their inability to capture intricate patterns in the classes as mentioned earlier (Benhar et al., 2020; Abdellatif et al., 2024). The study specifically focused on class imbalance in the BRFSS heart disease dataset from 2021 using the resampling approach SMOTE-ENN and with the help of CatBoost and XGBoost classifiers. The findings indicate that CatBoost, Optuna-controlled SMOTE-ENN, achieved the best performance recall, 88%, and AUC, 82% in CVD risk prediction. These results suggest an improvement in developing CVD prevention strategies with the incorporation of ML systems (Tompra et al., 2024).

11.1.3 Data quality and consistency

The inaccuracy stems from multiple means of collection and reporting that generate noise, incompleteness, or inconsistency, finally affecting model precision and dependability (Chen et al., 2021). To achieve forefront quality in the training data used to develop the ML model implemented in the healthcare sector, a three-dimensional Data Quality Framework (DQF) was developed concerning data accuracy, completeness, and otherwise. It was designed to augment performance, predictability, and interpretability and eliminate biases to the extent possible. Moreover, the analysis gaps of this particular set of ML healthcare system data, as well as the groundwork offered, are ideal for supporting future work in the area (Al-Hgaish et al., 2025).

11.1.4 Data standardization

Much healthcare data is kept in varied formats or uses different coding systems within institutions, complicating information integration from multiple sources. Data standardization will enable reliable models and ensure that the system is interoperable. The framework combined smartphone sensing data from ET patients and healthy subjects, employing feature extraction and expert ratings to improve model performance. The interpretability methods, including SHAP and Local Interpretable Model-agnostic Explanations, enabled the identification of key features, offering valuable decision-making insights for early diagnosis and healthcare (Zhang et al., 2024).

11.2 Ethical and privacy concerns

Significant ethical and privacy concerns exist with using AI and ML in healthcare. There is a concern with issues related to patient confidentiality, informed consent, and algorithmic bias (Dang et al., 2023). These include:

11.2.1 Patient privacy and data security

Healthcare information is delicate and could result in severe privacy breaches if misused. Safeguarding this information is of utmost importance, and measures like encryption, secure storage, and FL can help train models without having to share the raw data. The patient's privacy can also be somewhat protected (Chato and Regentova, 2023). The paper presented a method for real-time anomaly detection using FPGA (Field Programmable Gate Array) based IoT edge devices that employ LR classifiers. This system could

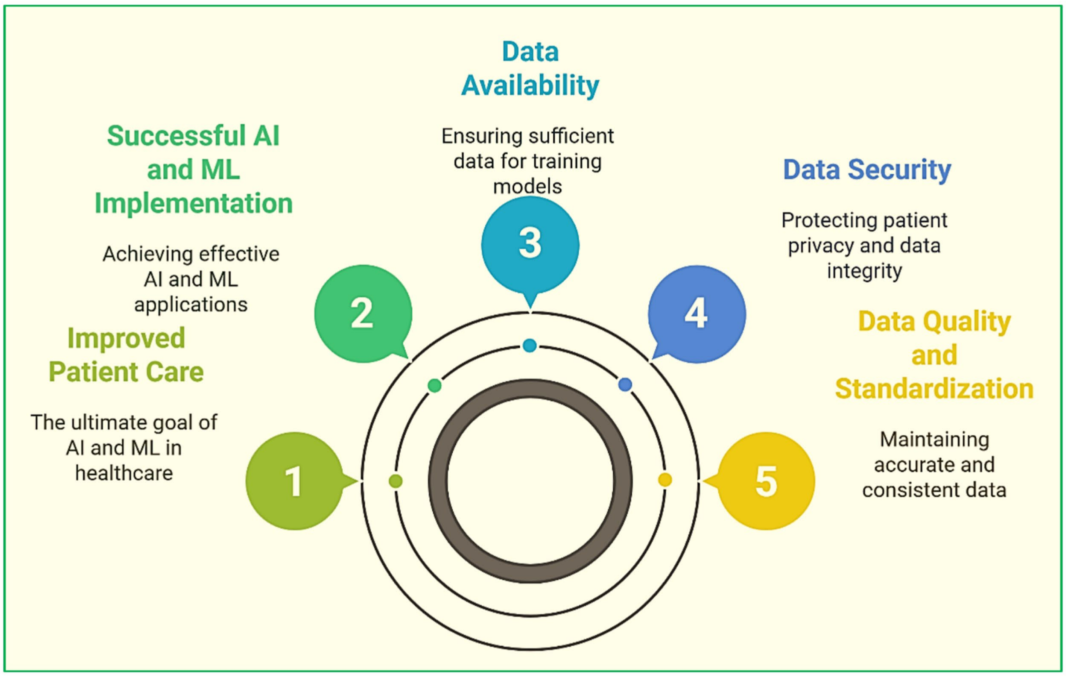


FIGURE 11
Key challenges and strategic solutions for AI and ML integration in healthcare data management.

monitor ingress traffic and predict anomalies at very low latency. It also used hardware-software partitions and introduced a different approach to the twin-based retraining of the LR classifier. The feasibility and timing analysis proved that the system is better than the implementation for software-only systems (Chaitanya et al., 2024).

11.2.2 Informed consent

As regards the development of AI instruments that utilize patient data, informed consent must be provided clearly and comprehensively. Patients must be made aware of how their data will be used, who will be able to access it, and the risks of using it (Linga et al., 2024). The project merges ancient Ayurvedic treatments with a new ML non-invasive Nadi Pariksha system that attempts to diagnose diseases through pulse reading. The system emphasized the analysis of pulses located at the wrist vata, pitta, and kapha regions employing signal processing and advanced ML with RF, which achieved 86.4 percent classification accuracy. The attempt was towards improving efficiency concerning the early detection of diseases, reduction in mortality, clinical validation, and a balance between patient privacy and ethical implementation (Priya et al., 2023).

11.2.3 Bias and fairness

If an AI model is trained with data that includes social constructs of race, gender, and class, ML models are prone to bias. In the healthcare sector, such biases can result in discrimination. Mitigating these biases is vital to guarantee fairness and that AI systems are unbiased towards different groups of patients. The paper illustrated the ethical, legal, and social aspects of integrating AI and healthcare. Focused issues included privacy, bias, transparency, and accountability of using AI in healthcare, all posing significant risks. Even as AI improved diagnostics and treatment processes with excellent economy, efficiency, and effectiveness, questions wanting answers persist concerning equity, safety, and social balance. This to be dealt with over time needed collaboration to mitigate possible risks while trying to achieve balance (Upreti et al., 2023).

11.2.4 Transparency and accountability

Considering all the ML models, especially DL systems, it has been said that they are 'black boxes' since very little is known about their inner workings. In the context of health care, this is an issue because all patients' care choices are supposed to be actionable and dependable on the part of the clinicians. To provide such accountability while at the same time reducing errors that can potentially harm a patient, attention to these issues is needed. The tutorial focused on the topic of responsible AI, which is concerned with proactively positive AI solutions that need explaining, justifying, and being fair and secure. The tutorial illustrated how processes of reproducibility, data provenance, and model management/monitoring of ML models are crucial to ensure responsible development. The purpose of the tutorial was to arm the attendees with the needed knowledge and skills to practice Responsibility AI (Paliwal et al., 2022).

11.3 Integration into clinical practices

AI and ML integration in clinical practice comes with numerous challenges (Gautam et al., 2023; Bai and Mardini, 2024).

11.3.1 Disruption of clinical workflow

Healthcare professionals may be unwilling to adopt AI-based tools because they fear disruption of established workflows. Clinicians are accustomed to traditional methods and may be wary of relying on new technologies, especially if perceived as complex or unreliable. AI tools should complement existing practices without overwhelming healthcare providers (Rush et al., 2019; Sufian et al., 2024).

11.3.2 Regulatory and legal barriers

It is a highly regulated field with the introduction of new technology. It is a field where regulatory standards and approval must be met for implementation. It may refer to the FDA or EMA. The process of acquiring approval for AI-based diagnostic tools is lengthy and pricey, which can delay adoption. There are legal issues concerning liability in case of errors (Handoko et al., 2024).

11.3.3 Lack of training and education

Healthcare professionals will be less likely to embrace these technologies without the appropriate education on the understanding or practical application of AI and ML. Providers will require educational tools and resources explaining the technologies and their potential benefits and limitations to support adoption. Investment in training programs is crucial to ensure clinicians' comfort with incorporating AI into practice. The paper proposed a novel FL framework for early COVID-19 detection using PSO. The framework achieved faster convergence and improved performance, with a 94.36% accuracy on a COVID-19 image dataset from multiple healthcare institutions. FL ensured data privacy by decentralizing patient information and sharing only aggregated model updates (Dasaradharami Reddy et al., 2024).

11.3.4 Interoperability

Healthcare systems have tended to operate independently with minimal integration across technologies and platforms. For effective adoption, AI systems must seamlessly integrate with EHRs, medical devices, and other systems in use within healthcare settings. The main barrier to interoperability is, perhaps, the standardization of data structures and rules negotiated by stakeholders in healthcare settings. Prabha et al. (2024) IoTs will likely change the healthcare domain, especially concerning providing personalized care, telemedicine, and remote care monitoring. Five classification schemes for IoT target energy consumption, privacy, and scalability issues while showing the importance of these problems for managing healthcare data with ML. Solving these problems positioned IoT as a viable means to increase the accessibility of healthcare services and improve the quality of life (Balasingam et al., 2017; Gevers-Montoro et al., 2022).

11.3.5 Trust and adoption

Even when the precision is verified, AI instruments encounter issues because of reliability problems and limited human participation. Clinicians and patients must rely on the trustworthiness of AI-based tools before they allow them to affect clinical or patient decision-making. This trust can be obtained by having a development process that is transparent and genuine validation combined with AI's ongoing performance monitoring. This tutorial concentrated on the concept of Responsible AI, which encompasses the need for transparency, accountability, fairness, and security that AI solution development requires. It highlighted AI's repetitiveness and the data provenance and control of ML models in responsible development. The tutorial concentrated on ensuring the audience is equipped with the knowledge and skills needed to practice Responsible AI (Paliwal et al., 2022).

AI, ML, and emerging technologies hold significant promise to redefine healthcare. However, they must first overcome challenges related to data quality, ethical issues, and clinical integration. These challenges will be prerequisites in determining how best to utilize these technologies for the optimal benefit of both patients and providers. Figure 12 highlights the primary challenges healthcare organizations face in adopting technology and the associated strategic solutions. It identifies key obstacles such as resistance to change, data security challenges, interoperability problems, and significant implementation costs, matched with proactive strategies that encourage an innovative culture, emphasize cybersecurity, support standardized interoperability, and consider alternative funding options. This organized overview underscores that effective integration of digital technologies in healthcare necessitates a comprehensive approach that tackles both technical and human aspects.

12 Future directions and emerging trends in ML-based heart disease prediction

12.1 Emerging trends in healthcare technology

The healthcare sector is experiencing rapid developments in several areas. It is expected to revolutionize how healthcare is delivered (Junaid et al., 2022). Some of these emerging trends include:

12.1.1 Personalized medicine

Faster healthcare delivery is now achievable due to technological progress in genomics, biomarker discovery, and AI analytics. With the implementation of IoT and ML in healthcare, there is a drastic improvement in diagnosis and care delivery, such as early kidney disease detection and proactive precision care (Loncaric et al., 2020). With the adoption of IoT, subtle risk factors were captured in real-time to assist an ML algorithm that supported personal risk models to provide timely interventions. Therefore, the model ensures continuous patient-provider communication is achieved, thus moving towards personalized healthcare while ensuring strong security measures (Ravi et al., 2025).

12.1.2 AI-driven diagnostics

AI and ML focus on improving the accuracy and timing of diagnosing diseases. From analyzing medical imaging like tumor scanning and cardiovascular monitoring to forecasting patient health through EHR logs, the precision of medicine is improved by automation. In the long term, AI promises to become a ready aid for quick and efficient diagnosis (Sufian et al., 2024; Yenurkar et al., 2024). This research utilizes SVM algorithms to distinguish between types of diabetes, Type 1 and 2, using linear, polynomial, and sigmoid functions. It enhanced the degree of classification incidence of diabetes and enabled the formulation of narrower treatment strategies. It showcased the promise that more developed computational

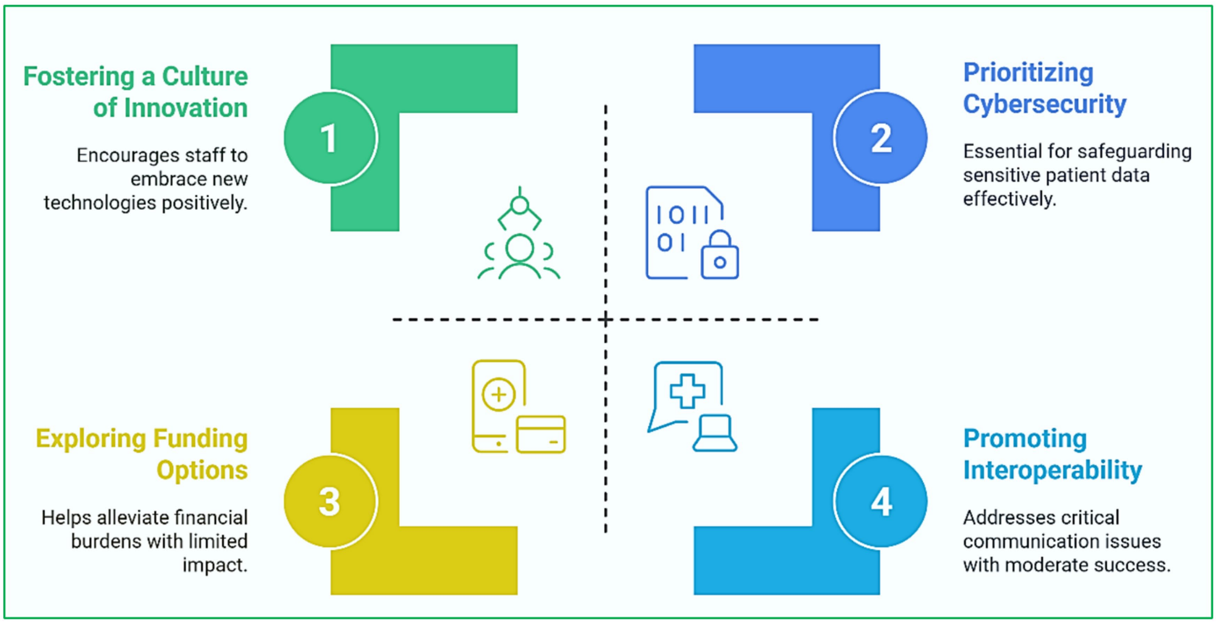


FIGURE 12
Key challenges in healthcare technology adoption: barriers and strategic solutions.

methods can have on medicine with AI-driven healthcare management tasks (Suvartha et al., 2024).

12.1.3 Telemedicine and remote monitoring

Incorporating telemedicine services and remote patient monitoring technologies transforms healthcare delivery, especially in low-resource regions or during public health crises. IoT devices, wearables, and telehealth platforms enable patients to be continuously monitored for their vital signs, medical conditions, and treatment responses (Ardeti et al., 2023; Wang et al., 2024). This development seems poised to continue offering more efficient and inexpensive healthcare services. This work investigated the application of IoT and ML to personalized telemedicine systems with the objective of patient health monitoring and prediction for better care delivery. The effectiveness of the ML model was assessed through recall, sensitivity, error rate, F-score, and other measures, which gave insight into prediction accuracy and complexity. The goal was to shift the paradigm of healthcare to support routine health monitoring and timely proactive interventions (Ramalingam et al., 2024).

12.1.4 Quantum computing (QC)

The emergence of QC promises to revolutionize fields like the modeling of complex diseases and genomics, even improving the standards of drug discovery. Quantum computers could simulate complex biological processes and novel therapeutic targets at speed rivaling thousands of classical computers (Ghazi Enad and Abed Mohammed, 2023). Predicting an enormous acceleration of medical research and precision treatment development is reasonable. In this research, two Quantum ML models, VQC and Pegasos QSVC, were implemented to predict lung cancer in smart healthcare systems. They were benchmarked against baseline models by measuring training accuracy and performing statistical analyses. The Pegasos QSVC model achieved the highest classification accuracy of 85% (Munshi et al., 2024b).

12.1.5 Federated learning (FL)

By not sharing sensitive patient information, FL allows training AI models collaboratively without compromising privacy, which helps gain insights. This innovation facilitates collaboration between healthcare institutions for research and model building while adhering to stringent data privacy regulations. As the fear around the security and privacy of healthcare information further increases, FL is a promising solution for creating AI models that require diverse information while ensuring privacy (Otoum et al., 2024). An ensemble FL approach for developing a DL model to classify data streams from distributed Internet of Medical Things (IoMT) environments was presented. Local federated models were deployed on IoMT devices, with ensemble learning performed on the cloud servers. The comparison results indicated that the ensemble federated model outperformed the primary federated models (Arya and Hanumat Sastry, 2022).

12.1.6 Interdisciplinary approaches for enhanced diagnostics

The merging of ML, AI, and contemporary technologies into healthcare has proven exceedingly useful and requires interdisciplinary teamwork. These technologies' potential is best capitalized with a multidisciplinary approach toward improving diagnostics. This includes:

- Cooperation Between Health Workers and Data Science Specialists: The adoption of AI in healthcare has proven successful through the collaboration of health workers and data science experts. The data science expert optimizes the ML models, and the healthcare expert adds value with their pathophysiology, clinical diagnosis, and therapeutics knowledge. Their collaboration empowers the development of AI-driven tools that are clinically relevant and accurate, hence ensuring enhanced patient and public healthcare.

- Hybridization of Medicine with Engineering: Biomedical engineers and other engineers are responsible for improving and developing new healthcare technology. Wearable devices, medical imaging systems, or diagnostic aids are some examples of other engineering branches that can be integrated. Engineering and medicine can be united to develop better devices or solutions that match the growing demands of patients and healthcare providers.
- Multidisciplinary Data Integration: Healthcare systems that integrate various data sources like EHRs, Genomics, Imaging, Wearables, and even Environmental Data can enhance diagnosis and treatment direction. Such fusion can be accomplished by interdisciplinary teams using integrated platforms that combine the various datasets to create a more wholesome understanding of the patient's health and enhance the efficacy of AI models.
- Ethics, Policy, and Technology Integration: To solve AI-related problems in healthcare, collaboration with ethics and policy experts should also take place. Data use's legal, regulatory, and ethical concerns, such as data privacy, bias, and accountability, must be addressed while technology is developed. Multidisciplinary teams will guarantee the realization of AI tools that are technologically accurate, ethically correct, and legally compliant.
- Patient and Community Involvement: Patients should participate in developing and adopting technology. When patients and communities are brought together through interdisciplinary teams, their issues, preferences, and needs will be addressed so that AI and ML technologies are patient-centric and culturally acceptable. This approach affirms that trust in AI is significantly improved within the frameworks in which diverse populations are formed.

To summarize, the advancement of healthcare technology rests on traditional sectors, which lie in ever-increasing exciting practices and cross-cutting collaborations that will improve diagnostics. New technology like AI, QC, and FL is expected to make healthcare more personalized, accessible, and efficient. Additionally, harnessing multidisciplinary learning will ensure that these results are scientifically valid and ethically responsible, resulting in significant health advantages for patients in all parts of the globe.

13 Conclusion

This review encompasses a wide range of ML applications for predicting heart disease, organized into five main themes: detection and diagnostics, ML models and algorithms, feature engineering and optimization, new technologies in healthcare, and cross-disease AI applications. The findings indicate that while DL models, especially hybrid CNN-LSTM architectures, tend to surpass traditional methods, the success of any model heavily relies on high-quality data, effective feature engineering, and clinical interpretability.

A rising trend in FL, monitoring systems based on wearables, and XAI practices that are gradually narrowing the gap between research and practical application. Nevertheless, significant challenges remain, such as ensuring model generalizability across different populations, maintaining data privacy, achieving interpretability, and integrating with EHRs.

Future studies should prioritize the practical implementation of AI systems via clinically validated trials, focusing on ethical and regulatory standards, and investigating interoperable, patient-centered AI platforms. Moreover, creating transparent models that healthcare providers can trust is essential for broad acceptance. By tackling these issues, ML-driven tools can transition from experimental concepts to game-changing solutions in cardiovascular healthcare.

13.1 Summary of key insights

- ▪ Diagnostic Accuracy: ML and AI have significantly increased accuracy towards diagnosing heart diseases, cancers, and even neurological disorders. The AI systems utilize EHRs, medical imaging, and large datasets alongside sensors to diagnose and improve patient outcomes.
- ▪ Use of Feature Engineering and Its Optimization: Advanced algorithms, GA, PSO, and GWO, have played a crucial role in improving prediction models, which enhances the accuracy with which disease detection can be made. The AI system was optimized for better interpretability and used to determine feature selection, reduce dimensionality, and use other AI-enabled techniques. Moreover, the AI system can select and interpret relevant features, reduce noise, and improve interoperability through Dimensionality Reduction.
- ▪ Emerging Technologies: Healthcare innovation is changing with QC, FL, and the IoT. QC can speed up complex drug discovery and disease modeling. FL alleviates privacy concerns because sensitive information is not expeditiously sent during collaborative model training. Moreover, the application of IoT devices enables proactive fitness healthcare management by providing real-time monitoring.
- ▪ Issues Relating to Data Quality and Ethical Considerations: The availability, quality, and privacy of healthcare data affect AI implementation in healthcare. The success of AI tools hinges on the availability of accurate and secure healthcare data. Other ethical issues, such as consent, privacy, and the algorithms' impartiality raisers, must also be addressed for AI to properly deploy in healthcare.
- ▪ Interdisciplinary Collaboration: Collaboration between healthcare personnel, such as doctors and nurses, data analysts and scientists, engineers, and ethics specialists, is essential for creating and implementing AI systems in healthcare. Interdisciplinary approaches ensure that AI technologies are technically sound and respond ethically and socially to the complex conditions surrounding patients and legal affairs.

13.2 Implications for research and practice

- *Research:*
  Significant work is currently being done to optimize AI model performance and data assimilation and address ethical concerns.

Improving the quality and representativeness of available healthcare datasets will enhance the generalizability of various AI models across different healthcare populations. Further research into other technologies, such as QC and FL, could contribute to advancing AI systems in healthcare. The lack of transparency and interpretability of AI models should also be a focus of future research to foster trust and facilitate smoother integration into clinical decision-making.

- *For Practice:*
  Older healthcare professionals and service providers need to adapt AI and ML-based technologies into their clinical routines to ensure improved diagnostic and treatment results. Sufficient instruction on working with AI systems must be provided alongside the implementation of AI tools to clinicians. Patients should also actively participate in incorporating the technologies so that the solutions work for them. Organizations providing healthcare must also coordinate in fostering interdisciplinary cooperation for the safe, ethical, and beneficial application of AI systems.

AI and ML can transform the health sector, but significant challenges remain in data management, ethics, and deployment. Responsible interdisciplinary collaboration facilitated through innovation has the power to surpass these challenges, which could result in improved health care globally through better utilization of these technologies.

Author contributions

RaK: Conceptualization, Formal analysis, Data curation, Investigation, Methodology, Validation, Visualization, Writing – original draft. SG: Investigation, Software, Writing – review & editing. RuK: Writing – review & editing. MJ: Writing – review & editing. SS: Writing – review & editing. SM: Writing – review & editing. PK: Writing – review & editing. AH: Writing – review & editing. SH: Writing – review & editing. JL: Conceptualization, Formal analysis, Funding acquisition, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2025.1583459/full#supplementary-material

References

Abdellatif, A., Mubarak, H., Abdellatef, H., Kanesan, J., Abdelltif, Y., Chow, C. O., et al. (2024). Computational detection and interpretation of heart disease based on conditional variational auto-encoder and stacked ensemble-learning framework. *Biomed. Signal Process. Control* 88:105644. doi: 10.1016/j.bspc.2023.105644

Abdullah, M. I., Hao, S. K., Abdullah, I., and Faizah, S. (2023). "Parkinson's disease symptom detection using hybrid feature extraction and classification model", in: 2023 IEEE 14th Control and System Graduate Research Colloquium, ICSGRC 2023 – Conference Proceeding), 93–98.

Ahsan, M. M., and Siddique, Z. (2022). Machine learning-based heart disease diagnosis: a systematic literature review. *Artif. Intell. Med.* 128:102289. doi: 10.1016/j.artmed.2022.102289

Al-Ajlouni, Y. A., Al Taani, O., Shamaileh, G., Nagi, Y., Tanashat, M., Al-Bitar, F., et al. (2024). The burden of cardiovascular diseases in Jordan: a longitudinal analysis from the global burden of disease study, 1990–2019. *BMC Public Health* 24:879. doi: 10.1186/s12889-024-18316-0

Al-Hgaish, A., Atan, R., Yaakob, R., and Osman, M. H. (2025). "Enhancing performance of machine learning models in healthcare: an analytical framework for assessing and improving data quality", in: Lecture notes in networks and systems, eds. V. Goar, M. Kuri and R. Kumar & T. Senjyu: Springer Science and Business Media Deutschland GmbH), 137–153.

Amann, J., Blasimme, A., Vayena, E., Frey, D., and Madai, V. I.Precise4Q consortium (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med. Inform. Decis. Mak.* 20:310. doi: 10.1186/s12911-020-01332-6

Ampavathi, A., and Saradhi, T. V. (2021). Multi disease-prediction framework using hybrid deep learning: an optimal prediction model. *Comput. Methods Biomech. Biomed. Engin.* 24, 1146–1168. doi: 10.1080/10255842.2020.1869726

Anand, A., Kadian, T., Shetty, M. K., and Gupta, A. (2022). Explainable AI decision model for ECG data of cardiac disorders. *Biomed. Signal Process. Control* 75:103584. doi: 10.1016/j.bspc.2022.103584

Ardeti, V. A., Kolluru, V. R., Varghese, G. T., and Patjoshi, R. K. (2023). An overview on state-of-the-art electrocardiogram signal processing methods: traditional to AI-based approaches. *Expert Syst. Appl.* 217:119561. doi: 10.1016/j.eswa.2023.119561

Arya, M., and Hanumat Sastry, G. (2022). "Ensemble federated learning for classifying IoMT data streams", in: 2022 IEEE 7th international conference for convergence in technology, I2CT 2022: Institute of Electrical and Electronics Engineers Inc.).

Assegie, T. A., Napa, K. K., Thulasi, T., Kumar, A. K., Priya, M. J. T. V., and Dhamodaran, V. (2024). Scalability and performance of decision tree for cardiovascular disease prediction. *IAES Int. J. Artif. Intell.* 13, 2540–2545. doi: 10.11591/ijai.v13.i3.pp2540-2545

Attalla, S. M., Ibrahim, J., Ads, H., Bakar, H. A., Hamid, N. A. A., and Johar, M. G. M. (2021). Medical students' perception for online learning during COVID-19 pandemic: a CASE study in a Malaysian private medical SCHOOL. *Int. J. Med. Toxicol. Legal Med.* 24, 75–80. doi: 10.5958/0974-4614.2021.00011.5

Bai, C., and Mardini, M. T. (2024). Advances of artificial intelligence in predicting frailty using real-world data: a scoping review. *Ageing Res. Rev.* 101:102529. doi: 10.1016/j.arr.2024.102529

Balasingam, M., Ebrahim, J., and Ariffin, I. A. (2017). Tele-echocardiography – made for astronauts, now in hospitals. *Indian Heart J.* 69, 252–254. doi: 10.1016/j.ihj.2017.01.010

Barfungpa, S. P., Samantaray, L., and Sarma, H. K. D. (2024). SMOTE-based adaptive coati Kepler optimized hybrid deep network for predicting the survival of heart failure patients. *Multimed. Tools Appl.* 83, 65497–65524. doi: 10.1007/s11042-023-18061-3

Baviskar, V., Verma, M., Chatterjee, P., Singal, G., and Gadekallu, T. R. (2023). Optimization using internet of agent based stacked sparse autoencoder model for heart disease prediction. *Expert. Syst.* 41:e13359. doi: 10.1111/exsy.13359

Benhar, H., Idri, A., and Fernández-Alemán, J. L. (2020). Data preprocessing for heart disease classification: a systematic literature review. *Comput. Methods Prog. Biomed.* 195:105635. doi: 10.1016/j.cmpb.2020.105635

Bhatt, H., Jadav, N. K., Kumari, A., Gupta, R., Tanwar, S., Polkowski, Z., et al. (2024). Artificial neural network-driven federated learning for heart stroke prediction in healthcare 4.0 underlying 5G. *Concurr. Comput. Pract. Exp.* 36. doi: 10.1002/cpe.7911

Chaitanya, C., Reddy, Y. A., Pasupuleti, H., Gvk, S., and Bapat, J. (2024). "Real-time anomaly detection at IoT-edge ingress port using FPGA based ML classifiers" in ACM international conference proceeding series. eds. S. Dua and V. Saxena (New York City: Association for Computing Machinery), 309–315.

Chato, L., and Regentova, E. (2023). Survey of transfer learning approaches in the machine learning of digital health sensing data. *J. Pers. Med.* 13:1703. doi: 10.3390/jpm13121703

Cheekati, V. R., Kaja, K. M., Pbv, R. R., Subbarao, G., and Chakravarthi, A. M. (2024). Amalgamate approaches can aid in the early detection of coronary heart disease. *J. Theor. Appl. Inf. Technol.* 102, 6669–6678.

Chen, X., Owen, C. A., Huang, E. C., Maggard, B. D., Latif, R. K., Clifford, S. P., et al. (2021). Artificial intelligence in echocardiography for anesthesiologists. *J. Cardiothorac. Vasc. Anesth.* 35, 251–261. doi: 10.1053/j.jvca.2020.08.048

Dala Ali, A. H. H., Harun, S. N., Othman, N., Ibrahim, B., Abdulbagi, O. E., Abdullah, I., et al. (2023). Determinants of inadequate empiric antimicrobial therapy in ICU Sepsis patients in Al-Madinah Al-Munawwarah, Saudi Arabia: a comparison of artificial neural network and regression analysis. *Antibiotics* 12:1305. doi: 10.3390/antibiotics12081305

Dang, T., Spathis, D., Ghosh, A., and Mascolo, C. (2023). Human-centred artificial intelligence for mobile health sensing: challenges and opportunities. *Royal Soc. Open Science* 10:230806. doi: 10.1098/rsos.230806

Das, S., Selvaraji, K., Dutta, M., Isa, M. Z. A., and Nataraja Moorthy, T. (2018). A pilot study on glaucoma and depth variation, a factor that affect the accuracy of eyewitness identification. *Int. J. Med. Toxicol. Legal Med.* 21, 239–242. doi: 10.5958/0974-4614.2018.00077.3

Dasaradharami Reddy, K., Srivastava, G., Zhu, Y., Supriya, Y., Yenduri, G., Victor, N., et al. (2024). "Federated learning using the particle swarm optimization model for the early detection of COVID-19" in Communications in Computer and Information Science. eds. B. Luo, L. Cheng, Z. G. Wu, H. Li and C. Li (Berlin: Springer Science and Business Media Deutschland GmbH), 425–436.

Decoodt, P., Liang, T. J., Bopardikar, S., Santhanam, H., Eyembe, A., Garcia-Zapirain, B., et al. (2023). Hybrid classical-quantum transfer learning for cardiomegaly detection in chest X-rays. *J. Imaging* 9:128. doi: 10.3390/jimaging9070128

Dehghani, M., Sharpe, L., and Khatibi, A. (2014). Catastrophizing mediates the relationship between fear of pain and preference for elective caesarean section. *Eur. J. Pain (United Kingdom)* 18, 582–589. doi: 10.1002/j.1532-2149.2013.00404.x

Dhingra, L. S., Aminorroaya, A., Sangha, V., Pedroso, A. F., Asselbergs, F. W., Brant, L. C. C., et al. (2025). Heart failure risk stratification using artificial intelligence applied to electrocardiogram images: a multinational study. *Eur. Heart J.* 46, 1044–1053. doi: 10.1093/eurheartj/ehae914

Dissanayake, K., and Johar, M. G. M. (2021). Comparative study on heart disease prediction using feature selection techniques on classification algorithms. *Appl. Comput. Intell. Soft Comput.* 2021, 1–17. doi: 10.1155/2021/5581806

Dissanayake, K., and Johar, M. G. M. (2023). Two-level boosting classifiers ensemble based on feature selection for heart disease prediction. *Indones. J. Electr. Eng. Comput. Sci.* 32, 381–391. doi: 10.11591/ijeecs.v32.i1.pp381-391

Dissanayake, K., Johar, M. G. M., and Ubeysekara, N. H. (2023). Data mining techniques in disease classification: descriptive bibliometric analysis and visualization of global publications. *Int. J. Comput. Digit. Syst.* 13, 289–301. doi: 10.12785/ijcds/130124

Draman, W. M. A. A. W., Chun-Phuoc, J., and Yajid, M. S. A. (2019). Halal certification of medical device in Malaysia: a Shariah analysis into the compliance process and legal requirements. *Int. J. Med. Toxicol. Legal Med.* 22, 94–99. doi: 10.5958/0974-4614.2019.00066.4

Elghanuni, R. H., Helmi, R. A. A., and Abdullah, M. I. (2022). "Improved detection of Facebook anomalies and abnormalities using graph-based and machine learning techniques", in: AIP Conference Proceedings).

Faizah, S., Nugraha, D., Abdulrazaq, M. N., Dionova, B. W., Abdullah, M. I., and Novianti, L. (2024). A rest tremor detection system based on internet of thing technology. *Indones. J. Electr. Eng. Comput. Sci.* 33, 476–484. doi: 10.11591/ijeecs.v33.i1.pp476-484

Faraji, A., Dehghani, M., and Khatibi, A. (2023). Familial aspects of fear of cancer recurrence: current insights and knowledge gaps. *Front. Psychol.* 14:1279098. doi: 10.3389/fpsyg.2023.1279098

Farhud, D. D., and Zokaei, S. (2021). Ethical issues of artificial intelligence in medicine and healthcare. *Iran. J. Public Health* 50, i–v. doi: 10.18502/ijph.v50i11.7600

Fatima, N., and Siddiqi, S. (2024). Acute myocardial infarction: prediction and patient assessment through different ML techniques. *Int. J. Intell. Syst. Appl. Eng.* 12, 106–121.

Fattepur, S., Nilugal, K. C., Rajendran, R., Asmani, F., and Yusuf, E. (2018). Anti-hyperlipidemic activity of methanolic extract of Boesenbergia pandurata (finger root) in experimental induced hypercholestrolemic Sprague dawley rats. *Asian J. Pharm. Clin. Res.* 11, 8–12. doi: 10.22159/ajpcr.2018.v11s3.29962

Ferdous Azam, S. M., and Sarwar, A. (2023). Consumer purchase intention towards traditional Chinese medicine. *Int. J. Bus. Glob.* 34, 391–404. doi: 10.1504/IJBG.2023.133273

Gautam, N., Mueller, J., Alqaisi, O., Gandhi, T., Malkawi, A., Tarun, T., et al. (2023). Machine learning in cardiovascular risk prediction and precision preventive approaches. *Curr. Atheroscler. Rep.* 25, 1069–1081. doi: 10.1007/s11883-023-01174-3

Gevers-Montoro, C., Deldar, Z., Conesa-Buendía, F. M., Lazar, E. A., Mahillo-Fernandez, I., Khatibi, A., et al. (2022). Pain catastrophizing mediates rapid benefits of accessing in-person chiropractic care during the COVID-19 lockdown. *Eur. J. Pain (United Kingdom)* 26, 463–479. doi: 10.1002/ejp.1872

Ghazi, H. F., Elnajeh, M., Abdalqader, M. A., Baobaid, M. F., and Ariffin, I. A. (2018). Non-communicable diseases and its association with body composition and nutrition among general population in Subnag Jaya, Selangor: community-based study. *Pak. J. Nutr.* 17, 255–259. doi: 10.3923/pjn.2018.255.259

Ghazi Enad, H., and Abed Mohammed, M. (2023). A review on artificial intelligence and quantum machine learning for heart disease diagnosis: current techniques, challenges and issues, recent developments, and future directions. *Fusion Pract. Appl.* 11, 08–25. doi: 10.54216/FPA.110101

Goud, P. S., Sastry, P. N., and Sekhar, P. C. (2024). A novel intelligent deep optimized framework for heart disease prediction and classification using ECG signals. *Multimed. Tools Appl.* 83, 34715–34731. doi: 10.1007/s11042-023-16850-4

Gros, C., De Leener, B., Badji, A., Maranzano, J., Eden, D., Dupont, S. M., et al. (2019). Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. *NeuroImage* 184, 901–915. doi: 10.1016/j.neuroimage.2018.09.081

Guleria, P., Naga Srinivasu, P., Ahmed, S., Almusallam, N., and Alarfaj, F. K. (2022). XAI framework for cardiovascular disease prediction using classification techniques. *Electronics [Online]* 11:4086. doi: 10.3390/electronics11244086

Gupta, A., Kumar, R., Arora, H. S., and Raman, B. (2022). C-CADZ: computational intelligence system for coronary artery disease detection using Z-Alizadeh Sani dataset. *Appl. Intell.* 52, 2436–2464. doi: 10.1007/s10489-021-02467-3

Hajiarbabi, M. (2024). Heart disease detection using machine learning methods: a comprehensive narrative review. *Journal of medical. Artif. Intell.* 7:21. doi: 10.21037/jmai-23-152

Handoko, M. L., de Man, F. S., Brugts, J. J., van der Meer, P., Rhodius-Meester, H. F. M., Schaap, J., et al. (2024). Embedding routine health care data in clinical trials: with great power comes great responsibility. *Neth. Hear. J.* 32, 106–115. doi: 10.1007/s12471-023-01837-5

Hassan, M. M., Zaman, S., Rahman, M. M., Bairagi, A. K., El-Shafai, W., Rathore, R. S., et al. (2024). Efficient prediction of coronary artery disease using machine learning algorithms with feature selection techniques. *Comput. Electr. Eng.* 115:109130. doi: 10.1016/j.compeleceng.2024.109130

Helmi, R. A. A., Elghanuni, R. H., and Abdullah, M. I. (2021). "Effect the graph metric to detect anomalies and non-anomalies on Facebook using machine learning models", in: 2021 IEEE 12th Control and System Graduate Research Colloquium, ICSGRC 2021 – Proceedings), 7–12.

Hulsen, T. (2023). Explainable artificial intelligence (XAI): concepts and challenges in healthcare. *AI [Online]* 4, 652–666. doi: 10.3390/ai4030034

Ibrahim, J., Attalla, S. M., Jayaram, J., and Ariffin, I. A. (2018). Adherence of medical students to medical ethics during clinical training – pilot study. *Int. J. Med. Toxicol. Legal Med.* 21, 208–211. doi: 10.5958/0974-4614.2018.00088.8

Ishaque, M., Johar, M. G. M., Khatibi, A., and Yamin, M. (2022). "Intrusion detection system using binary and multiclass deep neural network classification", in: Proceedings of the 2022 9th International Conference on Computing for Sustainable Global Development, INDIACom 2022), 749–753.

Ishaque, M., Johar, M. G. M., Khatibi, A., and Yamin, M. (2023). Hybrid deep learning-based intrusion detection system using modified chicken swarm optimization algorithm. *ARPN J. Eng. Appl. Sci.* 18, 1707–1718. doi: 10.59018/0723212

Jafari, M., Shoeibi, A., Khodatars, M., Ghassemi, N., Moridian, P., Alizadehsani, R., et al. (2023). Automated diagnosis of cardiovascular diseases from cardiac magnetic resonance imaging using deep learning models: a review. *Comput. Biol. Med.* 160:106998. doi: 10.1016/j.compbiomed.2023.106998

Jagannathan, R., Patel, S. A., Ali, M. K., and Narayan, K. V. (2019). Global updates on cardiovascular disease mortality trends and attribution of traditional risk factors. *Curr. Diab. Rep.* 19, 1–12. doi: 10.1007/s11892-019-1161-2

Jayakiran, K., Pranavi, M., Novika, M., Tejaswini, V., and Rajesh, N. (2019). Prediction of heart diseases in comparison with different machine learning algorithms. *Int. J. Innov. Technol. Explor. Eng.* 8, 328–332.

Jayasree, L., and Usha, D. (2022). Statistical analysis with machine and neural learning-based model on cardiovascular diseases and stroke prediction. *Int. J. Recent Innov. Trends Comput. Commun.* 10, 127–138. doi: 10.17762/ijritcc.v10i2s.5919

Jenefer, G. G., Deepa, A. J., and Linda, M. M. (2024). Diabetic prediction and classification of risk level using ODDTADC method in big data analytics. *J. Comb. Optim.* 47:80. doi: 10.1007/s10878-024-01179-x

Jiao, L., and Abdullah, M. I. (2024). Neutrosophic enhancement of YOLO-MD algorithm for automated metal surface Micro defect detection. *Int. J. Neutrosophic Sci.* 23, 308–316. doi: 10.54216/IJNS.230225

Jimmy, D. D., and Bakar, H. A. (2023). "Evaluation of urinary tract infection among neonates and children using urine FEME and urine culture", in: AIP Conference Proceedings).

Joshi, M., Pal, A., and Sankarasubbu, M. (2022). Federated learning for healthcare domain – pipeline, applications and challenges. *ACM Trans. Comput. Healthcare* 3, 1–36. doi: 10.1145/3533708

Jothi Prakash, V., Arul Antran Vijay, S., Ganesh Kumar, P., and Karthikeyan, N. K. (2024). A novel attention-based cross-modal transfer learning framework for predicting cardiovascular disease. *Comput. Biol. Med.* 170:107977. doi: 10.1016/j.compbiomed.2024.107977

Junaid, S. B., Imam, A. A., Balogun, A. O., De Silva, L. C., Surakat, Y. A., Kumar, G., et al. (2022). Recent advancements in emerging Technologies for Healthcare Management Systems: a survey. *Healthcare (Basel)* 10:1940. doi: 10.3390/healthcare10101940

Kayalvizhi, K., Kanchana, S., Silvia Priscila, S., Padmavathy, C., Ramana Murthy, B. S. N. V., and Thangavel, V. (2024). Optimized CNN-BiLSTM with attention: a high performance model for predicting heart disease using Cleveland and Framingham datasets. *J. Mach. Comput.* 4, 1195–1205. doi: 10.53759/7669/jmc202404110

Khan Minhas, A. M., Sedhom, R., Jean, E. D., Shapiro, M. D., Panza, J. A., Alam, M., et al. (2024). Global burden of cardiovascular disease attributable to smoking, 1990–2019: an analysis of the 2019 global burden of disease study. *Eur. J. Prev. Cardiol.* 31, 1123–1131. doi: 10.1093/eurjpc/zwae040

Khatibi, A., Dehghani, M., Sharpe, L., Asmundson, G. J. G., and Pouretemad, H. (2009). Selective attention towards painful faces among chronic pain patients: evidence from a modified version of the dot-probe. *Pain* 142, 42–47. doi: 10.1016/j.pain.2008.11.020

Khatibi, A., Sharpe, L., Jafari, H., Gholami, S., and Dehghani, M. (2015). Interpretation biases in chronic pain patients: an incidental learning task. *Eur. J. Pain (United Kingdom)* 19, 1139–1147. doi: 10.1002/ejp.637

Kishor, A., and Chakraborty, C. (2022). Artificial intelligence and internet of things based healthcare 4.0 monitoring system. *Wirel. Pers. Commun.* 127, 1615–1631. doi: 10.1007/s11277-021-08708-5

Kumar, K. A., and Belinda, M. J. C. M. (2024). A multi-layer acoustic neural network-based intelligent early diagnosis system for rheumatic heart disease. *Int. J. Image Graph.* 24:2450012. doi: 10.1142/S0219467824500128

Kwon, S. H., and Dong, L. (2022). Flexible sensors and machine learning for heart monitoring. *Nano Energy* 102:107632. doi: 10.1016/j.nanoen.2022.107632

Lenin, S. T., and Venkatasalam, K. (2024). Effective classification of heart disease using convolutional neural networks. *Circuits Syst. Signal Process.* 44, 911–935. doi: 10.1007/s00034-024-02851-7

Linga, B. G., Mohammed, S. G. A. A., Farrell, T., Rifai, H. A., Al-Dewik, N., and Qoronfleh, M. W. (2024). Genomic newborn screening for pediatric Cancer predisposition syndromes: a holistic approach. *Cancers* 16. doi: 10.3390/cancers16112017

Loncaric, F., Cikes, M., Sitges, M., and Bijnens, B. (2020). Comprehensive data integration—toward a more personalized assessment of diastolic function. *Echocardiography* 37, 1926–1935. doi: 10.1111/echo.14749

Majhi, B., and Kashyap, A. (2024). Explainable AI-driven machine learning for heart disease detection using ECG signal. *Appl. Soft Comput.* 167:112225. doi: 10.1016/j.asoc.2024.112225

Mandula, A., and Vijaya Kumar, B. S. (2024). Integrated feature selection and ensemble learning for heart disease detection: a 2-tier approach with ALAN and ET-ABDF machine learning model. *Int. J. Inf. Technol. (Singapore)* 16, 4489–4503. doi: 10.1007/s41870-024-02016-4

Maran, P. S. (2023). Multi-layered architecture convolution neural networks for diagnosing and predicting heart diseases on multi-modal data. *Malaysian J. Comput. Sci.* sp2023, 29–43. doi: 10.22452/mjcs.sp2023no1.3

Market Access, T. (2025). AI-ECG detects elevated cardiovascular risk in women.

Marzo, R. R., Chen, H. W. J., Abid, K., Chauhan, S., Kaggwa, M. M., Essar, M. Y., et al. (2022). Adapted digital health literacy and health information seeking behavior among lower income groups in Malaysia during the COVID-19 pandemic. *Front. Public Health* 10:998272. doi: 10.3389/fpubh.2022.998272

Matthews, D., Elgueta Cancino, E., Falla, D., and Khatibi, A. (2021). Exploring pain interference with motor skill learning in humans: a protocol for a systematic review. *BMJ Open* 11:e045841. doi: 10.1136/bmjopen-2020-045841

Mennella, C., Maniscalco, U., De Pietro, G., and Esposito, M. (2024). Ethical and regulatory challenges of AI technologies in healthcare: a narrative review. *Heliyon* 10:e26297. doi: 10.1016/j.heliyon.2024.e26297

Mienye, I. D., Obaido, G., Jere, N., Mienye, E., Aruleba, K., Emmanuel, I. D., et al. (2024). A survey of explainable artificial intelligence in healthcare: concepts, applications, and challenges. *Inform. Med. Unlocked* 51:101587. doi: 10.1016/j.imu.2024.101587

Mishra, I., and Mohapatra, S. (2023). An enhanced approach for analyzing the performance of heart stroke prediction with machine learning techniques. *Int. J. Inf. Technol. (Singapore)* 15, 3257–3270. doi: 10.1007/s41870-023-01321-8

Mohammadi, S., Dehghani, M., Sharpe, L., Heidari, M., Sedaghat, M., and Khatibi, A. (2012). Do main caregivers selectively attend to pain-related stimuli in the same way that patients do? *Pain* 153, 62–67. doi: 10.1016/j.pain.2011.08.021

Mohapatra, S., Mishra, I., and Mohanty, S. (2023). Stacking model for heart stroke prediction using machine learning techniques. *EAI Endorsed Trans. Pervasive Health Technol.* 9, 1–5. doi: 10.4108/eetpht.9.4057

Mohsin, S. S., Salman, O. H., Jasim, A. A., Al-Nouman, M. A., and Kairaldeen, A. R. (2025). A systematic review on the roles of remote diagnosis in telemedicine system: coherent taxonomy, insights, recommendations, and open research directions for intelligent healthcare solutions. *Artif. Intell. Med.* 160:103057. doi: 10.1016/j.artmed.2024.103057

Munshi, M., Gupta, R., Jadav, N. K., Polkowski, Z., Tanwar, S., Alqahtani, F., et al. (2024a). Quantum machine learning-based framework to detect heart failures in healthcare 4.0. *Softw. Pract. Experience* 54, 168–185. doi: 10.1002/spe.3264

Munshi, M., Gupta, R., Jadav, N. K., Tanwar, S., Nair, A., and Garg, D. (2024b). "Quantum machine learning-based lung Cancer prediction framework for healthcare 4.0" in: 2024 Asia Pacific Conference on Innovation in Technology, APCIT 2024: Institute of Electrical and Electronics Engineers Inc.).

Muthu Ganesh, V., and Nithiyanantham, J. (2022). Heuristic-based channel selection with enhanced deep learning for heart disease prediction under WBAN. *Comput. Methods Biomech. Biomed. Engin.* 25, 1429–1448. doi: 10.1080/10255842.2021.2013828

Muzammil, M. A., Javid, S., Afridi, A. K., Siddineni, R., Shahabi, M., Haseeb, M., et al. (2024). Artificial intelligence-enhanced electrocardiography for accurate diagnosis and management of cardiovascular diseases. *J. Electrocardiol.* 83, 30–40. doi: 10.1016/j.jelectrocard.2024.01.006

Nadeem, M. W., Goh, H. G., Khan, M. A., Hussain, M., Mushtaq, M. F., and Ponnusamy, V. A. P. (2021). Fusion-based machine learning architecture for heart disease prediction. *Comput. Mater. Continua* 67, 2481–2496. doi: 10.32604/cmc.2021.014649

Nancy, A. A., Ravindran, D., Raj Vincent, P. M. D., Srinivasan, K., and Gutierrez Reina, D. (2022). IoT-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning. *Electronics (Switzerland)* 11:2292. doi: 10.3390/electronics11152292

Napa, K. K., Kumar, A. K., Murugan, S., Mahammad, K., Assegie, T. A., and Okokpujie, K. (2024). Early prediction of chronic heart disease with recursive feature elimination and supervised learning techniques. *IAES Int. J. Artif. Intell.* 13, 730–736. doi: 10.11591/ijai.v13.i1.pp730-736

Naruka, V., Arjomandi Rad, A., Subbiah Ponniah, H., Francis, J., Vardanyan, R., Tasoudis, P., et al. (2022). Machine learning and artificial intelligence in cardiac transplantation: a systematic review. *Artif. Organs* 46, 1741–1753. doi: 10.1111/aor.14334

Natarajan, K., Vinoth Kumar, V., Mahesh, T. R., Abbas, M., Kathamuthu, N., Mohan, E., et al. (2024). Efficient heart disease classification through stacked ensemble with optimized firefly feature selection. *Int. J. Comput. Intell. Syst.* 17:174. doi: 10.1007/s44196-024-00538-0

Otoum, Y., Hu, C., Said, E. H., and Nayak, A. (2024). Enhancing heart disease prediction with federated learning and Blockchain integration. *Future Internet* 16:372. doi: 10.3390/fi16100372

Paliwal, M., Rao, D., and Tarcar, A. (2022). "Responsible AI Tutorial", in: ACM International Conference Proceeding Series: Association for Computing Machinery), 339–341.

Pathirana, S., Asirvatham, D., and Johar, M. G. M. (2018). "A multi-agent system for improving electroencephalographic data classification accuracy", in: IEEE Region 10 Humanitarian Technology Conference, R10-HTC.

Pathirana, S., Asirvatham, D., and M'D Johar, M. G. (2019). Applicability of multi-agent systems for electroencephalographic data classification. *Proc. Comput. Sci.* 152, 36–43. doi: 10.1016/j.procs.2019.05.024

Patil, R. S., Arjariya, T., and Gangwar, M. (2023). Detection of cardiac abnormalities and heart disease using machine learning techniques. *Int. J. Intell. Syst. Appl. Eng.* 11, 598–605.

Pouretemad, H. R., Khatibi, A., Ganjavi, A., Shams, J., and Zarei, M. (2009). Validation of addenbrooke's cognitive examination (ACE) in a persian-speaking population. *Dement. Geriatr. Cogn. Disord.* 28, 343–347. doi: 10.1159/000252772

Prabha, C., Garg, G., and Ahuja, B. (2024) "IoT in healthcare: applications, challenges and future concerns", in: 2024 OPJU international technology conference on smart computing for innovation and advancement in industry 4.0, OTCON 2024: Institute of Electrical and Electronics Engineers Inc..

Prieto-Avalos, G., Cruz-Ramos, N. A., Alor-Hernandez, G., Sánchez-Cervantes, J. L., Rodriguez-Mazahua, L., and Guarneros-Nolasco, L. R. (2022). Wearable devices for physical monitoring of heart: a review. *Biosensors* 12:292. doi: 10.3390/bios12050292

Priya, E., Pradeep, S., Krishna, K. S., and Saran, S. (2023). "Characterization of wrist pulse using machine learning techniques", in: 2023 Intelligent Computing and Control for Engineering and Business Systems, ICCEBS 2023: Institute of Electrical and Electronics Engineers Inc.).

Priyadarshi, R., Ranjan, R., Vishwakarma, A. K., Yang, T., and Rathore, R. S. (2024). Exploring the Frontiers of unsupervised learning techniques for diagnosis of cardiovascular disorder: a systematic review. *IEEE Access* 12, 139253–139272. doi: 10.1109/ACCESS.2024.3468163

Ramalingam, S., Dharani, S. P., Priya, R., Devi, E. P., and Pavithra, S. (2024). "Machine learning based remote health care data monitoring and prediction system using IoMT", in: 10th International Conference on Advanced Computing and Communication Systems, ICACCS 2024: Institute of Electrical and Electronics Engineers Inc.), 2113–2118.

Ravi, M., Krishnaiah, P., Dileep, C., Annapoorna, B., Reddy, M. J., and Satyanarayana, B. (2025). "Early detection of kidney disease risk factors through IoT-enabled machine learning systems" in Lecture notes in electrical engineering. eds. A. Kumar, V. K. Gunjan, S. Senatore and Y. C. Hu (Berlin: Springer Science and Business Media Deutschland GmbH), 1320–1333.

Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., et al. (2020). Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. *J. Am. Coll. Cardiol.* 76, 2982–3021. doi: 10.1016/j.jacc.2020.11.010

Roy, T. S., Roy, J. K., and Mandal, N. (2023). Design and development of electronic stethoscope for early screening of valvular heart disease prediction. *Biomed. Signal Process. Control* 86:105086. doi: 10.1016/j.bspc.2023.105086

Roy, T. S., Roy, J. K., and Mandal, N. (2024). Improvement in the performance of deep learning based on CNN to classify the heart sound by evaluating hyper-parameters. *Soft. Comput.* 28, 11587–11600. doi: 10.1007/s00500-024-09909-3

Rush, B., Celi, L. A., and Stone, D. J. (2019). Applying machine learning to continuously monitored physiological data. *J. Clin. Monit. Comput.* 33, 887–893. doi: 10.1007/s10877-018-0219-z

Sakthisudhan, K., Pamungkas, W., Isnawati, A. F., Ahmad, U. A., Mohan Raj, S., and Septiawan, R. R. (2024). The development of an IoT-based cyber-physical system for Parkinson's disease patients using ML prediction algorithm. *El-Cezeri J. Sci. Eng.* 11, 256–266. doi: 10.31202/ecjse.1419630

Saleh Alghamdi, N., Zakariah, M., Shankar, A., and Viriyasitavat, W. (2024). Heart disease prediction using autoencoder and DenseNet architecture. *Egypt. Inform. J.* 28:100559. doi: 10.1016/j.eij.2024.100559

Sau, A., Sieliwonczyk, E., Patlatzoglou, K., Pastika, L., McGurk, K. A., Ribeiro, A. H., et al. (2025). Artificial intelligence-enhanced electrocardiography for the identification of a sex-related cardiovascular risk continuum: a retrospective cohort study. *Lancet Digit. Health* 7, e184–e194. doi: 10.1016/j.landig.2024.12.003

Saw, S. N., Yan, Y. Y., and Ng, K. H. (2024). Current status and future directions of explainable artificial intelligence in medical imaging. *Eur. J. Radiol.* 183:111884. doi: 10.1016/j.ejrad.2024.111884

Seeli, D. J. J., and Thanammal, K. K. (2024). Quantitative analysis of gradient descent algorithm using scaling methods for improving the prediction process based on artificial neural network. *Multimed. Tools Appl.* 83, 15677–15691. doi: 10.1007/s11042-023-16136-9

Seneviratne, K., Hamid, J. A., Khatibi, A., Azam, F., and Sudasinghe, S. (2019). Multi-faceted professional development designs for science teachers' self-efficacy for inquiry-based teaching: a critical review. *Univ. J. Educ. Res.* 7, 1595–1611. doi: 10.13189/ujer.2019.070714

Sheik Abdullah, A., Naga Pranava Shashank, V., and Altrin Lloyd Hudson, D. (2024). Disseminating the risk factors with enhancement in precision medicine using comparative machine learning models for healthcare data. *IEEE Access* 12, 72794–72812. doi: 10.1109/ACCESS.2024.3400023

Shimpi, J. K., Shanmugam, P., and Stonier, A. A. (2024). Analytical model to predict diabetic patients using an optimized hybrid classifier. *Soft. Comput.* 28, 1883–1892. doi: 10.1007/s00500-023-09487-w

Singh, M., Kumar, A., Khanna, N. N., Laird, J. R., Nicolaides, A., Faa, G., et al. (2024). Artificial intelligence for cardiovascular disease risk assessment in personalised framework: a scoping review. *EClinicalMedicine* 73:102660. doi: 10.1016/j.eclinm.2024.102660

Singh, K., and Malhotra, D. (2023). Meta-health: learning-to-learn (Meta-learning) as a next generation of deep learning exploring healthcare challenges and solutions for rare disorders: a systematic analysis. *Arch. Comput. Methods Eng.* 30, 4081–4112. doi: 10.1007/s11831-023-09927-8

Siramshetty, V. B., Chen, Q., Devarakonda, P., and Preissner, R. (2018). The Catch-22 of predicting hERG blockade using publicly accessible bioactivity data. *J. Chem. Inf. Model.* 58, 1224–1233. doi: 10.1021/acs.jcim.8b00150

Srinivasulu, A., and Pushpa, A. (2020). Disease prediction in big data healthcare using extended convolutional neural network techniques. *Int. J. Adv. Appl. Sci.* 9, 85–92. doi: 10.11591/ijaas.v9.i2.pp85-92

Stafie, C. S., Sufaru, I. G., Ghiciuc, C. M., Stafie, I. I., Sufaru, E. C., Solomon, S. M., et al. (2023). Exploring the intersection of artificial intelligence and clinical healthcare: a multidisciplinary review. *Diagnostics (Basel)* 13:1995. doi: 10.3390/diagnostics13121995

Sufian, M. A., Hamzi, W., Zaman, S., Alsadder, L., Hamzi, B., Varadarajan, J., et al. (2024). Enhancing clinical validation for early cardiovascular disease prediction through simulation, AI, and web technology. *Diagnostics* 14:1308. doi: 10.3390/diagnostics14121308

Suvartha, M., Bhoomika, S., Kalyan, R., Nandan, T. M., Pavithra, G., and Manjunath, T. C. (2024) "Multi-type diabetes using machine learning – a new method of Prediction & Determination", in: 15th international conference on advances in computing, control, and telecommunication technologies, ACT 2024, eds. J. Stephen, P. Sharma, Y. Chaba, K. U. Abraham, P. K. Anooj and N. Mohammadet al.: Grenze Scientific Society International Journal of Engineering & Technology (GIJET) 10, 6805–6810.

Talaat, F. M., Elnaggar, A. R., Shaban, W. M., Shehata, M., and Elhosseini, M. (2024). CardioRiskNet: a hybrid AI-based model for explainable risk prediction and prognosis in cardiovascular disease. *Bioengineering* 11:822. doi: 10.3390/bioengineering11080822

Todd, J., Sharpe, L., Colagiuri, B., and Khatibi, A. (2016). The effect of threat on cognitive biases and pain outcomes: an eye-tracking study. *Eur. J. Pain (United Kingdom)* 20, 1357–1368. doi: 10.1002/ejp.887

Tompra, K. V., Papageorgiou, G., and Tjortjis, C. (2024). Strategic machine learning optimization for cardiovascular disease prediction and high-risk patient identification. *Algorithms* 17:178. doi: 10.3390/a17050178

Upreti, K., Vats, P., Nasir, M. S., Alam, M. S., Shahi, F. I., and Kamal, M. S. (2023). "Managing the ethical and sociologically aspects of AI incorporation in medical healthcare in India: unveiling the conundrum", in: Proceedings – 2023 IEEE world conference on applied intelligence and computing, AIC 2023: Institute of Electrical and Electronics Engineers Inc., eds. G. S. Tomar and J. Bansal, 293–299.

van der Velden, B. H. M., Kuijf, H. J., Gilhuijs, K. G. A., and Viergever, M. A. (2022). Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med. Image Anal.* 79:102470. doi: 10.1016/j.media.2022.102470

Vellore Pichandi, K., Janarthanan, V., Annamalai, T., and Arumugam, M. (2024). Enhancing healthcare in the digital era: a secure e-health system for heart disease prediction and cloud security. *Expert Syst. Appl.* 255:124479. doi: 10.1016/j.eswa.2024.124479

Venkata MahaLakshmi, N., and Rout, R. K. (2024). An intelligence method for heart disease prediction using integrated filter-evolutionary search based feature selection and optimized ensemble classifier. *Multimed. Tools Appl.* 83, 39841–39865. doi: 10.1007/s11042-023-16924-3

Vetrithangam, D., Krishnan, S. B., Shankar, S. S., and Chakrabarti, P. (2024). Integrating canonical correlation analysis with random Forest for heart disease prediction. *J. Mach. Comput.* 4, 1180–1194. doi: 10.53759/7669/jmc202404109

Vizhi, K., and Dash, A. (2020). Diabetes prediction using machine learning. *Int. J. Adv. Sci. Technol.* 29, 2842–2852.

Wang, R., Veera, S. C. M., Asan, O., and Liao, T. (2024). A systematic review on the use of consumer-based ECG wearables on cardiac health monitoring. *IEEE J. Biomed. Health Inform.* 28, 6525–6537. doi: 10.1109/JBHI.2024.3456028

Wei, Q., Binti Ariffin, I. A., and Tham, J. (2023). Comparison of effects of teacher training programs on TPACK and self-efficacy of primary school teachers. *Eurasian J. Educ. Res.* 2023, 31–47. doi: 10.14689/ejer.2023.107.003

Yajie, L., Johar, M. G. M., and Hajamydeen, A. I. (2023). Poultry disease early detection methods using deep learning technology. *Indones. J. Electr. Eng. Comput. Sci.* 32, 1712–1723. doi: 10.11591/IJEECS.V32.I3.PP1712-1723

Yenurkar, G., Mal, S., Nyangaresi, V. O., Kamble, S., Damahe, L., and Bankar, N. (2024). Revolutionizing chronic heart disease management: the role of IoT-based ambulatory blood pressure monitoring system. *Diagnostics* 14:1297. doi: 10.3390/diagnostics14121297

Zhang, L., Zhu, Y., Ni, Q., Zheng, X., Gao, Z., and Zhao, Q. (2024). Local/global explainability empowered expert-involved frameworks for essential tremor action recognition. *Biomed. Signal Process. Control* 95:106457. doi: 10.1016/j.bspc.2024.106457

Glossary

ASVM - Acoustic Support Vector Machine

ADASYN - Adaptive Synthetic Sampling

aTIG-FS - Adaptive Threshold Information Gain Feature Selection

AUC - Area Under the Curve

AI - Artificial Intelligence

ANN - Artificial Neural Network

ABCM - Attention-Based Cross-Modal Transfer Learning

Bi-LSTM - Bidirectional Long Short-Term Memory

CVD - Cardiovascular Disease

CNN - Convolutional Neural Network

DT - Decision Tree

DCGAN - Deep Convolutional Generative Adversarial Network

DL - Deep Learning

ENN - Edited Nearest Neighbor

ECG - Electrocardiogram

EHR - Electronic Health Record

XAI - Explainable Artificial Intelligence

XGB - Extreme Gradient Boosting

FL - Federated Learning

FDA - Food and Drug Administration

GDPR - General Data Protection Regulation

GANs - Generative Adversarial Networks

HIPAA - Health Insurance Portability and Accountability Act

KNN - K-Nearest Neighbor

LSTM - Long Short-Term Memory

ML - Machine Learning

NB - Naive Bayes

PCG - Phonocardiogram

PTB-XL - Physikalisch-Technische Bundesanstalt Extended ECG Dataset

PCA - Principal Component Analysis

QC - Quantum Computing

QSVC - Quantum Support Vector Classifier

RF - Random Forest

ROC - Receiver Operating Characteristic

RNN - Recurrent Neural Network

RFE - Recursive Feature Elimination

RBM - Restricted Boltzmann Machine

SCSO - Sand Cat Swarm Optimization

SHAP - SHapley Additive exPlanations

SVM - Support Vector Machine

SMOTE - Synthetic Minority Over-sampling Technique

VQC - Variational Quantum Classifier

WBAN - Wireless Body Area Network

ARTICLES FOR FACULTY MEMBERS

AN IMPROVED MLP-RFE MODEL FOR ENHANCED ACCURACY IN HEART DISEASE PREDICTION USING DEEP LEARNING

An interpretable machine learning model for predicting all-cause in-hospital mortality in patients with ischemic stroke and coronary heart disease: A multicenter study in China. / Huang, J., Peng, S., Xiang, T., Zhang, Y., Dong, Y., Wang, Q., & Liu, X.

Brain Network Disorders
Volume 2 (2026) Pages 49 - 58
https://doi.org/10.1177/22808000231180390
(Database: ScienceDirect)

Brain Network Disorders 2 (2026) 49–58

Contents lists available at ScienceDirect

Brain Network Disorders

journal homepage: www.sciencedirect.com/journal/brain-network-disorders

Original Article

An interpretable machine learning model for predicting all-cause in-hospital mortality in patients with ischemic stroke and coronary heart disease: A multicenter study in China

Jian Huang [a,g], Shengxian Peng [b,g], Tianyu Xiang [c,g], Yang Zhang [d], Yalin Dong [e], Qianqian Wang [e], Xiaozhu Liu [f,*]

[a] *Department of Ultrasound, Sir Run Run Shaw Hospital, Zhejiang University College of Medicine, Hangzhou, 310016, China*
[b] *Scientific Research Department, First People's Hospital of Zigong City, Zigong, Sichuan, 643000, China*
[c] *Information Center, University-Town Hospital of Chongqing Medical University, Chongqing, 401331, China*
[d] *Information Department, Kaijiang County People's Hospital, Dazhou, Sichuan, 636250, China*
[e] *College of Medical Informatics, Chongqing Medical University, Chongqing, 400016, China*
[f] *Department of Critical Care Medicine, Clinical and Research Center on Acute Lung Injury, Emergency and Critical Care Center, Beijing Shijitan Hospital, Capital Medical University, Beijing, 100038, China*

ARTICLE INFO

Keywords:
Cardio-cerebral infarction
Coronary heart disease
Interpretable machine learning
Ischemic stroke
Mortality
SHAP

ABSTRACT

Background: Timely and accurate prediction of clinical outcomes in patients with cardio-cerebral infarction (CCI) is essential for optimizing treatment strategies and improving prognosis. This study aimed to develop an interpretable machine learning model for predicting all-cause in-hospital mortality in patients with CCI.
Methods: A total of 4105 patients with CCI from a multicenter database were retrospectively analyzed. Nine machine learning (ML) models were constructed to predict in-hospital mortality. Feature selection was performed using the recursive feature elimination algorithm. Model performance was evaluated using the F1 score, area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and accuracy. The model with the highest predictive performance was further interpreted using the SHapley Additive exPlanations (SHAP) method to provide individualized explanations for predictions.
Results: Of the 4105 patients, 2071 (50.5 %) were female, with a median age of 79 years (interquartile range [IQR]: 72, 85). After feature selection, eight of the initial 43 clinical variables were retained for model development. Among the nine models, the Random Forest (RF) model achieved the highest predictive performance, with an AUC of 0.88 in the test cohort. Feature importance analysis identified serum potassium (K), N-terminal pro-B-type natriuretic peptide (NT-proBNP), troponin T, cholinesterase, creatinine, age, albumin (ALB), and platelet count (PLT) as the most influential predictors. SHAP visualization was applied to enhance the interpretability of the RF model.
Conclusions: The proposed ML-based risk prediction model demonstrated robust performance in predicting in-hospital mortality among patients with CCI. The integration of SHAP analysis provided transparent, patient-specific risk interpretation, thereby supporting clinical understanding of critical prognostic factors and enabling personalized decision-making.

1. Introduction

Cardio-cerebral infarction (CCI), defined as the coexistence of coronary heart disease (CHD) and ischemic stroke (IS), represents a significant global health concern. In 2019, cardiovascular diseases—including CHD and IS—affected an estimated 523 million individuals worldwide, resulting in 18.6 million deaths and 393 million disability-adjusted life years, reflecting a 1.4-fold increase compared with 1990.[1–3] CCI poses a

---

[*] Corresponding author. Department of Critical Care Medicine, Clinical and Research Center on Acute Lung Injury, Emergency and Critical Care Center, Beijing Shijitan Hospital, Capital Medical University, Beijing, 100038, China.

E-mail address: xiaozhuliu2021@163.com (X. Liu).

[g] Jian Huang, Shengxian Peng and Tianyu Xiang contributed equally to this work.

https://doi.org/10.1016/j.bnd.2025.12.003
Received 20 May 2025; Received in revised form 27 November 2025; Accepted 11 December 2025

particularly high burden due to its elevated mortality rate, long-term disability, and extensive healthcare demands.[4,5] Survivors of IS often experience profound functional impairments that markedly diminish their quality of life.[6] Reliable prognostic predictions not only provide patients with realistic expectations but also assist clinicians in selecting optimal rehabilitation strategies.[7,8] Despite extensive research on CHD and IS separately, few studies have examined their co-occurrence, limiting understanding of their shared etiologies and spatial distribution patterns. Emerging evidence highlights substantial overlap in modifiable risk factors and underlying pathophysiological mechanisms between the two conditions—including atherosclerosis, chronic inflammation, thrombosis, and oxygen supply–demand imbalance.[4] Therefore, identifying predictors of all-cause mortality in patients with CCI is critical for guiding targeted interventions and optimizing clinical management.

Machine learning (ML) techniques have been increasingly applied across various medical domains, including diagnosis, prognosis, and medical imaging.[9,10] Although several ML models have been developed to predict mortality in CHD or IS individually,[11,12] few have examined all-cause mortality in CCI. Tree-based ensemble models offer high predictive accuracy and can rank feature importance, whereas logistic regression (LR) provides interpretable associations between predictors and outcomes. However, the "black box" nature of many ML algorithms limits their clinical applicability.[13]

To address this challenge, interpretable ML (IML) methods—such as SHapley Additive exPlanations (SHAP)—have been developed to provide intuitive, individualized explanations for model predictions.[14,15] In this study, we developed an IML model to predict all-cause mortality in patients with CCI, compared its performance with that of traditional methods, and applied SHAP to enhance model interpretability. This approach supports clinical decision-making by identifying key risk factors and offering transparent, patient-specific prediction insights.

2. Methods

2.1. *Study population*

This was a retrospective, multicenter study designed to predict all-cause mortality in patients with CCI. Data for this study were collected from four major medical centers in China: the Second Affiliated Hospital of Chongqing Medical University, the Third Affiliated

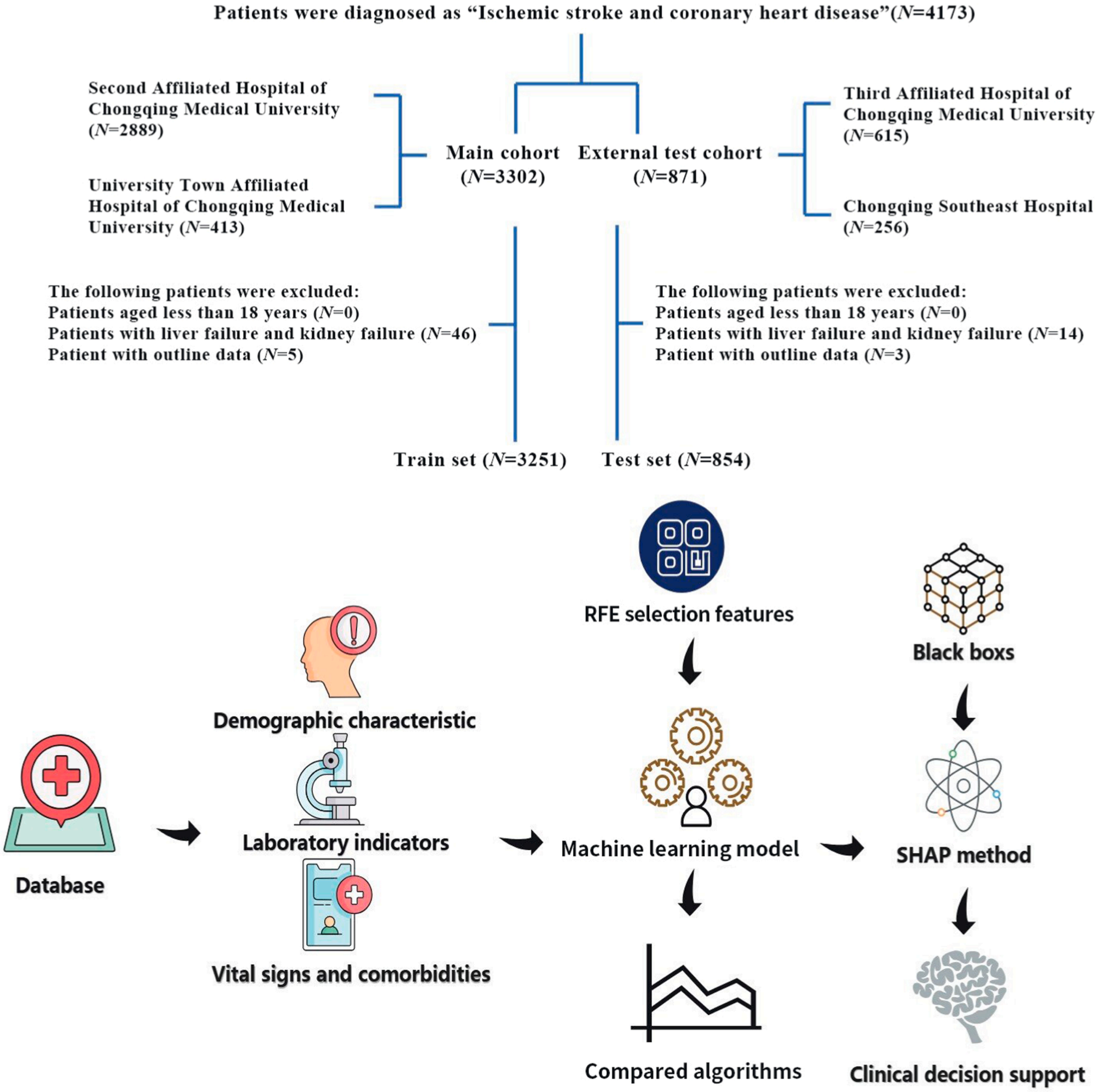


Fig. 1. Flowchart of patient selection. A total of 4105 patients with CHD and IS (CCI) diagnoses were included. CCI: Cardio-cerebral infarction; CHD: Coronary heart disease; IS: Ischemic stroke; RFE: Recursive feature elimination; SHAP: SHapley Additive exPlanations.

Hospital of Chongqing Medical University, University-Town Hospital, and Southeast Hospital. The Second and Third Affiliated Hospitals of Chongqing Medical University are the largest teaching hospitals in the region, whereas University-Town Hospital is an affiliated teaching hospital of Chongqing Medical University. Chongqing Southeast Hospital primarily serves a mixed population from both urban and rural areas. Data collection was conducted from January 2011 to February 2022 in accordance with predefined inclusion and exclusion criteria.

To improve the generalizability of the model, patient populations from the Second Affiliated Hospital of Chongqing Medical University and the University-Town Affiliated Hospital of Chongqing Medical University were assigned to the training set, whereas populations from Chongqing Southeast Hospital and the Third Affiliated Hospital of Chongqing Medical University were assigned to the independent test set. Study data were obtained from the electronic hospital records system and analyzed retrospectively under an ethics waiver granted by the hospital ethics committee. All patients diagnosed with CHD (International Classification of Diseases [ICD]-10: I25.1) were identified, and those with a concomitant diagnosis of IS (ICD-10: I63.0–I63.9) were included. Exclusion criteria were as follows: (i) age younger than 18 years; (ii) missing or incomplete data; and (iii) coexisting liver failure or kidney failure. A flowchart detailing the study selection process is presented in Fig. 1. This study adhered to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis guidelines. This study was approved by the Ethics Committee of Zigong People's Hospital (No. 2024-019), and conducted in accordance with the principles of the *Declaration of Helsinki*. Given its retrospective design, the requirement for informed consent was waived by the institutional review board, and all patient identifiers were removed to ensure data confidentiality.

2.2. Data collection

Following the variable selection approach proposed by Deshmukh *et al*,[16] 44 potential outcome-related variables were identified. Extracted data included patients' basic demographic information as well as other clinically relevant characteristics [Supplementary Material]. A modified k-nearest neighbor (KNN)-based classification technique was applied to impute missing values, thereby mitigating the negative impact of missing data on model classification performance.

2.3. Study outcomes

The primary outcome of interest was all-cause in-hospital mortality, defined as death from any cause during hospitalization.

2.4. Data preprocessing and feature selection

Using the DMwR2 R package, a KNN algorithm was applied to impute missing data for features with less than 30 % missing values. The KNN algorithm, a nonparametric ML approach, estimates missing values based on neighboring observations.[17,18] Outliers were addressed using the capping method: for each variable, values exceeding the 99th percentile were replaced with the 99th percentile, and those below the 1st percentile with the 1st percentile. Min–Max normalization was applied to account for the wide range of feature scales, and one-hot encoding was used to process categorical variables. Data from all centers were normalized uniformly to ensure consistency, as some of the applied algorithms required normalized quantitative data. After preliminary univariate screening, the final feature set was selected using the recursive feature elimination (RFE) method, which was based on a Random Forest (RF) algorithm combined with five-fold cross-validation (CV). RFE iteratively constructs a model, ranks features by importance, removes the least informative feature, and repeats the process until all features are ranked. This process yielded eight key variables: Serum potassium (K), N-terminal pro-B-type natriuretic peptide (NT-proBNP), cholinesterase, age, platelet count (PLT), troponin T, albumin (ALB), and creatinine.

2.5. Model development

ML models were developed using the caret and catboost packages. The original dataset included demographic variables (e.g., gender, age), vital signs, laboratory test results, and comorbidities. After preprocessing, 43 features were retained for analysis. Nine widely used algorithms—gradient boosting machine (GBM), LR, multilayer perceptron (MLP), naïve Bayes (NB), neural network (NN), RF, support vector machine (SVM), extreme gradient boosting (XGBoost), and categorical boosting (CatBoost)—were implemented to predict all-cause in-hospital mortality in patients with CCI. To enhance model stability, all continuous variables were standardized to have a mean of zero and a standard deviation of one. Hyperparameter tuning for each algorithm was performed using five-fold CV. Candidate models were evaluated based on receiver operating characteristic (ROC) curve performance or accuracy, and the optimal model for each algorithm was selected. The test set was excluded from model training and hyperparameter optimization and was used exclusively for final performance evaluation after model development.

2.6. Model interpretation and feature importance

ML models are often regarded as "black boxes" because the reasoning behind their predictions can be difficult to interpret for specific patient populations. To address this limitation, we incorporated SHAP, a framework introduced by Lundberg and Lee[19] for interpreting complex ML models. SHAP has been validated in prior studies as a theoretically rigorous approach that provides both local and global interpretability.[14] SHAP values decompose each model prediction into the contributions of individual features, quantifying the extent to which each variable affects the outcome in a consistent and interpretable manner. This approach facilitates the identification of the most influential predictors and clarifies their impact on the model's decision-making process, thereby enhancing transparency and supporting clinical trust. In this study, interpretability analysis was performed using the shapviz package. SHAP was applied to the final prediction model to identify key risk factors associated with in-hospital mortality in patients with CCI. Each variable was assigned a contribution value representing its relative importance in the final classification. The resulting feature-importance ranking provided a comprehensive understanding of variable relevance and mapped their influence on the model's predictive outcomes.

2.7. Statistical analysis

A total of 4105 patients were randomly divided into training and test cohorts in an 8:2 ratio using stratified sampling based on endpoint occurrence. Categorical variables were reported as frequencies with percentages, whereas continuous variables were summarized as mean ± standard deviation (SD) or median with interquartile range (IQR), depending on data distribution. Between-group differences were assessed using Pearson's χ^2 test or Fisher's exact test for categorical variables and one-way analysis of variance (ANOVA) or the Wilcoxon rank-sum test for continuous variables, as appropriate. Nine ML algorithms were trained on the training set and subsequently validated on the test set. Model discrimination in the test cohort was evaluated using sensitivity, specificity, F1 score, and the area under the ROC curve. Among the nine algorithms, the RF model demonstrated the highest predictive performance and was selected as the final model. Model calibration was evaluated using the Brier score.[20] Performance comparisons across the nine ML models were further examined using one-way ANOVA. To enhance interpretability, the SHAP approach was applied to the final model. All analyses were conducted using R software (version 4.1.2, https://www.r-project.org/).

Table 1
Baseline characteristics, laboratory parameters, vital signs, and statistical results of patients with CHD and IS (CCI).

Variables	Total ($n = 4105$)	Test set ($n = 854$)	Train set ($n = 3251$)	P-value
Age (years)	79 (72, 85)	80 (72, 85)	79 (72, 85)	0.314
Sex				0.857
Female	2071 (50)	428 (50)	1643 (51)	
Male	2034 (50)	426 (50)	1608 (49)	
Smoking history				<0.001
No	2994 (73)	568 (67)	2426 (75)	
Yes	1111 (27)	286 (33)	825 (25)	
Alcohol history				<0.001
No	2795 (68)	619 (72)	2176 (67)	
Yes	1310 (32)	235 (28)	1075 (33)	
Hypertension				0.818
No	789 (19)	167 (20)	622 (19)	
Yes	3316 (81)	687 (80)	2629 (81)	
Hyperlipidemia				0.804
No	3481 (85)	727 (85)	2754 (85)	
Yes	624 (15)	127 (15)	497 (15)	
AF				0.441
No	3051 (74)	644 (75)	2407 (74)	
Yes	1054 (26)	210 (25)	844 (26)	
Diabetes				0.598
No	2510 (61)	515 (60)	1995 (61)	
Yes	1595 (39)	339 (40)	1256 (39)	
COPD				0.291
No	4063 (99)	842 (99)	3221 (99)	
Yes	42 (1)	12 (1)	30 (1)	
SBP (mmHg)	136 (124, 151)	136 (123, 149)	137 (124, 151)	0.133
DBP (mmHg)	77 (69, 85)	76 (69, 84)	77 (70, 85)	0.062
NT-proBNP (pg/mL)	876.7 (271.4, 2791.4)	900.5 (329.0, 2853.8)	872.6 (260.3, 2773.0)	0.091
CRP (mg/L)	6.28 (5.00, 15.90)	5.99 (5.00, 13.49)	6.37 (5.00, 16.56)	0.906
WBC ($\times 10^{9}$/L)	6.84 (5.64, 8.47)	6.70 (5.56, 8.34)	6.87 (5.66, 8.50)	0.085
RBC ($\times 10^{12}$/L)	4.13 (3.75, 4.49)	4.10 (3.76, 4.46)	4.14 (3.75, 4.50)	0.131
HCT (%)	37.7 (34.2, 40.7)	37.4 (34.2, 40.2)	37.8 (34.2, 41.0)	0.105
NEU (%)	71.6 (65.0, 78.8)	71.2 (64.8, 78.4)	71.7 (65.1, 78.9)	0.364
LYM (%)	19.6 (13.5, 25.4)	19.85 (13.3, 25.6)	19.5 (13.6, 25.3)	0.792
PLT ($\times 10^{9}$/L)	175.1 (143.3, 210.6)	175.3 (143.2, 209.0)	175 0.0 (143.6, 210.9)	0.987
HB (g/L)	125 (113, 135)	124 (113, 134)	125 (113, 136)	0.171
ALT (U/L)	16 (11, 24)	17 (11, 25)	16 (11, 24)	0.087
AST (U/L)	21 (17, 28)	21 (17, 29)	21 (17, 28)	0.109
GGT (U/L)	27.9 (19.0, 45.6)	28.8 (19.0, 45.6)	27.2 (19.0, 45.6)	0.434
Cholinesterase (kU/L)	6.51 (5.28, 7.68)	6.46 (5.14, 7.52)	6.52 (5.30, 7.71)	0.127
ALB (g/L)	38.3 (35.5, 40.8)	38.2 (35.3, 40.8)	38.3 (35.6, 40.8)	0.312
Urea (mmol/L)	6.96 (5.58, 9.04)	7.04 (5.67, 9.15)	6.94 (5.57, 9.01)	0.177
Creatinine (μmol/L)	84.2 (67.7, 108.1)	84.9 (68.4, 110.7)	83.8 (67.5, 107.6)	0.253
Uric acid (μmol/L)	363.2 (300.0, 433.8)	364.1 (294.8, 429.8)	363.0 (300.8, 434.5)	0.606
TG (mmol/L)	1.18 (0.91, 1.61)	1.21 (0.93, 1.60)	1.17 (0.91, 1.61)	0.355
TC (mmol/L)	3.93 (3.35, 4.63)	3.91 (3.36, 4.56)	3.94 (3.34, 4.64)	0.489
LDL-C (mmol/L)	2.10 (1.65, 2.66)	2.04 (1.64, 2.63)	2.11 (1.66, 2.67)	0.176
PT (s)	13.3 (12.7, 14.2)	13.4 (12.6, 14.3)	13.3 (12.7, 14.2)	0.555
INR	1.04 (0.98, 1.12)	1.04 (0.98, 1.13)	1.04 (0.98, 1.12)	0.437
TT (s)	17.1 (16.0, 18.0)	17.1 (16.1, 18.0)	17.1 (16.0, 17.9)	0.263
APTT (s)	35.3 (32.0, 38.7)	35.6 (31.7, 38.8)	35.3 (32.1, 38.6)	0.977
Fibrinogen (g/L)	3.56 (3.00, 4.18)	3.56 (2.97, 4.20)	3.56 (3.01, 4.18)	0.962
Troponin T (ng/mL)	0.02 (0.01, 0.04)	0.02 (0.01, 0.03)	0.02 (0.01, 0.04)	0.859
Sodium (mmol/L)	140.4 (138.4, 142.0)	140.4 (138.4, 142.0)	140.4 (138.4, 142.1)	0.884
Potassium (mmol/L)	4.01 (3.81, 4.27)	4.06 (3.85, 4.30)	3.99 (3.80, 4.26)	<0.001
Calcium (mmol/L)	2.23 (2.14, 2.30)	2.22 (2.14, 2.30)	2.23 (2.14, 2.31)	0.292
Phosphorus (mmol/L)	1.09 (0.98, 1.21)	1.09 (0.96, 1.22)	1.09 (0.98, 1.20)	0.858
Chlorine (mmol/L)	103.0 (100.5, 105.5)	103.0 (100.4, 105.4)	103.0 (100.5, 105.6)	0.689
Antiplatelet				0.789
No	1165 (28)	246 (29)	919 (28)	
Yes	2940 (72)	608 (71)	2332 (72)	
Lipid-lowering agents				0.952
No	1246 (30)	258 (30)	988 (30)	
Yes	2859 (70)	596 (70)	2263 (70)	
Death				0.450
No	3934 (96)	814 (95)	3120 (96)	
Yes	171 (4)	40 (5)	131 (4)	

Data are presented as n (%) or median (Q1, Q3). ALB: Albumin; ALT: Alanine aminotransferase; APTT: Activated partial thromboplastin time; AST: Aspartate aminotransferase; AF: Atrial fibrillation; CHD: Coronary heart disease; COPD: Chronic obstructive pulmonary disease; CRP: C-reactive protein; CCI: Cardio-cerebral infarction; DBP: Diastolic blood pressure; GGT: Gamma-glutamyl transferase; HB: Hemoglobin; HCT: Hematocrit; INR: International normalized ratio; IS: Ischemic stroke; LDL-C: Low-density lipoprotein cholesterol; LYM: Lymphocytes; NEU: Neutrophils; NT-proBNP: N-terminal pro-B-type natriuretic peptide; PLT: Platelet count; PT: Prothrombin time; RBC: Red blood cells; SBP: Systolic blood pressure; TC: Total cholesterol; TG: Triglycerides; TT: Thrombin time; WBC: White blood cells.

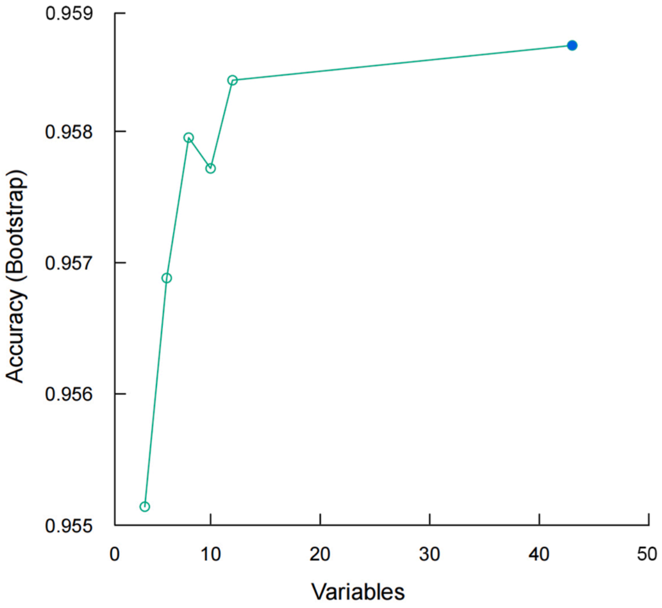


Fig. 2. Results of feature screening by RFE with 5-fold CV. CV: Cross-validation; RFE: Recursive feature elimination.

3. Results

3.1. Baseline characteristics

A total of 4173 patients with CCI were initially screened. Sixty-eight patients were excluded because they had insufficient demographic or clinical information, or had liver failure, kidney failure, or multi-organ dysfunction syndrome. Ultimately, 4105 patients with CCI were included in this study. The median age was 79 years (IQR: 72, 85), and 50 % were male. Among these patients, 171 died and 3934 survived their hospitalization. Additionally, 3481 patients had concomitant hyperlipidemia, accounting for 85 % of the cohort. Of the total dataset, 80 % ($n = 3251$) were used for model training and 20 % ($n = 854$) for model validation. Table 1 presents the clinicopathological characteristics of the study population.

3.2. Recursive feature elimination-selected variables

RFE with five-fold CV identified the top 8 most important variables from the initial 43 features: K, NT-proBNP, cholinesterase, age, PLT, troponin T, ALB, and creatinine [Fig. 2].

3.3. Model performance comparison

The performance of the nine ML models in the training and test sets is summarized in Tables 2 and 3. The RF model demonstrated the highest performance across both datasets. The RF model achieved the highest AUC and specificity in the training and test sets, respectively [Fig. 3]. In addition, the RF model yielded the lowest Brier score among the nine models, with values of 0.007 and 0.039 in the training and test sets, respectively. Based on these results, the RF model was selected as the final prediction model.

3.4. Visualization of feature importance

After recursive feature elimination, the selected variables were used to construct the RF model, and SHAP values were applied to compare the contributions of these variables to outcome events. The SHAP method enables both individualized patient-level explanations and global explanations across the entire validation cohort. Consequently, it is widely applied to visualize the effects of selected variables on mortality in prediction models. To generate global SHAP explanations, the SHAP values for each feature were averaged across all patients in the cohort. The clinical significance of a feature in model prediction increases with its mean absolute Shapley value. Fig. 4A illustrates the top eight risk factors ranked by mean absolute SHAP score, while Fig. 4B shows the eight most influential features of the model. In these plots, the y-axis represents feature ranking (relative importance), and the x-axis represents SHAP values (feature contributions to the predicted outcome). For each feature, individual patient contributions are visualized as colored dots: yellow indicates higher feature values associated with increased mortality risk, whereas purple indicates lower feature values associated with reduced risk. Interpretation of the SHAP plots demonstrated that an elevated predicted probability of all-cause in-hospital mortality was associated with older age, higher K, NT-proBNP, troponin T, and creatinine levels, as well as lower ALB and cholinesterase concentrations.

Table 2
Performance of nine models in the training set.

Model	F1-score	Accuracy	Sensitivity	Specificity	PPV	NPV
GBM	0.62	0.98	0.46	0.99	0.95	0.98
LR	0.06	0.96	0.03	0.99	0.33	0.96
MLP	0.31	0.96	0.20	0.99	0.74	0.97
NB	0.33	0.93	0.40	0.96	0.28	0.97
NN	–	0.96	0.00	1.00	0.00	0.96
RF	0.98	0.99	0.95	0.99	0.99	0.99
SVM	0.03	0.96	0.02	1.00	1.00	0.96
XGBoost	0.10	0.96	0.05	0.99	0.88	0.96
CatBoost	0.43	0.97	0.29	0.99	0.86	0.97

CatBoost: Categorical boosting; GBM: Gradient boosting machine; LR: Logistic regression; MLP: Multilayer perceptron; NB: Naïve Bayes; NN: Neural network; NPV: Negative predictive value; PPV: Positive predictive value; RF: Random Forest; SVM: Support vector machine; XGBoost: Extreme gradient boosting. –: NaN values, which were not calculated, likely due to an extremely unbalanced data structure with very few positive cases.

Table 3
Performance of nine models in the test set.

Model	F1-score	Accuracy	Sensitivity	Specificity	PPV	NPV
GBM	0.12	0.95	0.08	0.99	0.30	0.96
LR	0.04	0.95	0.03	0.99	0.20	0.95
MLP	0.08	0.94	0.05	0.99	0.17	0.95
NB	0.25	0.88	0.43	0.90	0.17	0.97
NN	–	0.95	0.00	1.00	–	0.95
RF	0.10	0.96	0.05	0.99	0.99	0.96
SVM	–	0.95	–	0.99	–	0.95
XGBoost	0.05	0.95	0.03	0.99	0.50	0.95
CatBoost	0.08	0.94	0.05	0.99	0.18	0.95

CatBoost: Categorical boosting; GBM: Gradient boosting machine; LR: Logistic regression; MLP: Multilayer perceptron; NB: Naïve Bayes; NN: Neural network; NPV: Negative predictive value; PPV: Positive predictive value; RF: Random Forest; SVM: Support vector machine; XGBoost: Extreme gradient boosting. –: NaN values, which were not calculated, likely due to an extremely unbalanced data structure with very few positive cases.

The relationship between feature values and SHAP values is illustrated in greater detail for patient age, K concentration, NT-proBNP, troponin T, cholinesterase, and ALB concentration in Fig. 5. These results indicate that higher concentrations of K, NT-proBNP, and troponin T, as well as older age and lower levels of cholinesterase and ALB, are significantly associated with increased mortality risk. Fig. 6A shows that both hyperkalemia and hypokalemia are associated with elevated SHAP values, indicating a higher model-predicted mortality risk. Fig. 6B demonstrates a positive association between age and SHAP values, with the effect most pronounced in very elderly patients, underscoring age as a major predictor of mortality. Fig. 6C reveals that low ALB levels are associated with elevated SHAP values, suggesting that poor nutritional

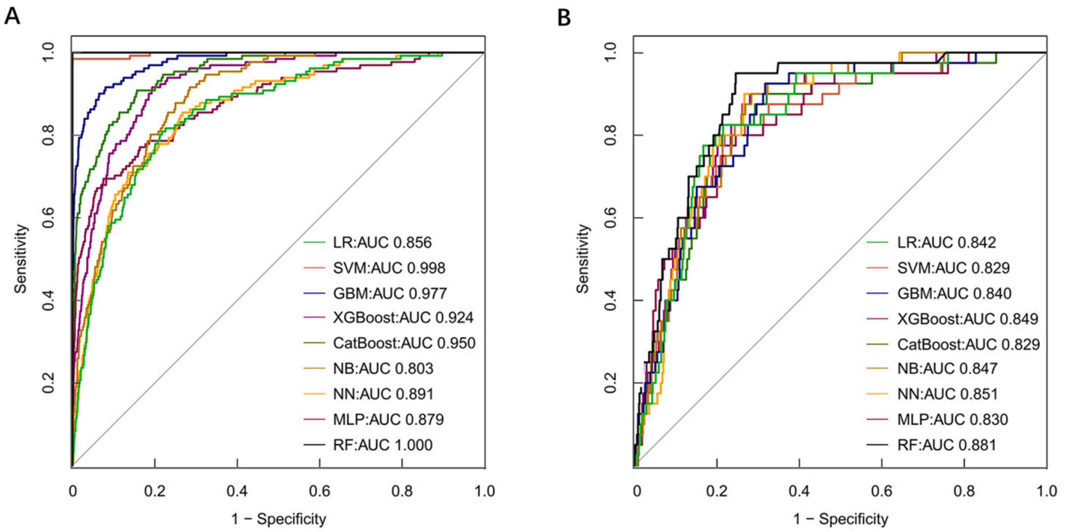


Fig. 3. ROC curves for nine ML models predicting all-cause mortality in patients with CCI. ROC curves are shown for the training set (A) and test set (B). AUC: Area under the receiver operating characteristic curve; CCI: Cardio-cerebral infarction; CatBoost: Categorical boosting; GBM: Gradient boosting machine; LR: Logistic regression; ML: Machine learning; MLP: Multilayer perceptron; NB: Naïve Bayes; NN: Neural network; RF: Random Forest; ROC: Receiver operating characteristic; SVM: Support vector machine; XGBoost: Extreme gradient boosting.

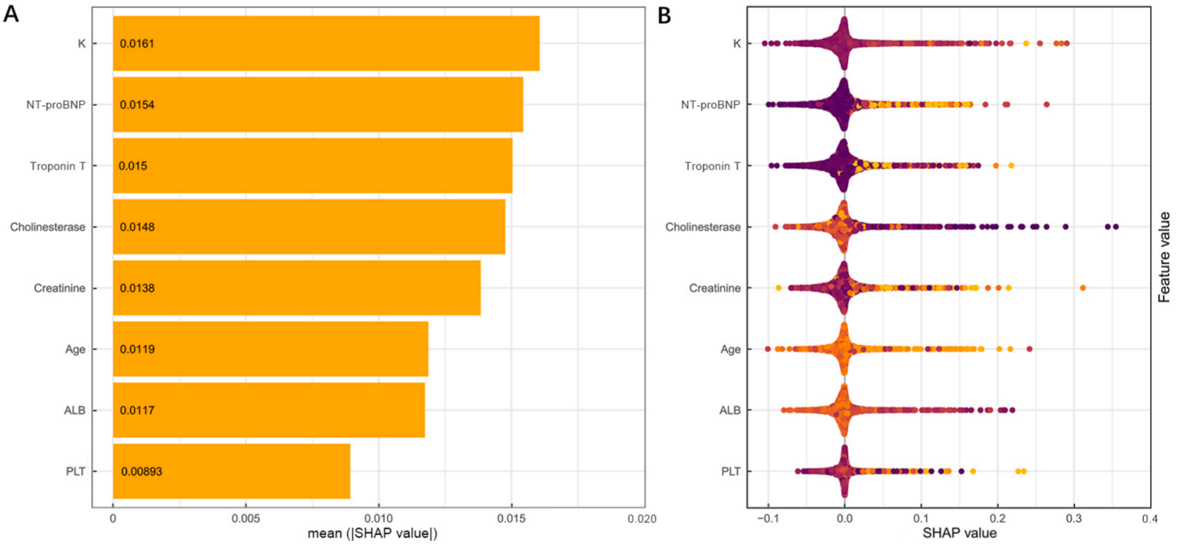


Fig. 4. Model interpretation: (A) ranking of the top eight variables by mean SHAP value; (B) ranking of the top eight risk factors using the optimal model. A higher SHAP value indicates a higher the risk of death. Yellow indicates higher feature values, and purple indicates lower feature values. ALB: Albumin; K: Potassium; NT-proBNP: N-terminal pro-B-type natriuretic peptide; PLT: Platelet count; SHAP: SHapley Additive exPlanations.

status may substantially contribute to mortality risk. Similarly, increased creatinine levels are linked to higher SHAP values, reflecting the model’s sensitivity to renal dysfunction.

3.5. *Interpretation of personalized predictions*

The SHAP force plot was used to visualize the Shapley value for each feature as a force that either increases (positive value) or decreases (negative value) the prediction from its baseline.[14] This approach clarifies the model’s predictions for individual patients by showing the relative contribution of each feature to the overall prediction. In addition, a novel visualization technique was employed to facilitate the interpretation of results. The interpretable RF model was used to analyze the SHAP values corresponding to K, NT-proBNP, troponin T, cholinesterase, creatinine, age, ALB, and PLT for individual patients in order to estimate their predicted probability of mortality. To demonstrate the ease of interpretation, two representative cases were examined: an 82-year-old man who passed away during hospitalization and an 81-year-old woman who survived until discharge [Fig. 7]. The direction of the arrows indicates the extent to which each feature contributes to the prediction. Yellow arrows indicate factors that increase the predicted probability of mortality, whereas red arrows indicate factors that decrease it. The final SHAP value, which corresponds to the prediction score, was calculated as the sum of contributions from all features added to the model baseline: Model prediction = Baseline + Σ (contributions from all features). Each SHAP value reflects both the direction and the magnitude of a feature’s influence on the prediction. The average predicted probability across the entire cohort was 0.50. For the representative male patient, the SHAP value was 0.655, and the predicted probability was 0.658—higher than the cohort mean—suggesting that this patient was at high risk, which was consistent with the observed outcome. In contrast, the representative female patient had a SHAP

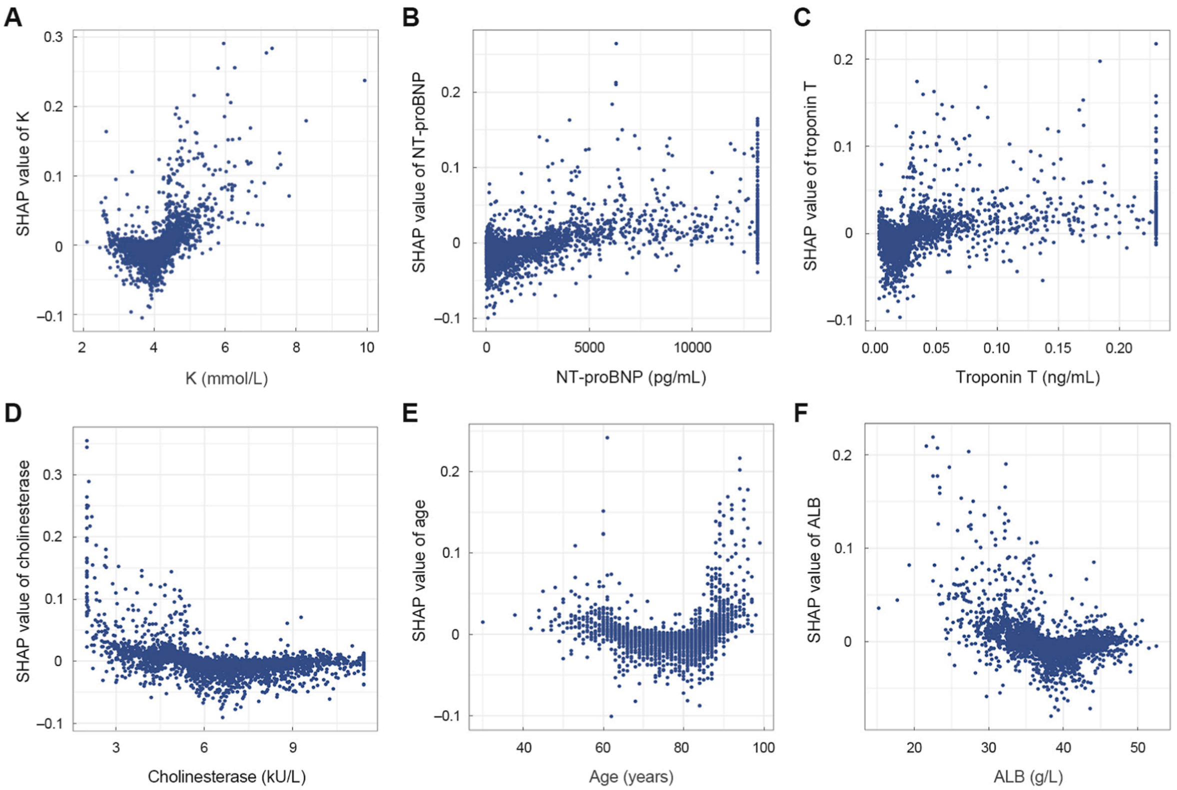


Fig. 5. Scatter plots showing the relationship between feature values and SHAP values. ALB: Albumin; K: Potassium; NT-proBNP: N-terminal pro-B-type natriuretic peptide; SHAP: SHapley Additive exPlanations.

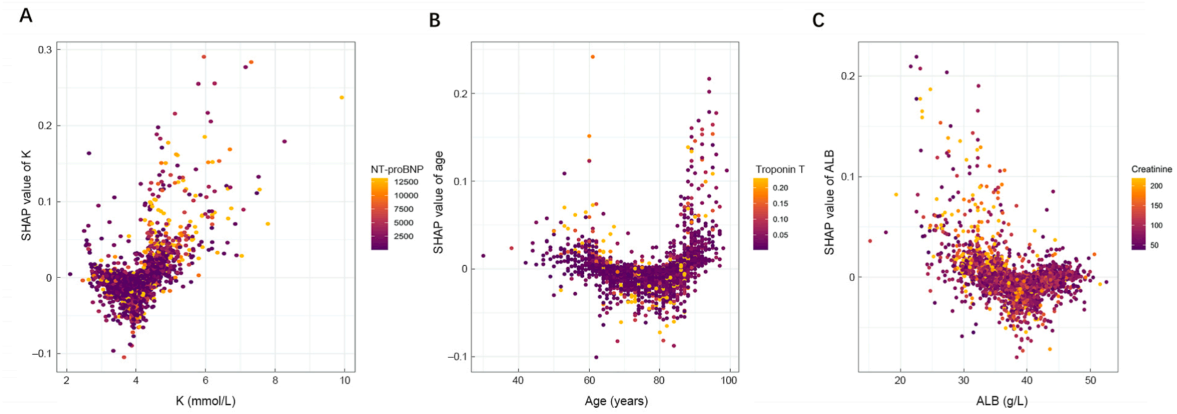


Fig. 6. SHAP interaction plots for all-cause mortality. (A) K *vs.* NT-proBNP. (B) Age *vs.* troponin T. (C) ALB *vs.* creatinine. ALB: Albumin; K: Potassium; NT-proBNP: N-terminal pro-B-type natriuretic peptide; SHAP: SHapley Additive exPlanations.

value of 0.0367 and a predicted probability of 0.491—lower than the cohort mean—indicating a lower mortality risk, again consistent with the clinical outcome.

3.6. *Clinical practicability of the Random Forest prediction model*

The decision curve analysis (DCA) curves for the RF model demonstrated excellent positive net benefits in both the training [Fig. 8A] and test [Fig. 8B] cohorts, supporting the model's clinical applicability.

4. Discussion

In this study, we developed and validated an interpretable ML model to predict all-cause in-hospital mortality in patients diagnosed with CCI. Among the nine ML classifiers evaluated, the RF model demonstrated the best overall performance and was selected to construct the ML-based risk score. With an average AUC exceeding 0.80, the ML risk score exhibited a substantial advantage over currently available risk prediction tools. These findings suggest that ML-based approaches hold promise for integration into clinical practice to enhance mortality risk assessment. Furthermore, by incorporating SHAP values and SHAP plots, we demonstrated that the ML model could identify and rank the most influential features, thereby providing a highly accurate and interpretable mortality prediction model for patients with CCI. The visual representation of feature importance and the cumulative ranking of clinically relevant variables are particularly valuable for clinicians, as they highlight the key contributors to model predictions and support informed decision-making. Our report offers insight into which features

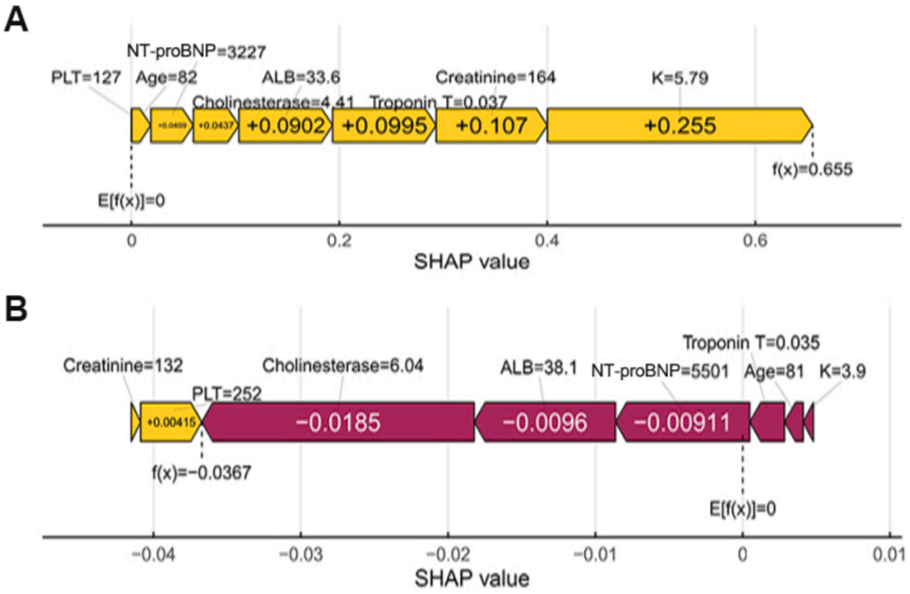


Fig. 7. Interpretation of model prediction results for two representative patients. (A) An 82-year-old male patient. (B) An 81-year-old female patient. ALB: Albumin; K: Potassium; NT-proBNP: N-terminal pro-B-type natriuretic peptide; PLT: Platelet count; SHAP: SHapley Additive exPlanations.

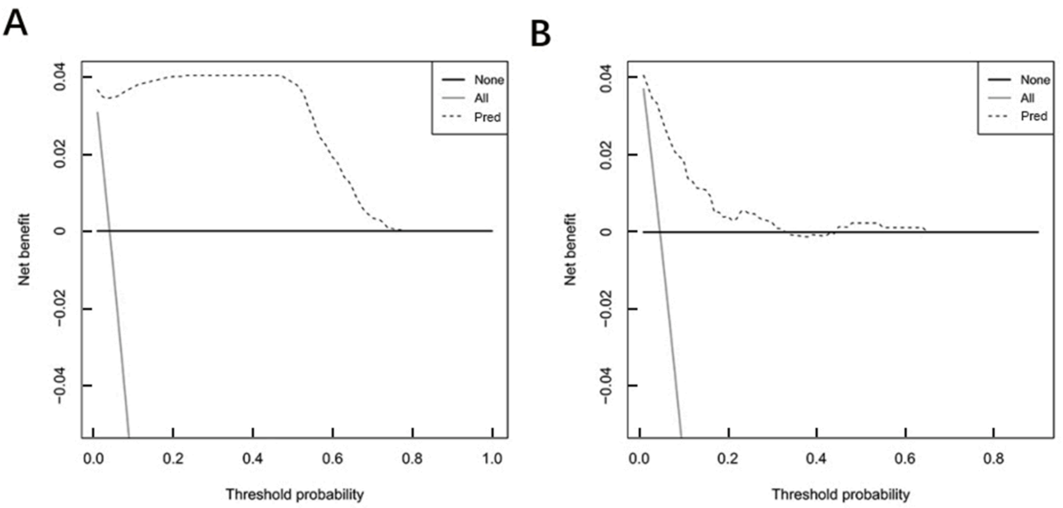


Fig. 8. Decision curves of the RF model predicting all-cause mortality in training set (A) and test set (B). The y-axis shows net benefit, while the x-axis shows the threshold probability. The gray line represents the strategy where all patients die, and the black line represents strategy where no patients die. When the threshold probability is lower than 20 %, the model provides greater net benefit. RF: Random Forest.

carry the greatest weight in the RF model's predictions, thereby improving transparency and fostering clinical trust.

ML algorithms have emerged as valuable tools in the era of electronic health records for detecting life-threatening conditions that may be overlooked by clinicians. ML, a subset of artificial intelligence, differs from regression-based methods because it does not make a priori assumptions about causal relationships between variables. ML has been extensively applied to the diagnosis and treatment planning of coronary artery disease (CAD) and IS.[21–23] These algorithms are capable of performing complex linear and nonlinear data analyses, uncovering patterns that may be missed by traditional regression techniques.[24] However, no prior studies have developed an ML-based prediction model specifically designed to predict all-cause mortality in patients with CCI. In this study, we constructed nine ML models to predict all-cause mortality in patients with CCI. The RF model achieved the highest AUC (0.881) in the test cohort, outperforming all other models. The inclusion of quantifiable clinical features likely contributed to the high evaluation metrics obtained. Among all features, K, NT-proBNP, cholinesterase, age, PLT, troponin T, ALB, and creatinine were particularly important for predicting outcomes.

Previous ML models have successfully incorporated feature selection and captured complex relationships among high-dimensional data.[14] However, when confronted with a large number of highly correlated features, such models are prone to overfitting, which reduces generalizability and makes predictions less stable due to noise in the data. In this study, we addressed this limitation by applying iterative stepwise regression to identify independent variables and using RFE to select the most important predictors. This approach enhanced both the accuracy and stability of the model's predictions.

Correctly interpreting ML prediction models, visually presenting predicted outcomes to clinicians, and demonstrating how individual features influence the target outcome (in this case, mortality) remain ongoing challenges. To address this, we applied the SHAP method to the RF model to determine whether each feature exerted a positive or negative effect on patient outcomes, while simultaneously optimizing predictive power and interpretability. The extent to which humans can understand the reasoning behind a ML model's prediction is referred to as the model's "explainability" or "interpretability."[25] When a predictive model offers a high degree of interpretability, clinicians can better understand the rationale underlying specific predictions, thereby enhancing their ability to make evidence-based clinical decisions in patients' best interests. By contrast, "black-box" models such as XGBoost often lack transparency, making it difficult to determine which patient features contribute to the predicted outcome. In comparison, interpretable models—such as sparse linear models—allow direct inference of feature contributions through their coefficients, assuming other features are held constant. Previous research applying ML algorithms in critical care has primarily focused on identifying risk factors and predicting outcomes, but often without sufficient emphasis on interpretability for clinical application. However, SHAP extends this approach by quantifying the contribution of each feature—including interactions between features—while providing consistent and locally accurate attributions within the prediction model. Nonetheless, predictive models often face challenges in generalizability due to variations in patient characteristics and clinical settings, and they may offer limited guidance for individualized treatment. In this study, we not only developed a predictive model by identifying key prognostic factors but also enhanced its interpretability. By applying SHAP, we demonstrated that each feature contributes uniquely to the model's output and that these contributions can be visualized [Fig. 4]. Furthermore, we quantified the magnitude of these contributions, allowing for a deeper understanding of their clinical implications [Fig. 5]. Although other interpretability methods—such as Local Interpretable Model-agnostic Explanations (LIME)[26]—are available, SHAP was selected because it offers strong theoretical guarantees of consistency and local accuracy. Based on cooperative game theory, SHAP uses an additive feature attribution framework that clarifies how each individual sample contributes to the overall prediction results [Fig. 4]. Notably, an interaction was observed between serum K and NT-proBNP: as NT-proBNP levels increased, the SHAP value corresponding to K also rose [Fig. 6]. This finding indicates that the predictive contribution of K is influenced by NT-proBNP levels, highlighting the synergistic impact of electrolyte imbalance and cardiac biomarkers on mortality risk assessment. Additionally, an interaction between age and troponin T was observed: in younger patients, elevated troponin T levels corresponded to higher SHAP values for age, suggesting that myocardial injury in younger individuals substantially increases mortality risk [Fig. 6]. This finding underscores the importance of vigilant clinical monitoring of high-risk biomarkers even in younger populations. Collectively, these results emphasize the critical roles of nutritional and renal status—as well as key cardiac and electrolyte biomarkers—in predicting patient outcomes [Fig. 6].

Conventional statistical methods, such as multivariable LR, estimate whether risk increases or decreases by calculating odds ratios and confidence intervals, thereby illustrating the effects of individual predictors. However, these approaches cannot fully capture complex quantitative relationships between features and outcomes, as associations may be nonlinear and individual features may contribute unequally to overall risk. Our findings are consistent with previous studies in patients with CAD and IS, which reported that reduced cholinesterase activity and elevated serum K levels were associated with worse prognoses.[27,28] Importantly, our model not only confirmed these associations but also visualized how risk changes continuously with increasing or decreasing variable values. Clinically, this interpretable framework allows physicians to easily evaluate model outputs and refine risk assessments as new patient data become available. Moreover, patients whose characteristics approach specific thresholds or "trigger points" should receive closer monitoring. This study further highlights that some parameters may indicate rising risk even when they remain within normal reference ranges. Identifying such "warning thresholds" could facilitate earlier preventive interventions and improve patient management.

Age, NT-proBNP concentration, ALB, PLT, troponin T, and creatinine level remain important predictors of mortality in patients with CCI. Our analysis provides more direct conclusions and a method of visually representing the relationships that generate predicted outcomes, which aligns with previous literature and clinical experience.[29–39] Notably, the differences in clinical characteristics between participants who died and those who survived closely correspond to our ranking of variable importance. Patients who did not survive hospitalization were generally older, had higher concentrations of NT-proBNP, troponin T, and creatinine, and lower concentrations of cholinesterase and ALB. According to the ML model, all these variables hold significant predictive importance. Clinicians should pay careful attention to these features when assessing the mortality risk of patients with CCI. Of particular note, K was ranked as the most influential predictor, highlighting its critical role in mortality risk stratification in this population. However, certain potential confounders—such as the stress hyperglycemia ratio (SHR) and inflammatory markers—were not included in this analysis. Lei *et al*[40] reported that elevated SHR ($>$1.355) was independently associated with increased mortality in critically ill cardiovascular patients, even after propensity score matching and multivariable adjustment.

Several limitations should be acknowledged. First, this study did not compare the proposed models with established and validated clinical risk calculators, as our primary objective was to evaluate the predictive performance of widely applied ML approaches. Second, with the rapid progress of artificial intelligence, deep learning methods are increasingly used in medical prognostic modeling. Future research should incorporate larger-scale, multi-level datasets to develop deep learning models for predicting the prognosis of CCI. Third, only structured data were analyzed in this study. Additional efforts are needed to extract and integrate unstructured information—including clinical notes, imaging biomarkers, environmental exposures, lifestyle factors, and other relevant variables—to enhance model performance and improve generalizability.

In this study, we developed and validated an ML-based risk stratification model that demonstrated high accuracy in predicting all-cause in-hospital mortality among patients with CCI. By incorporating SHAP analysis, the proposed model provided transparent, patient-specific explanations for risk prediction, enabling clinicians to better understand the contribution of critical features and thereby enhancing interpretability and clinical trust. Nevertheless, prospective validation and external testing are necessary to confirm the generalizability and clinical applicability of our findings. This individualized, interpretable prediction framework has the potential to be extended to risk assessment across a variety of disease contexts. Ultimately, the integration of advanced ML techniques with routinely collected clinical data may serve as a powerful tool to support clinical decision-making and optimize treatment strategies for patients with CCI.

Ethics statement

This study was approved by the Ethics Committee of Zigong People's Hospital (No. 2024-019), and conducted in accordance with the principles of the *Declaration of Helsinki*. Given its retrospective design, the requirement for informed consent was waived by the institutional review board, and all patient identifiers were removed to ensure data confidentiality.

Author contribution

Jian Huang and Xiaozhu Liu conceptualized the study. Jian Huang conducted the investigation and performed data analysis, while Xiaozhu Liu was responsible for data extraction. Shengxian Peng and Jian Huang drafted the initial version of the manuscript. Qianqian Wang, Yalin Dong, Tianyu Xiang, and Yang Zhang contributed to manuscript editing and critical revision. All authors reviewed and approved the final manuscript.

Funding

This work was supported by the Key Special Project for Technological Innovation and Application Development of the Chongqing Science and Technology Bureau (No. cstb2022tiad-kpx0169).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bnd.2025.12.003.

References

1. Wang Z, Ma Z, Xu S, et al. Expert consensus on the management of coronary heart disease with ischemic stroke in community. *Cardiovasc Dis Control*. 2022;20: 772–793. https://doi.org/10.3969/j.issn.1009-816x.2022.04.001.
2. World Health Organization. *The Global Burden of Disease: 2004 Update*. Geneva, Switzerland: World Health Organization; 2008. Available from: http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf. Accessed April 30, 2015.
3. Gersh BJ, Sliwa K, Mayosi BM, Yusuf S. Novel therapeutic concepts: the epidemic of cardiovascular disease in the developing world: global implications. *Eur Heart J*. 2010;31:642–648. https://doi.org/10.1093/eurheartj/ehq030.
4. An X, Liu Z, Zhang L, et al. Co-occurrence patterns and related risk factors of ischaemic heart disease and ischaemic stroke across 203 countries and territories: a spatial correspondence and systematic analysis. *Lancet Global Health*. 2025;13: e808–e819. https://doi.org/10.1016/S2214-109X(25)00013-0.
5. Rubinstein A, Colantonio L, Bardach A, et al. Estimación de la carga de las enfermedadescardiovascularesatribuible a factores de riesgomodificablesen Argentina [Estimate of the cardiovascular disease burden attributable to modifiable risk factors in Argentina]. *Rev Panamsalud Publica*. 2010;27:237–245. https://doi.org/10.1590/s1020-49892010000400001.
6. Alba AC, Agoritsas T, Jankowski M, et al. Risk prediction models for mortality in ambulatory patients with heart failure: a systematic review. *Circ Heart Fail*. 2013;6: 881–889. https://doi.org/10.1161/CIRCHEARTFAILURE.112.000043.
7. Goyal M, Ospel JM, Kappelhof M, Ganesh A. Challenges of outcome prediction for acute stroke treatment decisions. *Stroke*. 2021;52:1921–1928. https://doi.org/10.1161/STROKEAHA.120.033785.
8. Powers WJ, Rabinstein AA, Ackerson T, et al. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American heart association/American Stroke Association. *Stroke*. 2018;49: e46–e110. https://doi.org/10.1161/STR.0000000000000158.
9. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2:230–243. https://doi.org/10.1136/svn-2017-000101.
10. Van Calster B, Wynants L. Machine learning in medicine. *N Engl J Med*. 2019;380: 2588. https://doi.org/10.1056/NEJMc1906060.
11. Motwani M, Dey D, Berman DS, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J*. 2017;38:500–507. https://doi.org/10.1093/eurheartj/ehw188.
12. Heo J, Yoo J, Lee H, et al. Prediction of hidden coronary artery disease using machine learning in patients with acute ischemic stroke. *Neurology*. 2022;99: e55–e65. https://doi.org/10.1212/WNL.0000000000200576.
13. Zihni E, Madai VI, Livne M, et al. Opening the black box of artificial intelligence for clinical decision support: a study predicting stroke outcome. *PLoS One*. 2020;15: e0231166. https://doi.org/10.1371/journal.pone.0231166.
14. Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng*. 2018;2:749–760. https://doi.org/10.1038/s41551-018-0304-0.
15. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA*. 2017;318:517–518. https://doi.org/10.1001/jama.2017.7797.
16. Deshmukh F, Merchant SS. Explainable machine learning model for predicting GI bleed mortality in the intensive care unit. *Am J Gastroenterol*. 2020;115:1657–1668. https://doi.org/10.14309/ajg.0000000000000632.
17. Gromski PS, Xu Y, Kotze HL, et al. Influence of missing values substitutes on multivariate analysis of metabolomics data. *Metabolites*. 2014;4:433–452. https://doi.org/10.3390/metabo4020433.
18. Shah JS, Rai SN, DeFilippis AP, Hill BG, Bhatnagar A, Brock GN. Distribution based nearest neighbor imputation for truncated high dimensional data with applications to pre-clinical and clinical metabolomics studies. *BMC Bioinf*. 2017;18:114. https://doi.org/10.1186/s12859-017-1547-6.
19. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. New York, NY: Curran Associates Inc; 2017:4766–4777. https://doi.org/10.48550/arXiv.1705.07874.
20. Angraal S, Mortazavi BJ, Gupta A, et al. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. *JACC Heart Fail*. 2020;8:12–21. https://doi.org/10.1016/j.jchf.2019.06.013.
21. Nakanishi R, Slomka PJ, Rios R, et al. Machine learning adds to clinical and CAC assessments in predicting 10-year CHD and CVD deaths. *JACC, Cardiovasc Imaging*. 2021;14:615–625. https://doi.org/10.1016/j.jcmg.2020.08.024.
22. Johnson KM, Johnson HE, Zhao Y, Dowe DA, Staib LH. Scoring of coronary artery disease characteristics on coronary CT angiograms by using machine learning. *Radiology*. 2019;292:354–362. https://doi.org/10.1148/radiol.2019182061.
23. Abdar M, Książek W, Acharya UR, Tan RS, Makarenkov V, Pławiak P. A new machine learning technique for an accurate diagnosis of coronary artery disease. *Comput Methods Progr Biomed*. 2019;179:104992. https://doi.org/10.1016/j.cmpb.2019.104992.
24. Jabal MS, Joly O, Kallmes D, et al. Interpretable machine learning modeling for ischemic stroke outcome prediction. *Front Neurol*. 2022;13:884693. https://doi.org/10.3389/fneur.2022.884693.
25. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019;25:30–36. https://doi.org/10.1038/s41591-018-0307-0.
26. Miller T. Explanation in artificial intelligence: insights from the social sciences. *Artif Intell*. 2019;267:1–38. https://doi.org/10.1016/j.artint.2018.07.007.
27. Zhang Z, Beck MW, Winkler DA, et al. Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. *Ann Transl Med*. 2018;6:216. https://doi.org/10.21037/atm.2018.05.32.
28. Pun PH, Goldstein BA, Gallis JA, Middleton JP, Svetkey LP. Serum potassium levels and risk of sudden cardiac death among patients with chronic kidney disease and significant coronary artery disease. *Kidney Int Rep*. 2017;2:1122–1131. https://doi.org/10.1016/j.ekir.2017.07.001.
29. Li M, Chen Y, Zhang Y, et al. Admission serum cholinesterase concentration for prediction of in-hospital mortality in very elderly patients with acute ischemic stroke: a retrospective study. *Aging Clin Exp Res*. 2020;32:2667–2675. https://doi.org/10.1007/s40520-020-01498-z.
30. Idicula TT, Waje-Andreassen U, Brogger J, Naess H, Thomassen L. Serum Albumin in ischemic stroke patients: the higher the better. The Bergen Stroke study. *Cerebrovasc Dis*. 2009;28:13–17. https://doi.org/10.1159/000215938.
31. Wang C, Deng L, Qiu S, et al. Serum albumin is negatively associated with hemorrhagic transformation in acute ischemic stroke patients. *Cerebrovasc Dis*. 2019; 47:88–94. https://doi.org/10.1159/000498855.
32. Picker SM. Platelet function in ischemic heart disease. *J Cardiovasc Pharmacol*. 2013; 61:166–174. https://doi.org/10.1097/FJC.0b013e318279b78a.
33. Zhao YH, Gao H, Pan ZY, et al. Prognostic value of NT-proBNP after ischemic stroke: a systematic review and meta-analysis of prospective cohort studies. *J Stroke Cerebrovasc Dis*. 2020;29:104659. https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.104659.
34. Mishra RK, Judson G, Christenson RH, DeFilippi C, Wu AHB, Whooley MA. The association of five-year changes in the levels of N-terminal fragment of the prohormone brain-type natriuretic peptide (NT-proBNP) with subsequent heart failure and death in patients with stable coronary artery disease: the heart and soul study. *Cardiology*. 2017;137:201–206. https://doi.org/10.1159/000466682.
35. Tveit SH, Myhre PL, Hanssen TA, et al. Cardiac troponin I and T for ruling out coronary artery disease in suspected chronic coronary syndrome. *Sci Rep*. 2022;12: 945. https://doi.org/10.1038/s41598-022-04850-7.
36. Adamson PD, Hunter A, Madsen DM, et al. High-sensitivity cardiac troponin I and the diagnosis of coronary artery disease in patients with suspected angina pectoris. *Circ Cardiovasc Qual Outcomes*. 2018;11:e004227. https://doi.org/10.1161/CIRCOUTCOMES.117.004227.
37. Elgendy IY, Pepine CJ. High sensitivity cardiac troponin T and I and risk stratification of patients with stable CHD: is it time to incorporate this in routine clinical practice? *Int J Cardiol*. 2018;250:266–267. https://doi.org/10.1016/j.ijcard.2017.10.007.
38. Liu W, Zhu X, Tan X, et al. Predictive value of serum creatinine/cystatin C in acute ischemic stroke patients under nutritional intervention. *J Nutr Health Aging*. 2021; 25:335–339. https://doi.org/10.1007/s12603-020-1495-0.
39. Jiang WF, Deng ML. Prognostic impact of blood urea nitrogen/creatinine ratio changes in patients with acute ischemic stroke. *Clin Neurol Neurosurg*. 2022;215: 107204. https://doi.org/10.1016/j.clineuro.2022.107204.
40. Lei M, Li Y, Cheng L, et al. Stress hyperglycemia ratio as a biomarker for early mortality risk stratification in cardiovascular disease: a propensity-matched analysis. *Cardiovasc Diabetol*. 2025;24:286. https://doi.org/10.1186/s12933-025-02812-7.

ARTICLES FOR FACULTY MEMBERS

AN IMPROVED MLP-RFE MODEL FOR ENHANCED ACCURACY IN HEART DISEASE PREDICTION USING DEEP LEARNING

A Novel Ensemble Deep Learning Model for Coronary Heart Disease Prediction / Riyaz, L., Butt, M. A., & Zaman, M.

Revue d'Intelligence Artificielle
Volume 36 Issue 6 (2022) Pages 825–832
https://doi.org/10.18280/ria.360602
(Database: International Information and Engineering Technology Association)

29th April 2026

IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Revue d'Intelligence Artificielle
Vol. 36, No. 6, December, 2022, pp. 825-832
Journal homepage: http://iieta.org/journals/ria

A Novel Ensemble Deep Learning Model for Coronary Heart Disease Prediction

Lubna Riyaz[1*], Muheet A. Butt[1], Majid Zaman[2]

[1] Department of Computer Sciences, University of Kashmir, Srinagar 190006, India
[2] Directorate of IT & SS, University of Kashmir, Srinagar 190006, India

Corresponding Author Email: lubna.scholar@kashmiruniversity.net

https://doi.org/10.18280/ria.360602

Received: 9 July 2022
Accepted: 18 October 2022

Keywords:
coronary heart disease, ensemble learning, prediction, Framingham, multilayer perceptron

ABSTRACT

In the last decade, heart diseases have become the leading cause of deaths in the world. Various risk factors associated with heart disease include age, gender, cholesterol levels, blood pressure, glucose levels, chest pain, obesity, stress, family history, etc. with the help of which heart diseases can be predicted in any patient. In the past decade or so various efforts have been made by the researchers for effective prediction of various heart diseases. In this paper, a novel ensemble deep learning model has been proposed for efficient prediction of coronary heart disease. The dataset used for this purpose is collected from the Framingham heart disease database. Different performance evaluation metrics including precision, accuracy, recall and f1-score are being used for evaluating the performance of the proposed model. As per the experimental results, the proposed ensemble model outperformed most of the conventional machine learning techniques in terms of accuracy, precision, recall and f1 score for coronary heart disease prediction.

1. INTRODUCTION

Heart is a vital organ in the human body. It supplies blood to different body parts. A slight deviation in the normal working of the heart can affect the overall functioning of the body. Therefore heart health is necessary for a person to survive. In the last decade or so heart diseases have been the primary reason of deaths in the world. In USA alone, one in every four deaths occurs because of heart disease [1]. Various risk factors associated with heart disease include age, gender, cholesterol levels, blood pressure, glucose levels, chest pain, obesity, stress and family history etc. With the help of these factors one can predict the presence or absence of a heart disease in a patient. According to Centers for Disease Control and Prevention (CDC), the most common risk factors for heart disease include high blood pressure, high blood cholesterol levels and smoking [2]. Other risk factors include diabetes, obesity, unhealthy diet, physical inactivity and excessive consumption of alcohol. According to WHO (World Health Organization), the major underlying determinants of heart diseases include social, economic and cultural changes including globalization, urbanization and population ageing, poverty, stress and hereditary factors [3]. The various conditions of heart disease include coronary artery disease, valvular heart disease, cardiomyopathy, heart rhythm disturbances and various heart infections [4].

Coronary artery or the coronary heart disease is the most common type of heart disease in today’s world. It is the major cause of deaths in the United States. Coronary heart disease is sometimes also called as the ischemic heart disease or simply a heart disease. Cholesterol deposits or plaques in heart arteries are the major reason behind coronary heart disease. Coronary arteries are vessels whose job is to supply oxygen-rich blood to the heart. Coronary heart disease starts with deposition of fats on the artery walls known as atherosclerosis leading to narrowing of heart arteries. The heart shows symptoms of coronary heart disease when it is not getting sufficient oxygen rich blood, hence reducing the amount of blood flowing to the heart which in turn leads to chest pain and shortness of breath [5]. If the blood flow is completely blocked, it can result in a heart attack.

Coronary heart disease may take decades to develop and can go unnoticed for a long time. Thus, detection of coronary heart disease at an early stage is quite necessary. The symptoms for coronary heart disease usually include chest pain, shortness of breath, fatigue (tiredness), dizziness, nausea and weakness and in case women the symptoms can be little bit different causing discomfort in shoulders, indigestion problem, anxiety and cold sweat [6] and sometimes a patient will not be knowing that he/ she has a coronary heart disease until the patient gets a heart attack [6]. The risk of coronary heart disease increases if there is a family history of coronary heart disease [7]. Angina is a term that is normally used to describe the most common symptoms of coronary heart disease [8].

In the past decade, efforts have been made by various researchers for effective prediction of heart disease [9-36] and various other diseases [37, 38] using machine learning. In this study, a novel ensemble deep learning model has been proposed for effective prediction of coronary heart disease. The dataset is obtained from the Framingham heart disease database. In the first step, data preprocessing was done followed by selection of the most significant features from the dataset for the prediction purpose. Finally, a deep learning multilayer perceptron classifier was trained three times each time using a different instance of training and testing set. Then, in the final step, the outputs from these three classifiers instances were fed into an ensemble majority voting classifier to produce the final results. Upon observation, it was found that ensembling resulted in increase in accuracy of the classifier.

The rest of this paper is organized as follows. Section 2 gives the description of the literature. Section 3 describes ensemble learning and multilayer perceptron. Section 4 describes the methodology, section 5 is the results and discussion section and section 6 summarizes the conclusion.

2. LITERATURE REVIEW

In the literature various previous studies related to heart disease prediction were reviewed.

In a study, Shouman et al. [9] have used various types of decision trees for improving the prediction of heart disease. Sensitivity, specificity and accuracy are used to evaluate the model performance. In another study, Subbalakshmi et al. [10] have performed the prediction of heart disease using naïve bayes algorithm by developing a Decision Support in Heart Disease Prediction System (DSHDPS) along with a web based questionnaire application. However, the performance of the proposed system wasn't that much satisfiable. In another study, Anooj [21] proposed a clinical decision support system using weighted fuzzy rules for heart disease prediction and compared the experimental results with other studies.

In another study, Pattekari and Parveen [22] proposed an intelligent heart disease prediction system based on naïve bayes technique and implemented as a web based application. In a study, Pandey et al. [23] have developed a heart disease prediction system using decision tree machine learning algorithm. For performing the task of prediction, 14 most significant features have been selected. Then, J48 decision tree has been used to build a prediction model for heart disease prediction. As per the experimental results, pruned J48 algorithm with reduced error has performed better as compared to the simple pruned and unpruned techniques. In another study, Medhekar et al. [24] proposed a system for heart disease prediction using naïve bayes. Classification has been done for a 5 class classification problem by dividing the chances of heart disease into 5 categories i.e. no, low, average, high and very high. According to the results, the proposed system was able to achieve the average accuracy of 89.10% on Cleveland heart disease dataset. In another study Sonawane and Patil [25] proposed a heart disease prediction system using learning vector quantization algorithm that takes as input the 12 most significant features and performs the task of classification based on these 12 significant features. According to the results, it was found that the system was able to achieve the highest accuracy of 85.55%.

Sabarinathan and Sugumaran [26] proposed a heart disease prediction using decision tree classification technique. The proposed system was able to achieve an overall accuracy of 85% on the test set. Cong et al. [27] proposed a firefly based algorithm for heart disease prediction using reduced number of features. For reducing the number of features, the proposed firefly algorithm has been combined with the rough sets thereby reducing the dimensionality of the dataset. In a study Ismaeel et al. [28] performed heart disease prediction using extreme learning machine considering various factors related to heart disease. The proposed system can give a warning to a patient at the earliest possible stage. The data has been taken from the Cleveland dataset containing 300 records of patients. The proposed architecture has achieved the overall accuracy of 80% for predicting heart disease. In another study, Singh et al. [11] also proposed a study for heart disease prediction. As per the results, the proposed system was able to achieve an overall accuracy of 85.81% for heart disease prediction.

In a study Purushottam et al. [12] proposed a system for heart disease prediction using association rules. Training and testing of model has been done using the 10 fold cross validation for efficient prediction of heart disease. According to the results, the proposed system achieved an overall accuracy of 86.7% for diagnosing heart disease [12]. Another study carried out by Gupta et al. [13] performed efficient prediction of heart disease using naïve bayes. According to the results the system was able to achieve the prediction accuracy of 86.42%. Reddy and Khare [14] in their study performed prediction of heart disease using hybrid using oppositional firefly with BAT where firstly the significant features were selected from the dataset and then the fuzzy rules were being created for the said data. The dataset is obtained from the UCI machine learning repository. The proposed system was able to achieve an accuracy of 78% on the said dataset [14].

Haq et al. [15] proposed a system for heart disease prediction using intelligent hybrid architecture employing various machine learning techniques. A total of 7 machine learning techniques were used in this study. According to the results, the reduced set of features has showed a notable improvement in classification accuracies [15]. Vijayashree and Sultana [16] used PSO (Particle Swarm Optimization) with SVM for selecting features for heart disease prediction. For increasing the accuracy the number of features has been reduced. As per the results the proposed model has showed a significant improvement in the results [16].

In another study proposed by Mohan et al. [17], the authors have developed a novel method by hybridizing some of the common machine learning algorithms for efficient prediction of heart disease. Different combinations of available attributes have been tested. The proposed system was able to achieve the overall accuracy of 88.7% [17]. Singh and Kumar [18] proposed a model using various machine learning techniques for effective prediction of heart disease. The dataset was collected from the UCI machine learning repository. As per their results, the k nearest neighbour algorithm was able to produce better results among all classifiers [18].

Jindal et al. [19] proposed a study for heart attack prediction using machine learning techniques. Two machine learning classifiers logistic regression and k nearest neighbour were used for doing the prediction. According to the results, these two classifiers showed the better results than the other available machine learning classifiers [19]. Riyaz et al. [20] proposed a study for heart disease prediction using machine learning. According to their results, the ANN (Artificial Neural Network) technique produced better results as compared to others achieving an overall average accuracy of 86.91%. El-Hasnony et al. [36] proposed another machine learning model for heart disease diagnosis using multi label active learning based technique. As per the authors, the proposed technique was able to achieve an accuracy of 57.4 ± 4% [36].

Table 1 gives a description of the currently existing studies for heart disease prediction.

3. TECHNIQUES

3.1 Ensemble learning

In ensemble learning, multiple models (including classifiers and experts) are combined in order to solve a particular problem. It is basically done to improve the overall performance of the model by reducing the chance of selecting the wrong one. More applications include assigning confidence to a model decision, selecting optimal features and incremental learning etc.

Table 1. Existing studies

Author	Technique used	Accuracy achieved
Shouman et al. [9]	Decision Tree	84.10%
Pandey et al. [23]	Pruned J48 Decision Tree	75.73%
Medhekar et al. [24]	Naïve Bayes	89.10%
Sonawane et al. [25]	Learning Vector Quantization Algorithm	85.55%
Sabarinathan et al. [26]	Decision Tree	85%
Cong et al. [27]	Fuzzy logic system	88.3%
Ismaeel et al. [28]	Extreme Learning Machine	80%
Purushottam et al. [12]	Association Rules	86.7%
Gupta et al. [13]	Naïve Bayes	86.42%
Reddy et al. [14]	Fuzzy logic	78%
Haq et al. [15]	Logistic Regression	89%
Viijayashree et al. [16]	Particle Swarm Optimization	88.22%
Mohan et al. [17]	Hybrid model	88.7%
Singh et al. [18]	K Nearest Neighbors	87%
Jindal et al. [19]	K Nearest Neighbors	88.52%
El-Hasnony et al. [36]	Multi label active learning	57.4%

Ensemble classifier is achieved by aggregating various models (classifiers or experts) together sometimes also called as the multiple classifier systems. In ensembling the results from various models are combined into a single one in order to obtain the final results which are the improved ones. Ensembling can be understood from psychological point of view like before making any decision in our daily lives, suggestions from multiple experts are taken into consideration before making the final decision. Examples include prior to starting any medical procedure, before buying any item, article reviewing before acceptance, etc., same is the case with ensemble learning. In doing so, the main focus is on reducing the chance of selecting the wrong one. The process of ensemble learning is shown in Figure 1.

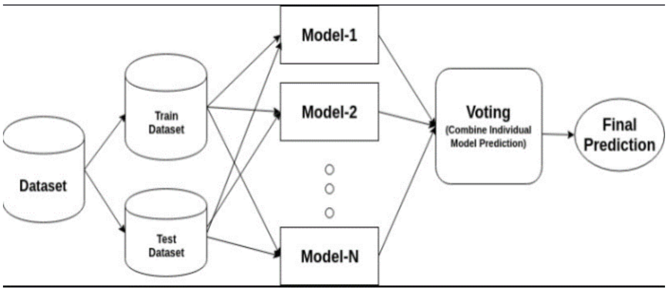


Figure 1. Ensemble learning

Firstly, the dataset is split into two subsets, training set and the testing set. Then, the models are trained using the said training set and eventually tested. Finally, the results from all models are being fed into the voting model in order to combine all the models and return the final prediction.

3.1.1 Bagging

Bagging, also known as bootstrap aggregation, is the ensemble learning method that is commonly used to reduce variance within a noisy dataset. In bagging, a random sample of data in a training set is selected with replacement-meaning that the individual data points can be chosen more than once.

After several data samples are generated, these weak models are then trained independently, and depending on the type of task-regression or classification, for example-the average or majority of those predictions yield a more accurate estimate. The algorithm for bagging is shown as under:

1. Select records from your dataset, with replacement, to get bootstrapped dataset.
2. Train a base classifier using this subset dataset, normally the base classifier used for this purpose is decision tree.
3. Keep repeating 1 and 2 ‘N’ times where N is a pre-chosen number.
4. Combine all N classifiers together into a single rule.

3.1.2 Boosting

Boosting works like bagging. A family of models is being created combined for getting more robust learner that could perform finer. Boosting sequentially combines various weak classifiers, each time more importance is being given to the wrongly classified records by preceding classifiers.

The algorithm for boosting is shown as under:

1. Initialize dataset while assigning same weights to all the observations.
2. Input this to the first classifier and locate the wrongly classified ones.
3. Increase weights of wrongly classified observations.
4. Update the weight of the dataset and send it back to the model.
5. Repeat steps 2-4 until all wrongly classified data points are fixed.
6. End

3.2 Multilayer perceptron

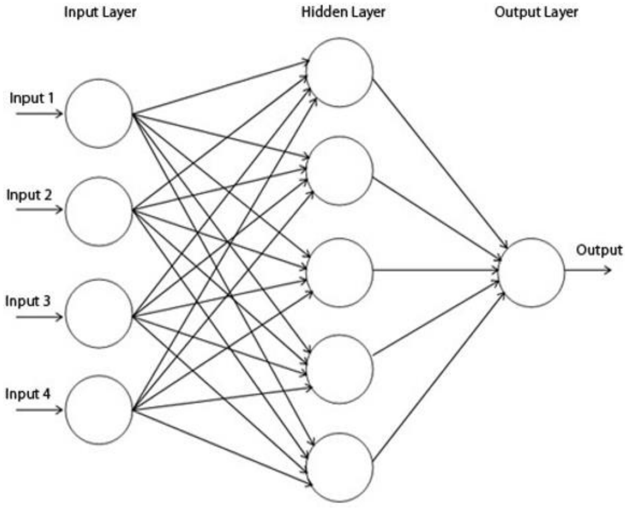


Figure 2. Multilayer perceptron

A multilayer perceptron (MLP) is a fully connected feed forward artificial neural network (ANN) that learns a function

$f(.): R^m \rightarrow R^o$ by training on a dataset, where m represents the input dimensions and o is the dimension of the output. Given a set of features $X = x_1, x_2, \ldots, x_m$ and a target y, an MLP can learn a non-linear function approximation for classification or regression.

The term MLP is used ambiguously, sometimes to mean any feed forward neural network and sometimes to specifically refer to networks composed of multiple perceptron layers. If the multilayer perceptron contains only a single layer, then it is called as the vanilla neural network. A multilayer perceptron consists of three sets of layers containing nodes: an input layer, one or more hidden layers and an output layer. Multilayer perceptron uses the backpropagation supervised learning technique. A multilayer perceptron is used when the data is not linearly separable. Figure 2 shows one instance of a multilayer layer perceptron neural network with only one hidden layer.

The nodes in case of a multilayer perceptron are connected as a linked directed graph starting from input and ending at the output layers. Except for the input nodes, all other nodes have an activation function. The input data enters through the input layer, propagates through one or more hidden layers and finally reaches the output layer which returns the final output. The outputs from one layer act as input to the next layer. Each node of the hidden layer converts the values from the previous layer with a weighted linear summation $w_1x_1 + w_2x_2 + \cdots + w_m m$, followed by a non-linear activation function $g(.): R \rightarrow R -$ like a hyperbolic tan function. For training the network, the multilayer perceptron uses backpropagation.

4. METHODOLOGY

The dataset for this study was collected from the Framingham heart disease database which is an ongoing heart disease study from the people of Framingham, USA. The dataset contains 4240 records of patients with 15 attributes and one target attribute. The description of the said dataset is given in Table 2. Figure 3 shows the correlation matrix of attributes.

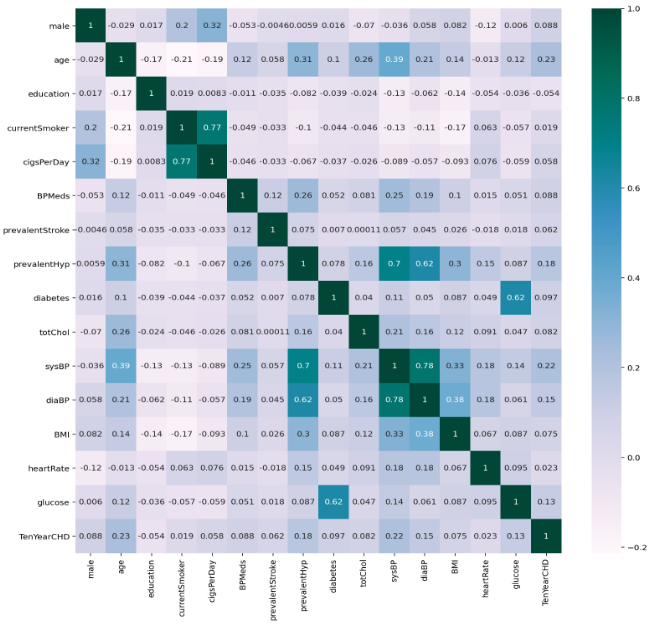


Figure 3. Correlation matrix

Prior to training and testing of the model, proper preprocessing of the data was done. In the very first step null values were identified in the data. Out of 4240 records of patients, there were 582 records with missing values. For each attribute, the missing values were handled differently. For attribute education, the records with missing values were considered illiterate and hence were filled with the value “0”. The cigsPerDay attribute was applicable to those patients only for whom the value of the currentSmoker attribute was equal to 1, hence the missing values for this attribute were filled by calculating the median of cigsPerDay attribute values for those records only for which the value of currentSmoker attribute was equal to 1.

Table 2. Framingham heart disease dataset description

S. No.	Attribute	S. No.	Attribute
1	Gender of the patient (male)	10	Total Cholesterol levels
2	Age (age)	11	Systolic BP
3	Education	12	Diastolic BP
4	Whether Current Smoker	13	Body Mass Index (BMI)
5	No. of cigarettes smoking per day	14	Heart Rate
6	Whether on Blood Pressure medication	15	Glucose levels
7	History of any stroke	16	TenYearCHD (Target variable)
8	Whether Hypertensive		
9	Whether having Diabetes		

The missing values in case of attribute BPMeds were assumed that they were not taking any blood pressure medicine at all, hence its missing values were filled with value 0. In case of attributes totChol and BMI, the missing values were filled by calculating the median values of the rest of the records for respective attributes. For attribute heartRate, there was only one record with missing value and that too of a current smoker, hence its missing value was filled with the median of heartRate attribute values of all those records for which the value of currentSmoker attribute was equal to 1. For attribute glucose, the majority of the missing values belonged to the non-diabetic category, hence its missing values were filled by calculating the median of attribute glucose values for non-diabetic patients. Upto this point all the records with null values were successfully handled in our dataset.

Then, in the next step detection of outliers was performed. Removable outliers were found in totChol and sysBP columns and were removed accordingly since outliers in data can affect the overall performance of the model. Figure 4 shows the plot representation of the outliers present. A total of three removable outliers were detected in our dataset, two in totChol (600 & 696) and one in sysBP column (295).

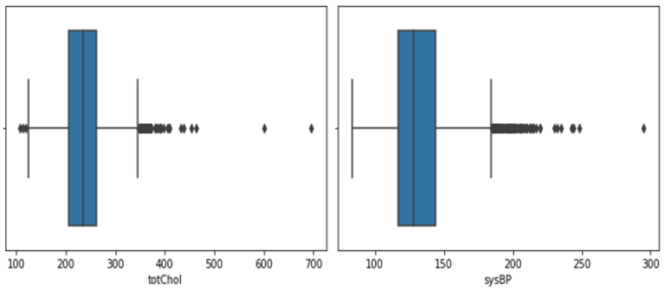


Figure 4. Plots showing outliers

After successful removal of outliers, the dataset was then left with 4237 records of patients only. Then in the next step the dataset was tested for imbalance, it was observed that the dataset was highly imbalanced since the number of records with target value 0 were much higher in number than those with target value 1 (Figure 5).

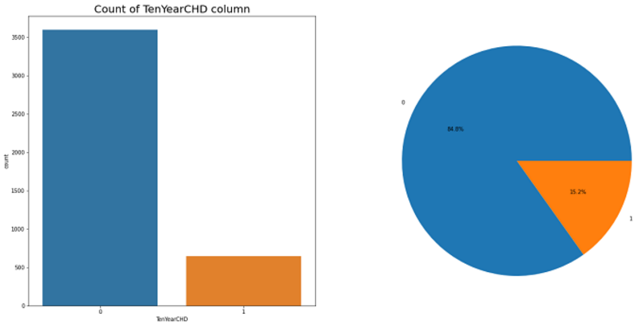


Figure 5. Plot showing number of instances for each of the two target values

Hence resampling of the data was done by oversampling the positive (value 1) cases in such a way that number of positive and number of negative cases became equal to each other thereby increasing the size of the dataset to 7190 records of patients. Figure 6 shows the block diagram of the proposed model.

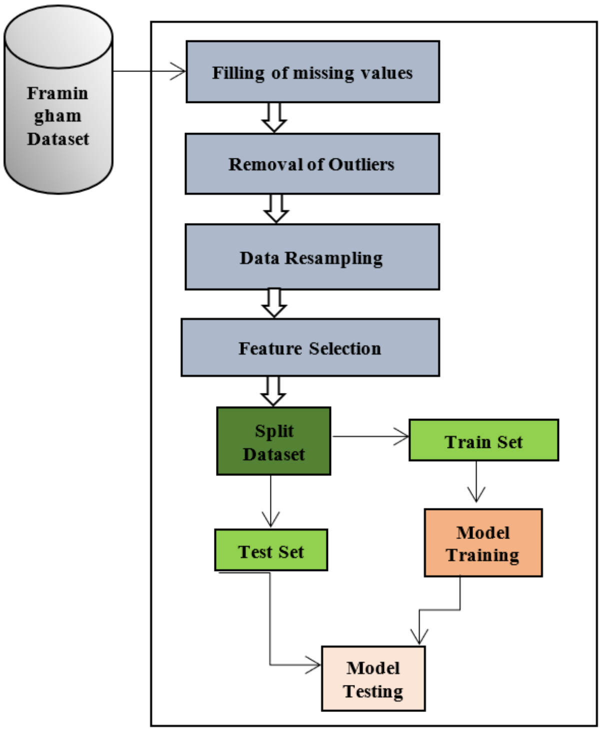


Figure 6. Block diagram of the proposed model

After resampling of the available dataset, the task of feature selection was done by selecting the most significant features from the dataset based on the IG (Information Gain) value for each attribute. Table 3 shows the list of attributes in the descending order of their information gain values. Out of the total attributes, the top ten most significant attributes were selected from the dataset for prediction purpose while dropping the remaining ones.

The top ten most significant features were selected for coronary heart disease prediction including BMI, sysBP, totChol, age, diaBP, heartRate, prevalentHyp, glucose, cigsPerDay and male, and the features which were not considered included education, prevalentStroke, BPMeds, currentSmoker and diabetes. After feature selection, the dataset was then split into two subsets, training set (80%) and testing set (20%) three times, each time using a different combination of records for training and testing of the multilayer perceptron (MLP) classifier. Data scaling was also performed in order to scale our features to a given range. Hyper parameter tuning for each instance of the multilayer perceptron classifier was also performed in order to increase the prediction accuracy. The best hyper parameters found for each instance of a multilayer perceptron classifier are shown in Table 4. From the table it can be seen that the best hyper parameters chosen for 'activation', 'alpha', and 'solver' remained same throughout all instances that is tanh, 0.0001and lbfgs respectively, however, for hyper parameter 'learning rate' only it varied.

Table 3. List of attributes in the descending order of their information gain values

S. No.	Attributes
1	BMI
2	sysBP
3	totChol
4	age
5	diaBP
6	heartRate
7	prevalentHyp
8	glucose
9	cigsPerDay
10	male
11	education
12	prevalentStroke
13	BPMeds
14	currentSmoker
15	diabetes

Table 4. Best parameters found for multilayer perceptron after Hyperparameter tuning

	activation	alpha	learning rate	solver
mlp1	tanh	0.0001	constant	lbfgs
mlp2	tanh	0.0001	invscaling	lbfgs
mlp3	tanh	0.0001	constant	lbfgs

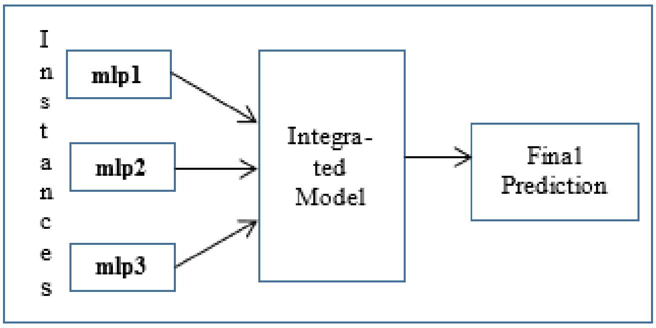


Figure 7. Integrated model

The results achieved by each of the three multilayer perceptron instances are shown in Tables 5, 6 and 7 respectively. Finally, the results from the three multilayer perceptron layers were fed into an ensemble majority voting integrated classifier (as shown in Figure 7) which performed the final prediction based on the output from these three classifiers. The final results after ensembling are shown in Table 8.

5. RESULTS AND DISCUSSION

Data preparation and experimental steps have already been explained. Now, in this section we present the experimental results and discussion part. Tables 5-7 show the experimental results for each of the three multi-layer perceptron instances for coronary heart disease prediction. Precision, recall, f1 score, support and accuracy performance metrics were used to evaluate the model efficiencies. Precision is the ratio of correctly predicted positive records to the total predicted positive records. Recall is the ratio of correctly predicted positive records to all the records in actual positive class. F1 score is the weighted average of precision and recall. Support is the number of actual occurrences of each class in the given dataset. And accuracy is the ratio of correctly predicted records to the total records in the dataset. Table 5 shows the results of the first instance of the multilayer perceptron classifier. The results achieved were as 93% precision, 92% recall, 92% f1-score and 92% accuracy. For second instance the results achieved were as 92% precision, 91% recall, 91% f1-score and 91% accuracy and for third instance the results obtained are as 91% precision, 90% recall, 90% f1-score and 90% accuracy as shown in Tables 5-7.

Table 5. Results achieved by instance mlp1

	Precision	**Recall**	**F1-score**	**Support**
0	0.99	0.85	0.91	725
1	0.86	0.99	0.92	713
accuracy			0.92	1438
macro average	0.93	0.92	0.92	1438
weighted average	0.93	0.92	0.92	1438

Table 6. Results achieved by instance mlp2

	Precision	**Recall**	**F1-score**	**Support**
0	0.99	0.84	0.91	737
1	0.85	0.99	0.91	701
accuracy			0.91	1438
macro average	0.92	0.91	0.91	1438
weighted average	0.92	0.91	0.91	1438

Table 7. Results achieved by instance mlp3

	Precision	**Recall**	**F1-score**	**Support**
0	0.98	0.82	0.89	704
1	0.85	0.98	0.91	734
accuracy			0.90	1438
macro average	0.91	0.90	0.90	1438
weighted average	0.91	0.90	0.90	1438

Table 8. Results achieved after ensembling

	Precision	**Recall**	**F1-score**	**Support**
0	0.99	0.88	0.93	725
1	0.89	0.99	0.94	734
accuracy			0.94	1438
macro average	0.94	0.94	0.94	1438
weighted average	0.94	0.94	0.94	1438

The results from these three multi-layer perceptron classifier instances were then fed into a majority voting ensemble classifier to perform the final prediction. The results obtained are shown in Table 8.

The final results achieved after ensembling the three multi-layer perceptron instances were as follows: precision (94%), recall (94%), f1-score (94%) and accuracy (94%) as shown in table (Table 8). From the results it can be clearly seen that after ensembling of classifiers, the precision, recall, f1-score and accuracy of the model increased upto a large extent. This is due to the fact that in ensembling the final output is decided based on the outputs from multiple classifiers instead of any single individual classifier, multiple classifiers are trained at the same time and then the final prediction is done by combining the outputs from all the individual classifiers to produce the final output hence resulting in improved model efficiencies.

Table 9 compares the results of the proposed ensemble model with some of the common machine learning classifiers.

Table 9. Comparison of results with other machine learning classifiers

	Precision	**Recall**	**F1-Score**	**Accuracy**
Logistic Regression	68%	68%	68%	68%
Decision Tree	92%	91%	91%	91%
SVM	69%	68%	68%	68%
Multinomial Naïve Bayes	62%	61%	61%	61%
K Nearest Neighbors	80%	78%	78%	78%
AdaBoost	70%	69%	69%	69%
Gradient Boosting	72%	71%	71%	71%
Multilayer Perceptron	93%	92%	92%	92%
Proposed Ensemble Model	**94%**	**94%**	**94%**	**94%**

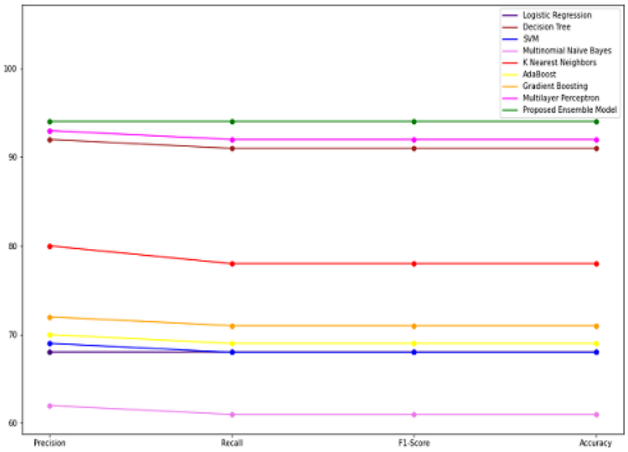


Figure 8. Performance comparison

From Table 9, it is evident that the proposed model outperformed most of the individual machine learning classifiers in terms of accuracy, precision, recall and f1 score for coronary heart disease prediction using common preprocessed data. Figure 8 shows the graphical representation of the performance comparison between some of the individual machine learning classifiers and the proposed ensemble classifier. Therefore, it was concluded that ensembling improves the overall efficiency of any machine learning classifier by combining the outputs from multiple classifiers. Hence this study suggests use of ensemble learning for coronary heart disease prediction.

6. CONCLUSIONS

Heart disease is the most prominent cause of deaths in the world today. As per WHO, it accounts for almost 18 million deaths per year. Therefore, detection of heart disease at an early stage is vital if we want to save human lives in near future. The focus of this study was solely on prediction of one of the types of heart disease known as the coronary heart disease using ensemble deep learning. Coronary heart disease is mainly caused by the buildup of plaque on the artery walls. Firstly the dataset was collected from the Framingham heart study database which is an ongoing cardiovascular study of the people from Framingham city in US. Then, the data was tested for any missing value and handled accordingly. In the next step outliers were removed from the dataset. In the next step the dataset was balanced. Feature selection was also performed by selecting the 10 most significant features based on their IG values for coronary heart disease prediction. Finally, the dataset was split into train and test set for training and testing of classifiers. In the last step, performance evaluation of the proposed ensemble model was done by comparing the results of the proposed ensemble deep learning model with some of the existing machine learning classifiers. According to the results, it was evident that the proposed ensemble deep learning model outperformed most of the existing machine learning classifiers in terms of accuracy, precision, recall and f1-score for coronary heart disease prediction. Hence, it was concluded that ensembling improves the overall efficiency of the individual models by combining the results from multiple classifiers into a single one thereby returning the final predicted output. In future, a more robust dataset can be obtained by collecting the data from multiple sources instead of a single one using data fusion approach for further improving the efficiency. In addition to this, model efficiencies can further be improved by assigning different weights to different attribute instead of a single general weight using some feature weighting technique for more improved results.

REFERENCES

[1] https://www.healthline.com/health/heart-disease

[2] https://www.cdc.gov/heartdisease/about.htm

[3] https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[4] https://www.medicinenet.com/heart_disease_coronary_artery_disease/article.htm

[5] https://www.mayoclinic.org/diseases-conditions/coronary-artery-disease/symptoms-causes/syc-20350613

[6] https://my.clevelandclinic.org/health/diseases/16898-coronary-artery-disease

[7] https://www.cancer.gov/publications/dictionaries/cancer-terms/def/coronary-heart-disease

[8] https://www.bhf.org.uk/informationsupport/conditions/coronary-heart-disease

[9] Shouman, M., Turner, T., Stocker, R. (2011). Using decision tree for diagnosing heart disease patients. In Proceedings of the 9-th Australasian Data Mining Conference (AusDM'11), Ballarat, Australia, pp. 23-30.

[10] Subbalakshmi, G., Ramesh, K., Rao, M.C. (2011). Decision support in heart disease prediction system using naive bayes. Indian Journal of Computer Science and Engineering (IJCSE), 2(2): 170-176.

[11] Singh, Y.K., Sinha, N., Singh, S.K. (2017). Heart disease prediction system using random forest. In Advances in Computing and Data Sciences First International Conference, ICACDS 2016, Ghaziabad, India, pp. 613-623. https://doi.org/10.1007/978-981-10-5427-3_63

[12] Purushottam, Saxena, Sharma, R. (2016). Efficient heart disease prediction system. Procedia - Procedia Comput. Sci., 85: 962-969. https://doi.org/10.1016/j.procs.2016.05.288

[13] Gupta, N., Ahuja, N., Malhotra, S., Bala, A., Kaur, G. (2016). Intelligent heart disease prediction in cloud environment through ensembling. Expert Systems, 34(3): e12207. https://doi.org/10.1111/exsy.12207

[14] Reddy, G.T., Khare, N. (2017). An effcient system for heart disease prediction using hybrid OFBAT with rule-based fuzzy logic model. Journal of Circuits, Systems and Computers, 26(4): 1750061. https://doi.org/10.1142/S021812661750061X

[15] Haq, A.U., Li, J.P., Memon, M.H., Nazir, S., Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Information Systems, 2018: 3860146. https://doi.org/10.1155/2018/3860146

[16] Vijayashree, J., Sultana, H.P. (2018). A machine learning framework for feature selection in heart disease classification using improved particle swarm optimization with support vector machine classifier. Programming and Computer Software, 44(6): 388-397. https://doi.org/10.1134/S0361768818060129

[17] Mohan, S., Thirumalai, C., Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7: 81542-81554. https://doi.org/10.1109/ACCESS.2019.2923707

[18] Singh, A., Kumar, R. (2020). Heart disease prediction using machine learning algorithms. 2020 International Conference on Electrical and Electronics Engineering (ICE3), Gorakhpur, India, pp. 452-457. https://doi.org/10.1109/ICE348803.2020.9122958

[19] Jindal, H., Agrawal, S., Khera, R., Jain, R., Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. IOP Conf. Ser. Mater. Sci. Eng., 1022: 012072. https://doi.org/10.1088/1757-899X/1022/1/012072

[20] Riyaz, L., Butt, M.A., Zaman, M., Ayob, O. (2022). Heart disease prediction using machine learning techniques: A quantitative review. International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing, pp. 81-94. https://doi.org/10.1007/978-981-16-3071-2_8

[21] Anooj, P.K. (2012). Clinical decision support system : Risk level prediction of heart disease using weighted fuzzy rules. J. King Saud Univ. - Comput. Inf. Sci., 24(1): 27-40. https://doi.org/10.1016/j.jksuci.2011.09.002

[22] Pattekari, S.A., Parveen, A. (2012). Prediction system for heart disease using naive bayes. International Journal of Advanced Computer and Mathematical Sciences, 3(3): 290-294.

[23] Pandey, A.K., Pandey, P., Jaiswal, K.L., Sen, A.K.

(2013). A heart disease prediction model using decision tree. IOSR Journal of Computer Engineering (IOSR-JCE), 12(6): 83-86.

[24] Medhekar, D.S., Bote, M.P., Deshmukh, S.D. (2013). Heart disease prediction system using naive Bayes. Int. J. Enhanced Res. Sci. Technol. Eng, 2(3): 1-5.

[25] Sonawane, J.S., Patil, D.R. (2014). Prediction of heart disease using learning vector quantization algorithm. In 2014 Conference on IT in Business, Industry and Government (CSIBIG), pp. 1-5. https://doi.org/10.1109/CSIBIG.2014.7056973

[26] Sabarinathan, V., Sugumaran, V. (2014). Diagnosis of heart disease using decision tree. International Journal of Research in Computer Applications & Information Technology, 2(6): 74-79.

[27] Cong, N., Meesad, P., Unger, H. (2015). Expert systems with applications a highly accurate firefly based algorithm for heart disease prediction. Expert Syst. Appl., 42(21): 8221-8231. https://doi.org/10.1016/j.eswa.2015.06.024

[28] Ismaeel, S., Miri, A., Chourishi, D. (2015). Using the Extreme Learning Machine (ELM) technique for heart disease diagnosis. In 2015 IEEE Canada International Humanitarian Technology Conference (IHTC2015), pp. 1-3. https://doi.org/10.1109/IHTC.2015.7238043

[29] https://scholar.google.co.th/citations?view_op=view_citation&hl=th&user=_wkCqD8AAAAJ&citation_for_view=_wkCqD8AAAAJ:M3NEmzRMIkIC

[30] Mohd, R., Butt, M.A., Baba, M.Z. (2020). GWLM–NARX: Grey Wolf Levenberg–Marquardt-based neural network for rainfall prediction. Data Technologies and Applications, 54(1): 85-102. https://doi.org/10.1108/DTA-08-2019-0130/FULL/PDF

[31] Altaf, I., Butt, M.A., Zaman, M. (2022). Disease detection and prediction using the liver function test data: A review of machine learning algorithms. International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing, pp. 785-800. https://doi.org/10.1007/978-981-16-2597-8_68

[32] https://www.academia.edu/80328577/A_hybrid_adaptive_grey_wolf_levenberg_marquardt_GWLM_and_nonlinear_autoregressive_with_exogenous_input_NARX_neural_network_model_for_the_prediction_of_rainfall?f_ri=10005

[33] https://www.academia.edu/78642235/Is_Decision_Tree_Obsolete_in_Its_Original_Form_A_Burning_Debate

[34] Riyaz, L., Butt, M.A., Zaman, M. (2022). Improving coronary heart disease prediction by outlier elimination. Appl. Comput. Sci., 18(1): 70-88. https://doi.org/10.35784/ACS-2022-6

[35] Riyaz, L., Butt, M.A., Zaman, M., Ayob, O. (2022). Heart disease prediction using machine learning techniques: A quantitative review. International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing, pp. 81-94. https://doi.org/10.1007/978-981-16-3071-2_8

[36] El-Hasnony, I.M., Elzeki, O.M., Alshehri, A., Salem, H. (2022). Multi-label active learning-based machine learning model for heart disease prediction. Sensors, 22(3): 1184. https://doi.org/10.3390/s22031184

[37] Swathi, K., Kodukula, S. (2022). XGBoost clssifier with hyperband optimization for cancer predicton based on geneselection by using machine learning techniques. Revue d'Intelligence Artificielle, 36(6): 665-670. https://doi.org/10.18280/ria.360502

[38] Srivastava, S., Kumar, G., Mishra, R.K., Kulshrestha, N. (2020). A complex diffusion based modified fuzzy C-means approach for segmentation of ultrasound image in presence of speckle noise for breast cancer detection. Revue d'Intelligence Artificielle, 34(4): 419-427. https://doi.org/10.18280/ria.340406

ARTICLES FOR FACULTY MEMBERS

AN IMPROVED MLP-RFE MODEL FOR ENHANCED ACCURACY IN HEART DISEASE PREDICTION USING DEEP LEARNING

Classification models combined with Boruta feature selection for heart disease prediction / Manikandan, G., Pragadeesh, B., Manojkumar, V., Karthikeyan, A. L., Manikandan, R., & Gandomi, A. H.

Informatics in Medicine Unlocked
Volume 44 101442 (2024) Pages 1-12
https://doi.org/10.1016/j.imu.2023.101442
(Database: ScienceDirect)

Informatics in Medicine Unlocked 44 (2024) 101442

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Classification models combined with Boruta feature selection for heart disease prediction

G. Manikandan [a], B. Pragadeesh [a], V. Manojkumar [a], A.L. Karthikeyan [a], R. Manikandan [a], Amir H. Gandomi [b,c,*]

[a] *School of Computing, SASTRA Deemed University, Thanjavur, Tamil Nadu, India*
[b] *Faculty of Engineering & Information Technology, University of Technology Sydney, Sydney, NSW, Australia*
[c] *University Research and Innovation Center (EKIK), Óbuda University, 1034, Budapest, Hungary*

ARTICLE INFO

Keywords:
Decision tree
Logistic regression
Support Vector Machine
Boruta
Heart disease
Feature selection

ABSTRACT

Cardiovascular disease (CVD), generally called heart illness, is a collective term for various ailments that affect the heart and blood vessels. Heart disease is a primary cause of fatality and morbidity in people worldwide, resulting in 18 million deaths per year. By identifying those who are most vulnerable to heart diseases and ensuring they receive the appropriate care, premature demise can be prevented. Machine learning algorithms are now crucial in the medical field, especially when using medical databases to diagnose diseases. Such efficient algorithms and data processing techniques are applied to predict various diseases and offer much potential for accurate heart disease prognosis. Therefore, this study compares the performance logistic regression, decision tree, and support vector machine (SVM) methods with and without Boruta feature selection. The Cleveland Clinic Heart Disease Dataset acquired from Kaggle, which consists of 14 features and 303 instances, was used for the investigation. It was found that the Boruta feature selection algorithm, which selects six of the most relevant features, improved the results of the algorithms. Among these classification algorithms, logistic regression produced the most efficient result, with an accuracy of 88.52 %.

1. Introduction

Cardiac arrest is one of the most common disorders with a significant fatality rate. The World Health Organization (WHO) states that one out of every three persons has high blood pressure, and more people in cities than in rural areas suffer from heart diseases [1]. Furthermore, people in their forties and fifties are more vulnerable to heart illness. Multiple elements, including individual lifestyle choices, medical history, and genetic predisposition, play a pivotal role in the emergence of heart illness. The likelihood of heart disease is increased by risk factors, such as smoking, binge drinking, sedentary behavior, depression, obesity, chronic stress, genetic disorders, and already existing cardiac problems [2]. In order to take preventative measures to avoid further complications and even death, the ability to swiftly, effectively, and accurately detect cardiac illness is essential.

Providing cost-effective, premium medical care is a primary hurdle for healthcare institutions. To receive quality care, a patient's problems must be appropriately identified and effective remedies must be provided. To resolve these challenges, it is crucial to integrate cutting edge technologies. This involves exploring methods, like utilizing data analysis thereby improving the decision making processes of healthcare professionals. By utilizing the potential of these technologies healthcare systems can optimize their ailments accordingly, which ultimately lays the foundation, for more precise and efficient patient care. Patients' information and treatments are maintained as a massive medical data warehouse. Instead of being informed by data-based understanding, medical decisions typically rely on the healthcare professional's career stint, training, and instinct.

The value of data in the ever-changing world of today cannot be emphasized. It acts as the cornerstone for efficient problem-solving by facilitating the extraction of priceless insights, the discovery of patterns, and the creation of crucial connections. These innovations improve prediction accuracy while enabling a deeper comprehension of complex, non-linear interactions within huge datasets. This, in turn, equips medical practitioners to make knowledgeable decisions and choose the best treatments from the onset of an illness. With its ability to

---

* Corresponding author. Faculty of Engineering & Information Systems, University of Technology Sydney, Sydney, Australia.
E-mail addresses: manikandan@it.sastra.edu (G. Manikandan), bpragadeesh.25@gmail.com (B. Pragadeesh), manojkumarvg95@gmail.com (V. Manojkumar), 112002karthik@gmail.com (A.L. Karthikeyan), srmanimt75@gmail.com (R. Manikandan), gandomi@uts.edu.au (A.H. Gandomi).

https://doi.org/10.1016/j.imu.2023.101442
Received 13 August 2023; Received in revised form 31 December 2023; Accepted 31 December 2023

autonomously adapt and improve models, machine learning is at the forefront of this revolution, empowering medical professionals to make informed choices and select optimal therapies right from the outset of a condition. This dynamic field continues to push the boundaries of what is possible, promising even greater advancements in healthcare and beyond [3]. Machine learning, which employs data to find intricate patterns and connections that human observation might overlook, is extremely helpful in the prognostication of heart disease. It gives medical practitioners the capacity to decide quickly and intelligently, which improves the diagnosis and prognosis of heart illness. Moreover, machine learning is being incorporated into cardiac care in ways that go beyond diagnosis and are essential to customized therapy regimens. These systems analyze patient-specific data to make personalized treatment and intervention recommendations that optimize outcomes and minimize potential adverse effects.

Accordingly, it is necessary to estimate the possibility that heart disease will become evident in a given set of populations through efficient data mining techniques; that is, techniques to examine data in order to identify significant patterns and practical insights in the dataset, which is a compilation of different data pertinent to the cause along with the resulting feature for every patient that is being monitored. In addition, in order to address the outcomes surpassing some insignificant features in the dataset, feature selection techniques are integrated, in which features are chosen based on its importance with respect to the resulting feature [7]. These estimation evaluations help in developing a machine-learning model to prognosticate the heart disease in the testing data with the help of the trends or insights learned from training data.

In this study, the proposed system predicts heart disease by employing decision tree, support vector machine classifier and logistic regression along with the Boruta feature selection algorithm based on the Cleveland Clinic Heart Disease Dataset acquired from Kaggle.

Increasing the precision of heart disease diagnosis is the primary motivation. Machine learning algorithms are invaluable in the medical field, leveraging large medical databases to predict and diagnose diseases. Examining the performance of non-ensemble classification models is a further motive. While ensemble techniques such as Random Forest and XGBoost have excelled in a number of applications, it is crucial to ascertain whether less complex models such as logistic regression and decision trees can also accurately prognosticate cardiac disease.

The remainder of this dissertation is structured as follows. Section 2 presents the literature review. The suggested system is explored in Section 3. In Section 4, benchmarks for the classification methods are provided, and Section 5 concludes the paper.

2. Literature survey

In the medical field, patient data are widely available and are analyzed using several machine learning methods. Particularly, healthcare professionals examine this information to make potent diagnostic decisions. Obtaining valuable data from vast databases is called data mining. Through analysis, medical data mining provides clinical support by employing various classification algorithms and tests these algorithms to see if they can anticipate patients' cardiac issues [4]. For instance, Weng et al. concluded that machine learning approaches can predict heart diseases more efficiently [5]. Numerous information extraction approaches are used, regression, association rules, clustering, and different classification algorithms like K-nearest neighbor, Naive Bayes, Random Forest, and Decision Trees.

Many studies have used data mining and machine learning approaches to diagnose heart diseases [6]. To classify heart illness into normal or abnormal, Reddy et al. implemented SVM, RF, KNN, Neural Networks, and NB along with a number of feature selection algorithms, such as recursive feature elimination (RFE), learning vector quantization (LVQ) model, and correlation matrix [7]. The results suggest that RF delivered the foremost results. Pillai et al. employed a recurrent neural network (RNN), genetic algorithm, and K-mean to prognosticate cardiac diseases [8]. RNN yielded the highest accuracy, whereas K-mean resulted in the least accuracy. Neural networks are computer representations of the human brain made up of artificial neurons that are connected to one another. They are employed in machine learning to process data and learn from it, which enables them to perform a variety of tasks. Leveraging sequential information, RNN excels in capturing temporal dependencies, making it well-suited for tasks like time-series analysis. However, some of the lower accuracy of the clustering algorithm K-means may be due to its inability to handle complex, non-linear patterns. Moreover, the utilization of genetic algorithms in conjunction with RNN showcases the potential for hybrid approaches in enhancing predictive accuracy. Genetic algorithms contribute to the optimization of model parameters, complementing the strengths of RNN and addressing its limitations.

To categorize heart illness, Raza used an ensemble learning model, multilayer perceptron (MLP), NB, as well as LR [9]. MLP, an artificial neural network, is made up of numerous layers of interconnected nodes (neurons). The MLP modifies its internal weights to reduce prediction errors by training with labeled 2data and backpropagation, allowing it to learn intricate patterns and make precise predictions. MLP's ability to capture non-linear relationships makes it particularly effective for intricate medical data patterns Because of its hierarchical structure, it can automatically extract features, which improves its ability to identify subtle symptoms of cardiac disease. Through backpropagation and other training techniques, MLP is able to improve its prediction accuracy over time. Furthermore, the incorporation of Naive Bayes (NB) and Logistic Regression (LR) alongside MLP in the ensemble model contributes to a comprehensive analysis of heart disease patterns. NB, based on probabilistic principles, excels in handling diverse and incomplete medical datasets, providing valuable insights into potential risk factors. LR, on the other hand, offers interpretability and simplicity, enhancing the model's transparency for clinical decision-makers. The outcomes reveal that ensemble learning techniques outperform other classifiers when it comes to cardiac illness prediction.

The authors of [10] proposed improved cardiac disease prediction using two algorithms (XGBoost and LR). The results revealed that LR XGBoost, had an accuracy of 84.46 %, while LR performed better with an accuracy of 85.68 %. To diagnose cardiac illness, Bhatet et al. created a model that implements an MLP with a backpropagation technique. This developed framework minimized error and attained a maximum accuracy of 80.99 % [11]. To predict heart pathology, Abushariah et al. deployed an artificial neural network (ANN) with an adaptive neuro-fuzzy inference system (ANFIS), which resulted in a minimum value of accuracy 75.93 %, while ANN had a maximum accuracy of 87.04 % [12]. ANFIS, a hybrid model, combines the interpretability of fuzzy logic systems with the flexibility of neural networks. Fuzzy inference systems are a type of structure that ANFIS uses to model intricate interactions between input and output data. ANFIS adjusts the fuzzy rule parameters in response to variations in data patterns by employing a hybrid learning technique. When fuzzy logic and neural networks are coupled, it performs better in handling uncertainty and non-linear relationships. In healthcare, ANFIS has shown promise for complex diagnostic tasks by effectively capturing and modeling intricate dependencies within medical datasets. Furthermore, because of its flexibility, ANFIS is especially valuable in healthcare environments where data may show dynamic and evolving patterns. Strong performance in real-world scenarios is ensured by its capacity to self-adjust to varying complexities in medical datasets, offering a dependable basis for challenging diagnostic tasks.

Hasan et al. made use of MLP with backpropagation and SVM in order to prognosticate heart disease [13]. As per the results, MLP produced a maximum accuracy of 98 %. On two datasets (Z-Alizadesh Sani and Cleveland Clinic Heart Disease Dataset), Sapra et al. investigated six machine learning procedures for identifying cardiac disorders, including LR, deep learning (DL), DT, RF, SVM, and ensemble learning (gradient

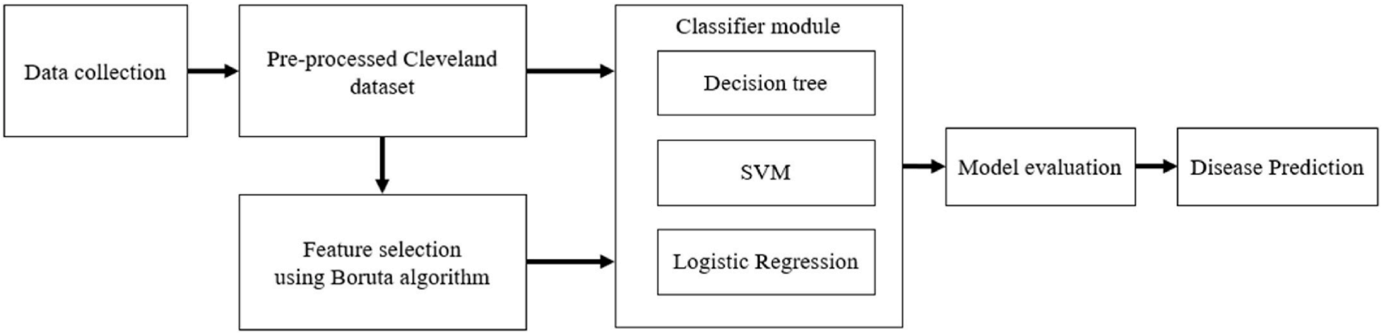


Fig. 1. Proposed system for heart disease prediction.

boosted tree) [14]. Compared to the other approaches, the gradient-boosted tree produced the highest accuracy of 84 %. Chen et al. applied ANN to the prognosis of heart disease utilizing various features [15].

Haq et al. leveraged the minimal-redundancy-maximal-relevance (mRMR), relief, and shrinkage and selection operator (LASSO) on the LR, ANN, KNN, NB, SVM, DT, and RF predictive modeling algorithms to predict cardiac illness [16]. LASSO is a widely used regularisation technique in machine learning and statistics. It is primarily employed to prevent overfitting and enhance model interpretability. Through the implementation of regularisation, LASSO promotes the identification of vital features, thereby augmenting the model's capacity to identify pertinent patterns in the prognostication of cardiac illness. This regularisation technique is particularly useful in clinical applications, as therapists can simplify complex models and thus better extract actionable insights from the data. The goal of mRMR is to reduce redundancy among selected features while shortlisting a subset of highly informative attributes for a given task. When compared to other approaches, LR with relief generated a maximum accuracy of 89 %.

Kumar et al. utilized a combination of ML techniques, such as logistic regression, random forest, and gradient boosting machines for the prognosis of heart disease [17]. Feature selection was performed using the relief algorithm, which achieved an accuracy of 92 %. The findings demonstrate that gradient boosting machines outperformed the competing algorithms. Sharma and Parmar experimented with a deep-learning neural network model in order to identify the best classification algorithm for prognosticating heart illnesses [18]. The outcome shows that hyper-parameter optimization using Talos yielded an accuracy of 90.78 %.

Muhammad et al. investigated several machine learning classifiers and employed four feature selection approaches, including Fast Correlation-Based Filter (FCBF), which ranks and determines the most beneficial features for enhanced model performance by combining correlation with the target variable and mutual information, Minimal Redundancy Maximal Relevance (mRMR), Least Absolute Shrinkage and Selection Operator (LASSO), and Relief, to enhance the diagnosis and recognition of cardiac illnesses. Relevance and informativeness are ensured by FCBF's ability to identify key features by taking into account mutual information and correlation with the target variable. Through the astute integration of these variables, FCBF improves the model's ability to prognosticate cardiac conditions. Using FCBF in feature selection helps clinicians make more informed decisions by providing a more focused and comprehensible representation of the data [19]. The experimental findings reveal that the Extra-Tree Classifier (ETC) scored better than other classifiers, achieving an accuracy of 92.09 % using 10-fold cross-validation. In addition, KNN exhibited the best performance at k = 7 and attained an accuracy of 85.55 %. An effective convolutional neural network (CNN) architecture was designed by Dutta et al. [20]. The results show that their CNN model effectively detected the presence of coronary heart disease cases on the testing data, which accounted for 85.70 % of the entire dataset, with a classification accuracy of 77 %.

A stacked model using KNN, RF, and SVM was demonstrated to be the best-performing model in Shorewala's study with an accuracy of 75.1 %, whereas the accuracy of bagged models was proven to be enhanced by 1.96 % [21]. Implementing a TANFIS classifier, Sekar et al. presented a highly successful clinical assistance approach to heart illness prediction [22]. Equipped with an Internet of Things (IoT) infrastructure, the TANFIS classifier successfully predicted cardiac illness with an astounding 99.76 % accuracy.

In the study by Rajendran and Karthi, an entropy-based feature engineering strategy in combination with the ensembling of machine learning classifiers was proposed [23]. The experimental findings show that the ensemble model, which includes naive Bayes (NB) and logistic regression (LR), performed exceptionally well, achieving an accuracy of 92.7 %. Employing the random forest (RF) approach, Chang et al. developed an artificial intelligence model for prognosticating heart disease [24]. The Python-based model exhibited an accuracy of almost 83 % over the training data.

A stacking classifiers model was established by Mohapatra et al. as an effective approach for discovering heart abnormalities and prognosticating heart illness [25]. The model, which combines heterogeneous learners at the base and meta levels, had good scores for precision, sensitivity, and specificity and an accuracy of 92 %.

A comprehensive analysis of machine learning (ML) strategies employed in medical applications was carried out by Shehab et al. [26]. The survey focused on the application of ML to improve reliability, efficiency, and accuracy in medical diagnosis, notably in computer-aided diagnosis (CAD) systems. The review covered more than 200 research articles and examined the advantages and disadvantages of using machine learning to address healthcare issues.

Diwan et al. proposed an ensemble machine learning approach combined with feature selection to prognosticate heart illnesses [27]. Of the various feature selection techniques that were tested, classification and regression trees (CART) yielded the optimum accuracy (87.65 %).

Recent research, exemplified by Ansari et al.'s study, highlights the remarkable success of algorithms like k-nearest neighbor and random forest, setting the stage for further advancements in predictive accuracy and feature selection methodologies. Various machine learning algorithms, along with feature selection techniques, were employed to prognosticate heart disease [29]. This expanding repertoire of machine learning algorithms showcases the dynamic landscape of cardiac prognostication, offering a diverse toolkit for both researchers and clinicians.

Saketha Rama et al. carried out a pragmatic comparison of classifiers, employing diverse medical datasets [30]. Remarkably, the SVM classifier demonstrated a 59 % accuracy on the Diabetic dataset by meticulously weighing individual model opinions, whereas the RF eclipsed these results with an accuracy of 0.9974 on the breast cancer Wisconsin dataset, leveraging its ensemble methodology to consolidate prevailing viewpoints.

Thus, machine learning models mentioned in the survey can serve as invaluable decision support tools for clinicians by aiding in the early

Table 1
Comparison of various works using feature selections.

Reference number	Feature selection algorithm	Selected number of features
21	LASSO	9
07	Correlation matrix and recursive feature elimination	8
04	Correlation based feature selection with Best First Method	10
27	Pearson's Correlation coefficient method, Chi-selector, Recursive Feature Elimination (rfe) selector, Embedded logistic regression selector	7

diagnosis of heart disease by incorporating the predictive models into their electronic health records (EHR) systems for real-time risk assessments during patient encounters. They can use these assessments to guide diagnosis, treatment planning, patient monitoring, and direct resource allocation. Clinicians can also encourage patients to adopt mitigation strategies by counselling them about the risks of heart disease. This seamless integration of predictive models into clinical workflows empowers healthcare professionals to make data-informed decisions, ultimately improving patient outcomes and reducing the burden of heart disease.

3. Proposed system

The proposed system uses the Boruta feature selection algorithm to obtain the pertinent features from the Cleveland Clinic Heart Disease dataset acquired from Kaggle, which is a standard benchmark dataset. This dataset has been extensively used in numerous research papers and serves as a common reference point for evaluating similar systems. The proposed system's workflow is shown in Fig. 1.

3.1. Feature selection

To find the set of most pertinent features or variables that contribute to forecast accuracy, feature selection is essential to machine learning-based prediction models (Fig. 2). Boruta has drawn attention among the many feature selection algorithms because of its capacity to manage high-dimensional datasets and find statistically important features.

The Boruta method, a variation of the Random Forest algorithm, evaluates the relevance of each feature using statistical testing and shadow characteristics. Boruta distinguishes between essential and irrelevant features by evaluating the significance of each original feature in relation to its corresponding shadow feature. In this study, the top 6 features were used to predict heart disease, which includes age, type of chest pain (cp), maximum heart rate achieved (thalach), ST depression induced by exercise relative to rest (oldpeak), number of major vessels (0–3) colored by flourosopy (ca), and displays the thalassemia (thal).

By using machine learning algorithms like Boruta on these datasets, it may be possible to find important parameters linked to heart disease and boost the precision of predictive models.

The Boruta algorithm proceeds as follows:

1. Include all of the original features to begin the feature set.
2. By randomly permuting the values of the original features, create shadow features.
3. Train a machine learning model using both the original and shadow features combined.
4. By contrasting it with the relevance of its shadow aspects, evaluate the significance of each original feature. Numerous metrics, including Gini impurity, information gain, and feature weights, can be used to quantify relevance.
5. Using a predetermined threshold, decide whether a characteristic is statistically significant or not. A feature is deemed statistically significant if its importance exceeds the sum of the importance of its shadow characteristics.
6. Eliminate unnecessary features from the feature list.
7. Continue with steps 2 through 6 until all features have been accepted or denied.

Shadow features are created by randomly permuting the values of the original features while keeping the target variable intact. By comparing the importance of each original feature with its corresponding shadow feature, Boruta determines if a feature is statistically significant or not. At the end of the Boruta algorithm, the confirmed features are considered appropriate for the prediction task, while the rejected attributes are deemed unimportant.

Boruta has several advantages. It can handle various types of machine learning models, including both tree-based and non-tree-based algorithms, as well as interactions and redundancy among features effectively. Furthermore, Boruta is relatively robust to noise and can provide stable results even with noisy datasets. However, it is important to note that Boruta may not always identify the absolute best subset of features, as its performance can be influenced by the choice of the importance threshold and quality of the feature importance metric used. Additionally, Boruta can be computationally costly, particularly for datasets with a lot of features.

It is a useful tool for minimizing the degree of dimensionality of datasets and improving the interpretability and performance of machine learning models. In this research, we provide a thorough investigation of machine learning-based heart disease prognostication employing Boruta feature selection. The aim of this work is to establish a dependable and effective prediction model that can help with heart disease early diagnosis and risk assessment. We utilized a dataset with a variety of clinical and demographic information, allowing for an in-depth examination of potential risk factors. A brief comparison of various feature selection algorithms is shown in Table 1.

3.2. Decision tree

Based on a set of training data, this supervised learning technique builds a tree-like model, where each non-leaf node, branch and leaf node represent an attribute, decision rule, and predicted value or class label, respectively. The splitting process is stopped when an obvious terminating condition is met, such as attaining a maximum depth, obtaining a minimum number of samples per leaf, or increasing the number of splits does not increase the predicted accuracy.

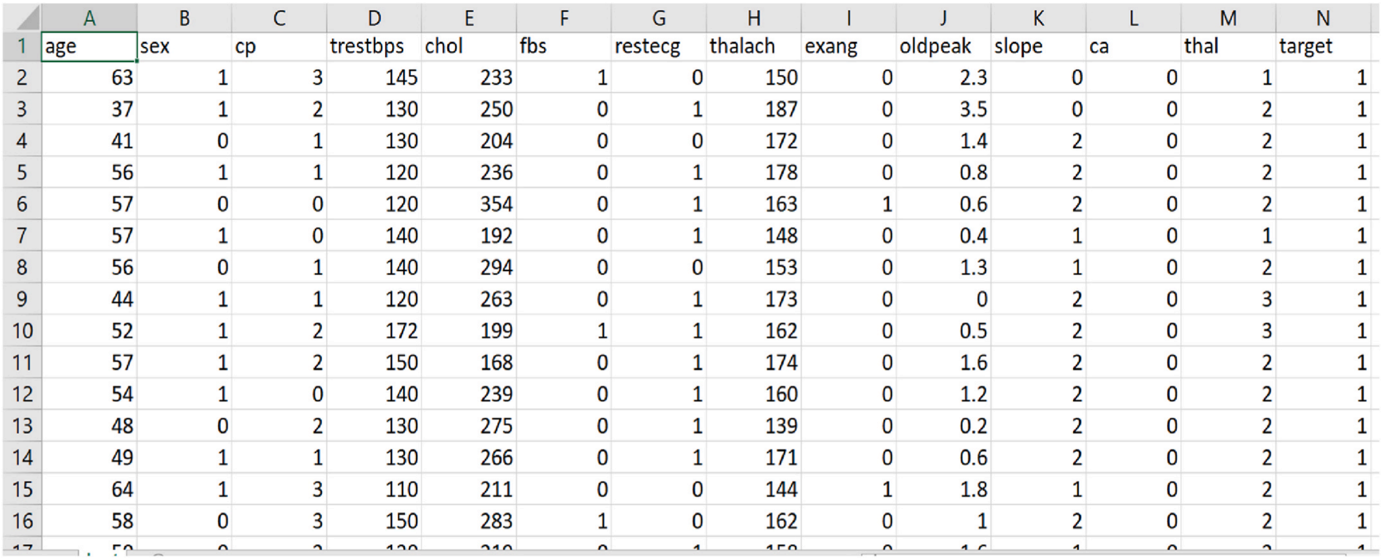
The decision tree can then be used to make predictions on unobserved or test data by moving down the tree, depending on the input instance's feature values, until a leaf node is reached. While generating homogenous subsets at each internal node, the decision tree algorithm seeks to minimize the entropy and Gini impurity, which are indices of impurity or disorder within a set of samples.

Gini impurity is a measurement of impurity or disorder present in a group of samples that is frequently employed in decision tree algorithms as a splitting criterion. It is calculated using the following formula:

$$\mathrm{Gini}(p) = 1\text{-}\Sigma\left(\mathrm{pi}^2\right), \tag{1}$$

where Gini(p) represents the Gini impurity for a set of samples; and pi is the likelihood of that instance falling under a specific class.

Boruta uses the decision tree classifier to evaluate the significance of chosen features by contrasting them with shadow features that were created at random. This evaluation process, which involves stopping criteria during tree construction and making predictions on test data, helps identify key elements for heart disease prediction, improving the accuracy and interpretability of the final model. In order to avoid overfitting, the decision tree has been constructed using training data and criteria such as minimum samples per leaf and maximum depth. By modifying decision rules and values in forecasts, the model improves its



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
2	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
3	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
4	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
5	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
6	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
7	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
8	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
9	44	1	1	120	263	0	1	173	0	0	2	0	3	1
10	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
11	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
12	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
13	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
14	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
15	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
16	58	0	3	150	283	1	0	162	0	1	2	0	2	1

Fig. 2. Cleveland Clinic Heart Disease Dataset.

generalisation and more closely matches the supervisor's preferences.

3.3. Logistic regression

When used in supervised learning, the logistic regression technique first creates a linear model from the supplied training data. By modifying weights and biases to reduce a cost function, the model learns the correlations between the independent and the dependent variables. This is done to predict the dependent variable by examining numerous independent factors, thereby obtaining the target variable results.

With the help of a sigmoid function, the logistic regression model translates the results of this linear model into values between 0 and 1. This sigmoid function guarantees that the predictions of the model represent probabilities.

It is a statistical approach that studies the causal relationship between a collection of binary variable dependencies and a group of autonomous variables. It is an efficient decision-making instrument.

The Logistic regression algorithm first trains a linear model with the given data, whose outputs are then mapped onto values that fall between 0 and 1 using a sigmoid function:

$$z = b + w_1x_1 + w_2x_2 + w_3x_3 + \ldots.. + w_nx_n \tag{2}$$

The sigmoid function is applied to the output of a linearly trained learning model, as follows:

$$y' = \left(\frac{1}{1 + e^{-z}}\right) \tag{3}$$

The terms of LR are described below.

Logistic function: The logistic function shows the correlation between the variable dependencies and autonomous variables. The input values are mapped to a probability value that always lies within the range of 0–1.

Odds: Ratio of a particular event taking place to the same event not taking place.

Log - odds: Natural log of the ratio of an event occurring to the same event not happening.

$$z = \log\left(\frac{y}{1 - y}\right) \tag{4}$$

Maximum likelihood estimation: The procedure for estimating the logistic regression model's coefficients enhances the likelihood of actually obtaining the data given by the model.

By fine-tuning pertinent variables like these, the system can produce predictions that are more legitimate.

- Learning Rate (α): Modify the gradient descent step size to strike a balance between rapid convergence and overshooting.
- Regularisation Strength (λ): To balance overfitting and model complexity, adjust λ for L1 or L2 regularisation.
- Number of Iterations (Epochs): Adjust the training iterations to guarantee optimal model convergence without going overboard with overfitting.

3.4. Support vector machine

Support vector machine (SVM) trains the problem on labeled data to find an optimal hyperplane that best segregates the data points into two classes, expanding the margin between them. This hyperplane is positioned in a high-dimensional feature space, and with supervised learning, which involves adjusting parameters to minimize misclassifications while considering a penalty parameter (C) which is responsible for achieving a wider margin and allowing some misclassification. SVM takes into account misclassifications during the fine-tuning phase. In order to improve model accuracy, these misclassifications are analyzed and minimized during the evaluation process. It evaluates the sensitivity of decision boundaries to changes in input data. This analysis ensures that the model is not overly sensitive to noise or outliers, aligning decisions with the supervisor's desired robustness.

This algorithm is majorly employed for binary classification, where the objective is to divide data items into two classes according to their attributes. Points are positioned to represent the data being learned in a high-dimensional feature space, with each attribute corresponding to a particular attribute or feature of the data points. After that, the algorithm locates a hyperplane, a decision boundary that maximally divides the data points into distinct classes, which enhances generalization and performance on untested data.

Since data might not always be properly separated by a hyperplane in real-world situations, y including a penalty parameter C in the optimization objective function, Soft Margin SVM permits some misclassification. Soft Margin SVM's objective function can be translated as follows:

Minimize subject to the condition that for all i, y_i * (wT * x_i + b) ≥ 1-_i, (5)

where:

Table 2
Metrics of various algorithms with and without Boruta Feature Selection.

Classifier algorithm	Without feature selection				With Boruta feature selection			
	Accuracy	Precision	F1 score	Recall	Accuracy	Precision	F1 score	Recall
Decision tree	75.41	84.00	73.68	65.62	80.33	85.71	80.00	75.00
Logistic regression	88.52	87.88	89.23	90.62	88.52	87.88	89.23	90.62
Support Vector Machine	81.97	81.82	83.08	84.38	83.61	82.35	84.85	87.50
Random forest	83.61	84.38	84.38	84.38	80.33	77.78	82.35	87.50
XGBoost	81.97	86.21	81.97	78.12	77.05	76.47	78.79	81.25

Table 3
Comparison of related works.

Reference number	Method	Accuracy
15	ANN	80 %
10	XGBoost	85.68 %
	Logistic regression	84.46 %
11	MLP with backpropagation technique	80.99 %
14	LR, Deep learning, RF, DT, SVM, Gradient boosting	84 %
20	An effective convolutional neural network (CNN)	85.70 %
21	A stacked model involving KNN, random forest classifier, and SVM	75.1 %
24	Random forest algorithm	83 %
Proposed approach with Boruta Feature Selection	**LR**	**88.52 %**
	SVM	83.61 %
	DT	80.33 %

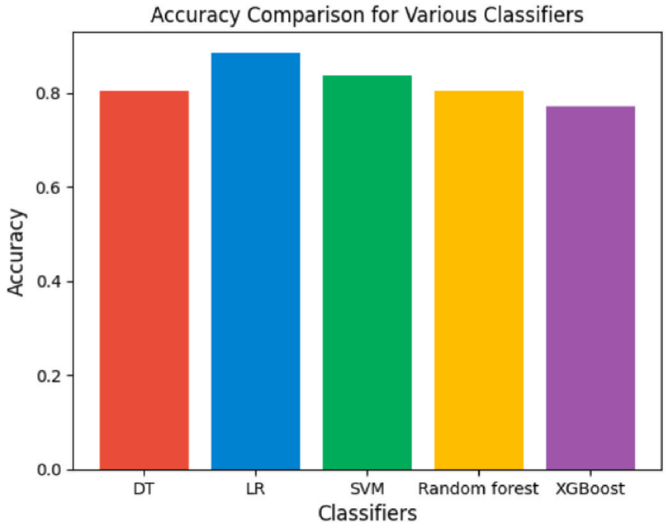


Fig. 3. Accuracy comparison of various algorithms.

- ||w|| is the Euclidean norm of the weight vector w.
- The trade-off between achieving a higher margin and permitting misclassifications is controlled by the penalty parameter C.
- Slack variables like _I allow for classification errors. Each instance's level of misclassification is quantified.
- The i-th training instance's feature vector and matching class label are denoted by the variables x_i and y_i, respectively.

By minimizing this objective function, SVM finds the optimal hyperplane that maximizes the margin while considering the trade-off between misclassifications and the margin width.

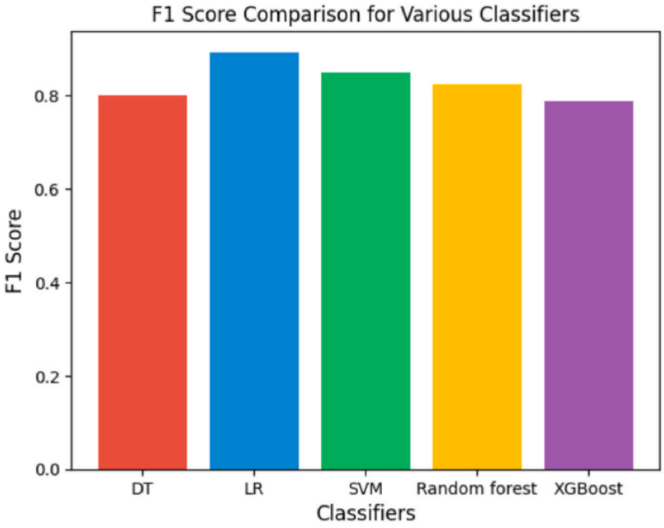


Fig. 4. F1 score comparison of various algorithms.

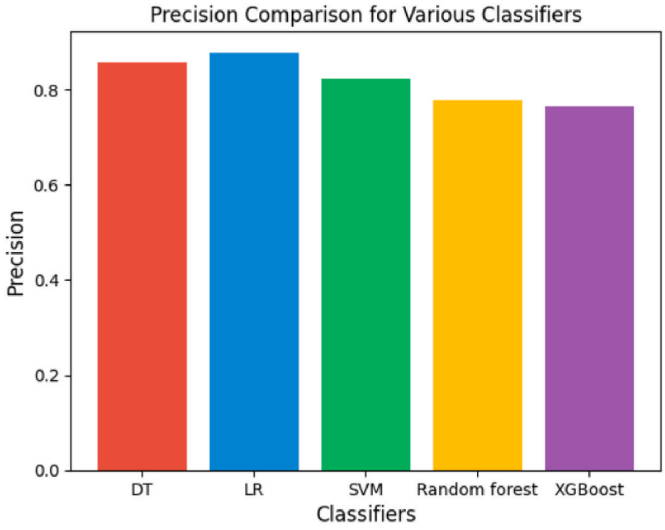


Fig. 5. Precision comparison of various algorithms.

4. Experimental analysis

4.1. Dataset description

As mentioned, the Cleveland Clinic Heart Disease Dataset, collected via Kaggle, was applied in this work [28]. The dataset contains 303 instances with 14 attributes, encompassing clinical factors like AGE, SEX, CP (type of chest pain), and TRESTBPS (resting blood pressure); routine test data like CHOL (serum cholesterol in mg/dl), FBS (fasting blood sugar), and RESTECG (resting electrocardiographic results); and

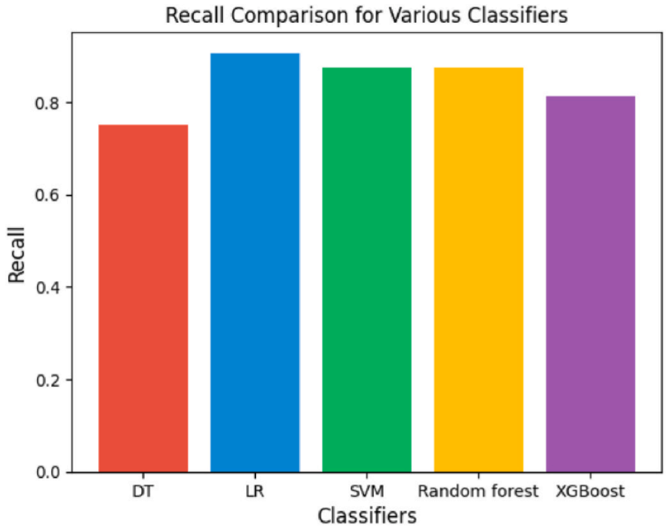


Fig. 6. Recall comparison of various algorithms.

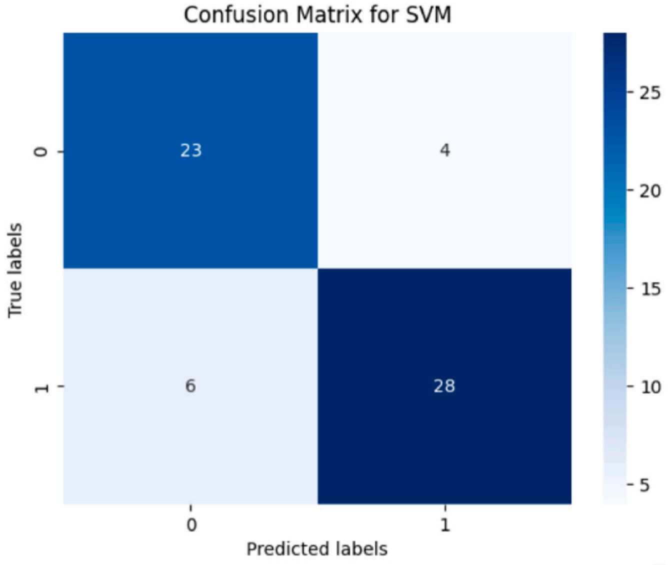


Fig. 7. Confusion matrix of SVM before feature selection.

the exercise electrocardiography test with features like THALACH (maximum heart rate achieved), EXANG (exercise-induced angina), SLOPE (the slope of the peak exercise ST segment), and OLDPEAK (ST depression induced by exercise relative to rest), THAL and NUM. The dataset has no null values, so preprocessing was not required. The snapshot of the dataset is shown in Fig. 2.

4.2. Evaluation process

A confusion matrix, also called an error matrix, is a visual representation of a model's performance that displays the tally of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) results. A confusion matrix serves as a comprehensive overview of a classification model's performance. Various performance metrics, including F1 score, recall (sensitivity), specificity, accuracy, and precision, are calculated using a confusion matrix. These metrics aid in determining the types of errors and accuracy of model in classifying instances of various classes.

Accuracy evaluates how well the model's predictions perform overall:

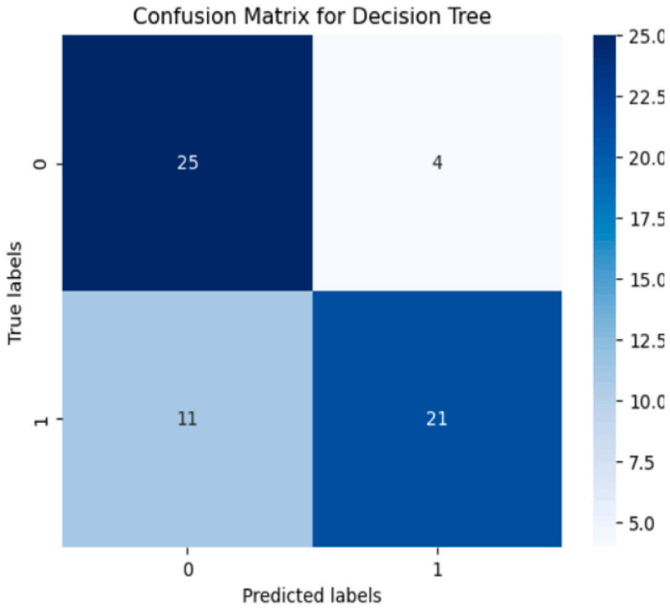


Fig. 8. Confusion matrix of decision tree before feature selection.

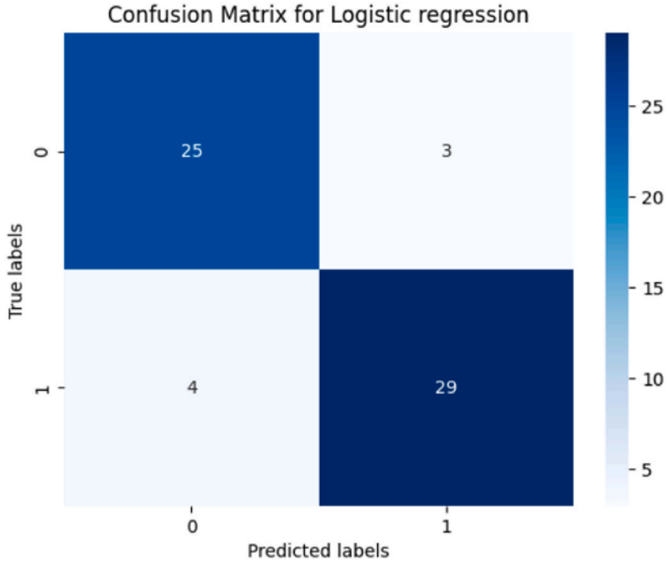


Fig. 9. Confusion matrix of logistic regression before feature selection.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{6}$$

Precision can be described as the relative number of accurately predicted positive instances to all positive instances that were projected:

$$Precision = \frac{TP}{(TP + FP)} \tag{7}$$

Recall is the share of accurately predicted positive instances out of all actual positive instances:

$$Recall = \frac{TP}{(TP + FN)} \tag{8}$$

F1 score is a rendition statistic that balances precision and recall. By combining how precise the model is and recall into a single quantity, it provides a precise assessment of the model's correctness:

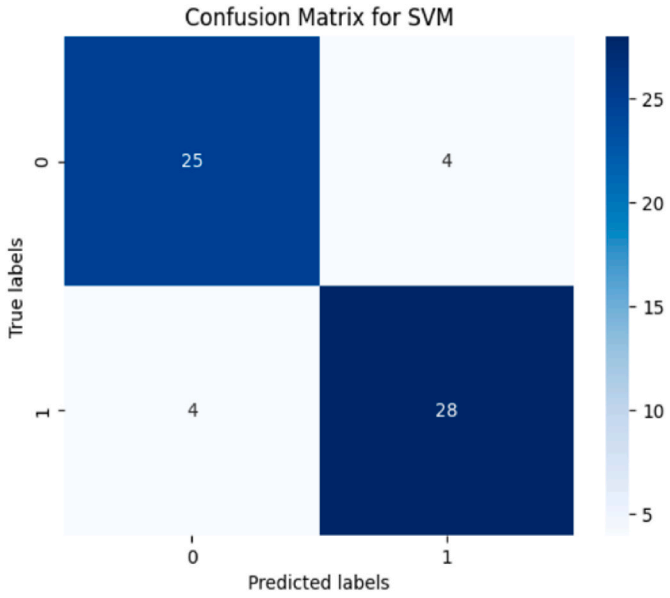


Fig. 10. Confusion matrix of SVM with Boruta feature selection.

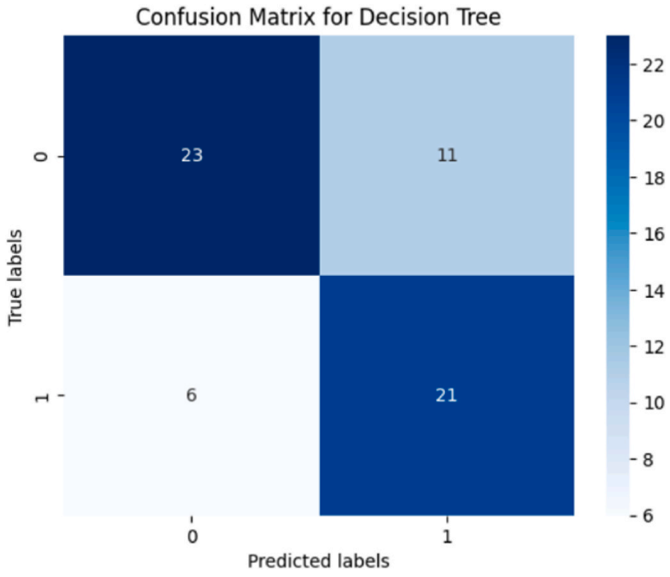


Fig. 11. Confusion matrix of decision tree with Boruta feature selection.

$$F1 = \frac{2 * (precision * recall)}{(precision + recall)} \tag{9}$$

Table 2 displays the performance metrics of various classification algorithms with and without Boruta feature selection algorithm, which aids in locating pertinent features for prediction (Table 1). Accuracy is expressed as a percentage of examples that were properly classified by each algorithm as shown in Table 3.

Based on the results, it is apparent that the performances of Decision Tree and Support Vector Machine algorithms were all enhanced by Boruta feature selection. The performance of Logistic Regression, which previously demonstrated great accuracy and balanced metrics before feature selection, was not considerably impacted. While the performance of Random Forest and XGBoost are diminished by Boruta feature selection.

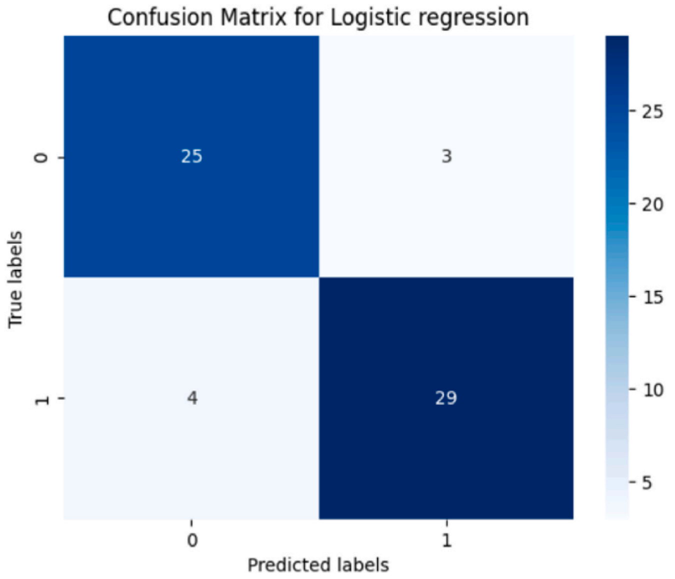


Fig. 12. Confusion matrix of logistic regression with Boruta feature selection.

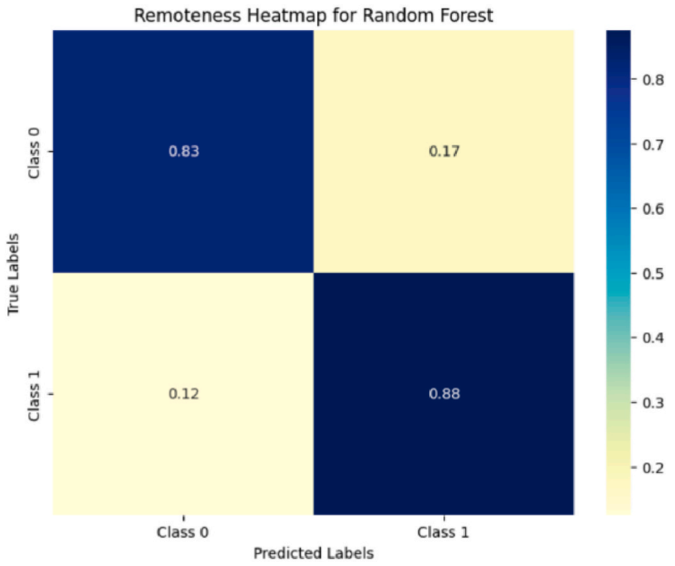


Fig. 13. Remoteness heatmap for random forest.

Fig. 3 demonstrates that among the different algorithms, logistic regression has the optimum accuracy. From Fig. 4, it is clear that logistic regression also has the best F1 score. Moreover, Figs. 5 and 6 further confirm the LR also achieved the highest precision recall, respectively. According to these results, logistic regression gives the best values, followed by SVM for all metrics except precision (Fig. 7).

4.3. Results and discussion

The outcomes emphasize the significance of ML algorithms for CVD prognosis. The implementation of the Boruta feature selection algorithm proved to be beneficial in enhancing the performance of the models. Selecting the pertinent features from the dataset helped improve the evaluation metrics of the models. Figs. 8–12 show the obtained confusion matrix with and without feature selection. The discussion should include how these machine-learning models can be employed in a clinical setting. It is crucial to consider how these models can assist healthcare professionals in making decisions. They can help with risk stratification by assisting medical practitioners in selecting individuals who require a more complete assessment and early intervention. In the context of clinical application, it is crucial to emphasize that the

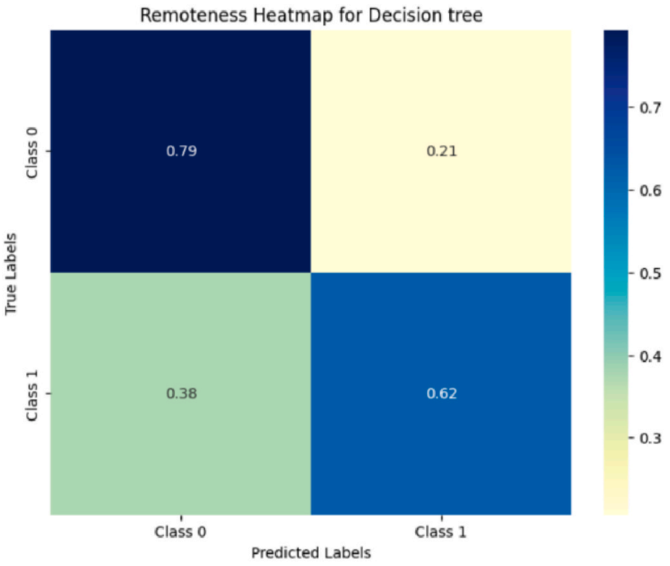


Fig. 14. Remoteness heatmap for decision tree.

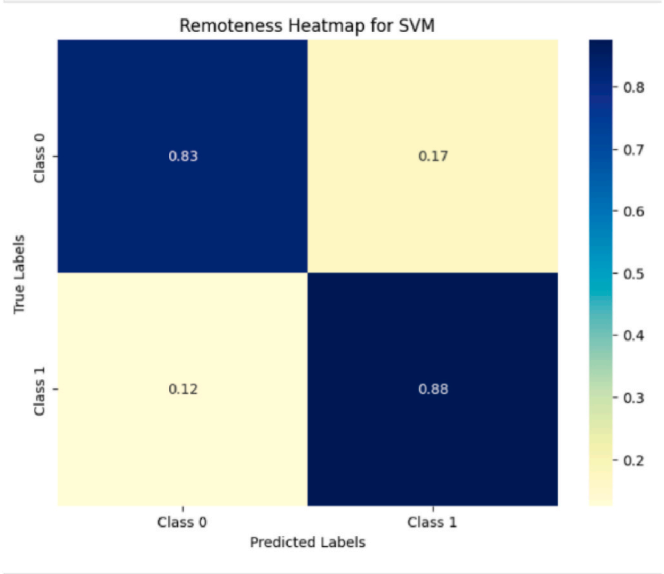


Fig. 15. Remoteness heatmap for SVM

algorithmic output should not be taken at face value. In order to ensure these models remain reliable and effective in a variety of clinical circumstances, their responsible use necessitates regular supervision and reevaluation of the rules that are produced from the learning set.

Here, we have performed the Kruskal-Wallis test, a non-parametric statistical test that is employed to determine if there are statistically significant differences among Machine learning techniques. Figs. 13–17 show the obtained Remoteness Heatmap for various algorithms. Figs. 18–22 highlight the fetched calibration plot and Figs. 22–27 reveal the SHAP Plot for the used approaches. In this case, this test is considered for the three non-ensemble methods discussed earlier. The outcome of this test is a p-value(performance), from which it can be inferred that there is no significant difference among the algorithms since the dataset is concise and algorithms used here are supervised (efficient) in a way they produce almost similar results.

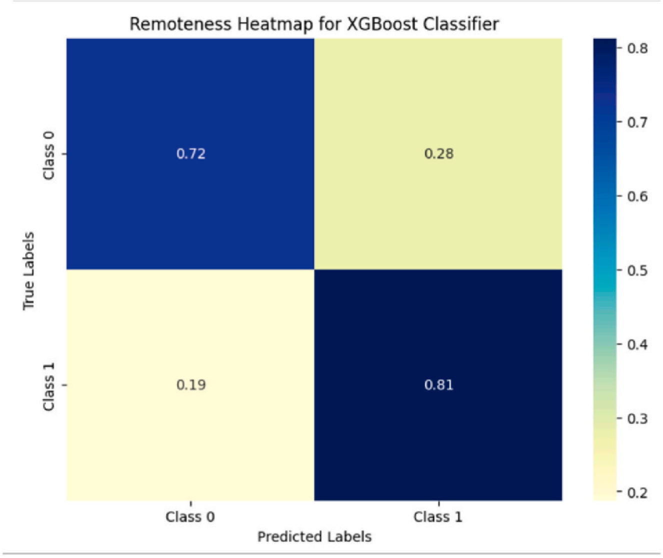


Fig. 16. Remoteness heatmap for XGBoost.

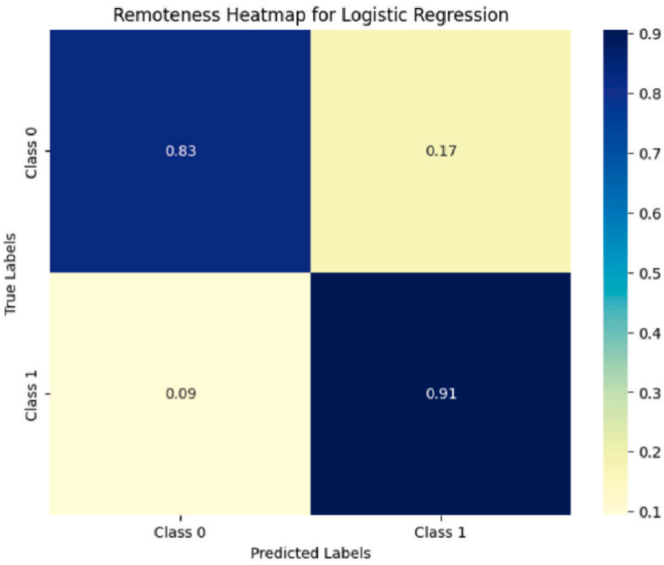


Fig. 17. Remoteness heatmap for logistic regression.

5. Conclusion

This work aimed to compare various machine learning algorithms to achieve the precise prognosis of heart diseases. The Boruta feature selection algorithm was used to enhance the results of the compared machine learning models, which were trained on the selected features. The accuracy, precision, recall and F1 scores of each model were evaluated, and their confusion matrices were studied. Regarding performance, logistic regression exhibited the highest accuracy, decision tree and support vector machine algorithms performed better when Boruta feature selection was employed. The comparatively modest dataset size is one of this study's limitations. Although beneficial, the Cleveland Clinic Heart Disease Dataset, which consists of 303 records and 14 attributes, may have limits in terms of generalizability. We can extend the work in the future by employing a diverse dataset with more instances and attributes. Future studies are suggested to use various feature selection strategies. Additionally, such works can be extended to cover diagnosing and classifying additional illnesses, such as diabetes and cancer.

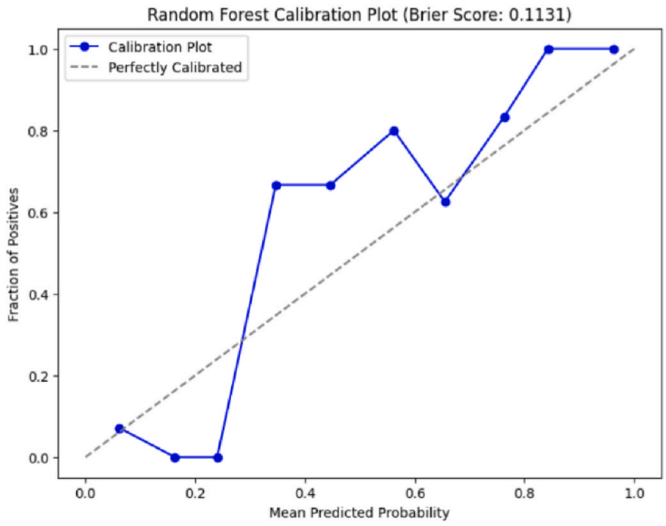


Fig. 18. Calibration plot for Random forest.

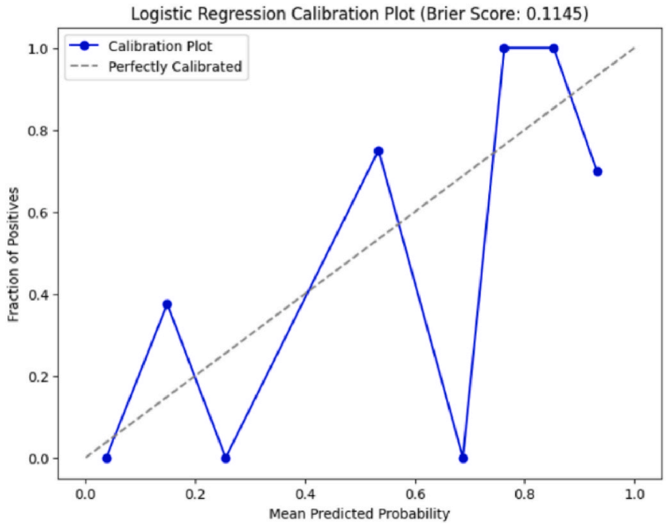


Fig. 19. Calibration plot for Logistic Regression.

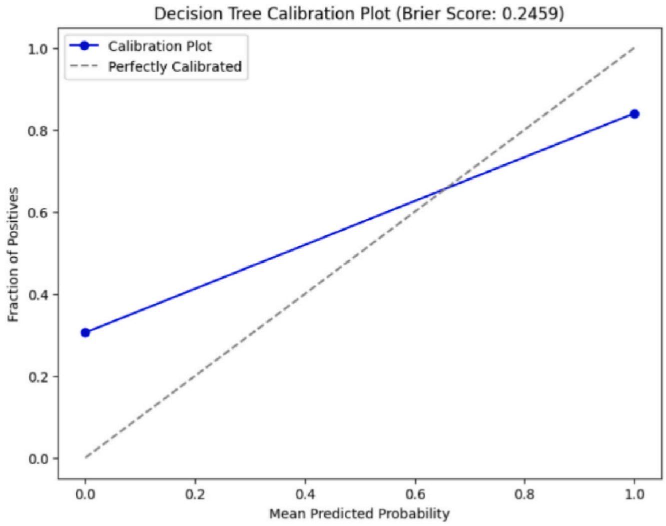


Fig. 20. Calibration plot for Decision tree.

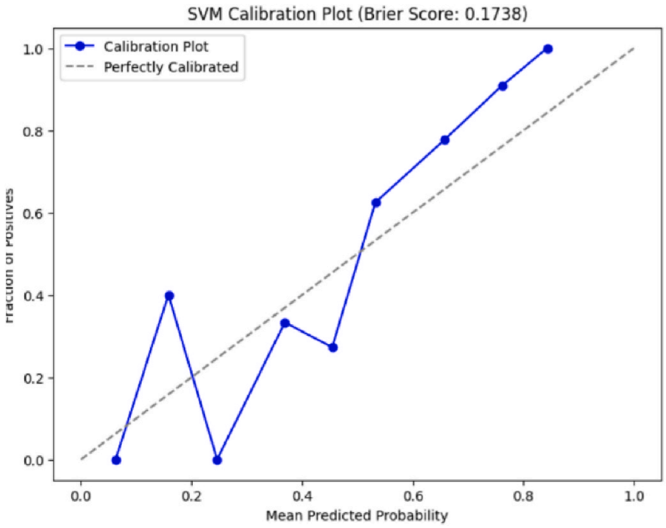


Fig. 21. Calibration plot for SVM.

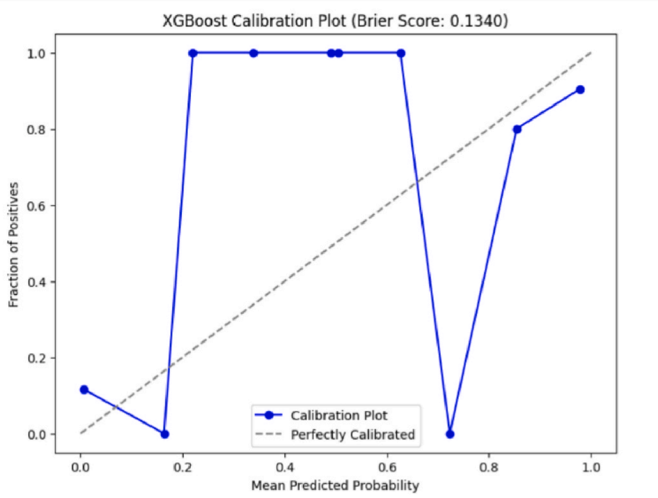


Fig. 22. Calibration plot for XGBoost.

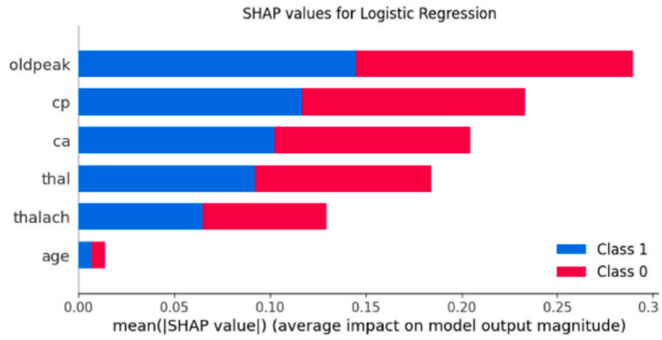


Fig. 23. SHAP plot for Logistic Regression.

Funding

Open Access funding is provided by Obuda University.

Data availability

The link is shared in the article.

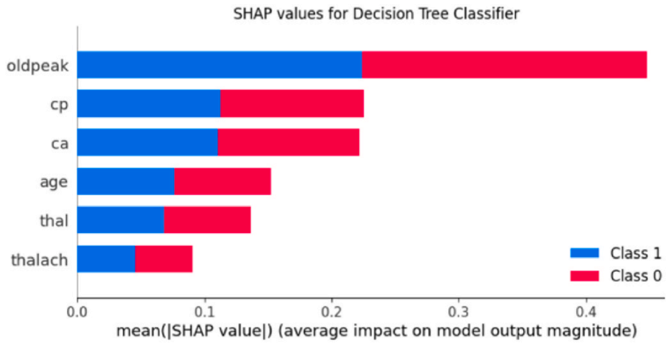


Fig. 24. SHAP plot for Decision tree.

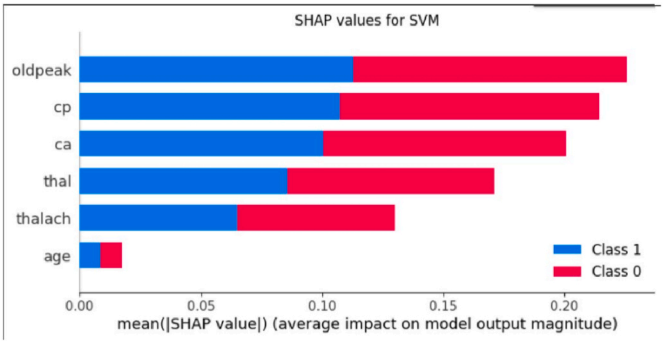


Fig. 25. SHAP plot for SVM.

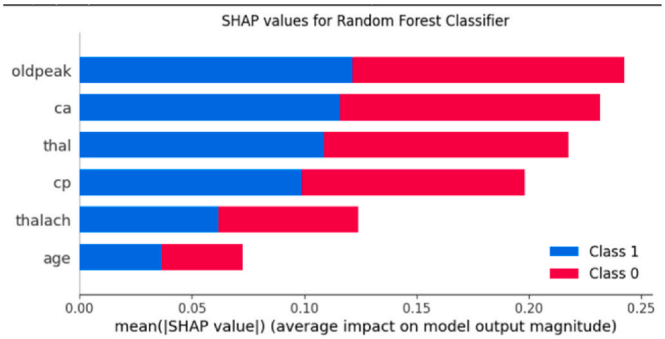


Fig. 26. SHAP plot for Random forest.

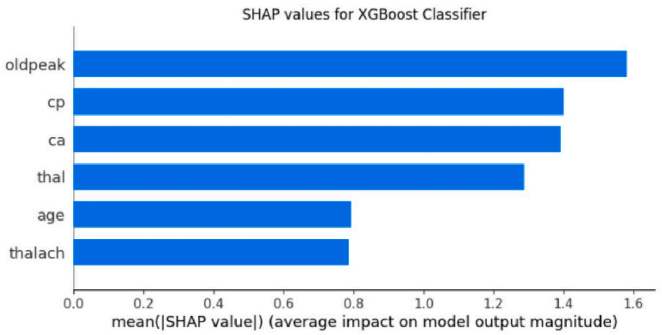


Fig. 27. SHAP plot for XGBoost.

CRediT authorship contribution statement

G. Manikandan: Writing – original draft. **B. Pragadeesh:** Conceptualization. **V. Manojkumar:** Visualization. **A.L. Karthikeyan:** Methodology. **R. Manikandan:** Validation. **Amir H. Gandomi:**Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

I would congratulate the authors for having a good spirit and providing their valuable time over this research. I would also thank Informatics in Medicine Unlocked for providing us the platform.

References

[1] Seckeler MD, Hoke TR. The worldwide epidemiology of acute rheumatic fever and rheumatic heart disease. Clin Epidemiol 2011:67–84.
[2] Gaziano TA, Bitton A, Anand S, Abrahams-Gessel S, Murphy A. Growing epidemic of coronary heart disease in low-and middle-income countries. Curr Probl Cardiol 2010;35(2):72–115.
[3] Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine learning improve cardiovascular risk prediction using routine clinical data? PLoS One 2017;12(4):e0174944.
[4] Ramalingam VV, Dandapath A, Raja MK. Heart disease prediction using machine learning techniques: a survey. Int J Eng Technol 2018;7(2.8):684–7.
[5] Patel J, TejalUpadhyay D, Patel S. Heart disease prediction using machine learning and data mining technique. Heart Dis 2015;7(1):129–37.
[6] Pham V, De Hemptinne Q, Grinda JM, Duboc D, Varenne O, Picard F. Giant coronary aneurysms, from diagnosis to treatment: a literature review. Arch Cardiovascul Dis 2020;113(1):59–69.
[7] Reddy NSC, Nee SS, Min LZ, Ying CX. Classification and feature selection approaches by machine learning techniques: heart disease prediction. Int J Innov Comp 2019;9(1).
[8] Pillai NSR, Bee KK, Kiruthika J. Prediction of heart disease using rnn algorithm. Int Res J Eng Technol 2019;5.
[9] Raza K. Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule. In: U-healthcare monitoring systems. Academic Press; 2019. p. 179–96.
[10] Nalluri S, Vijaya Saraswathi R, Ramasubbareddy S, Govinda K, Swetha E. Chronic heart disease prediction using data mining techniques. In: Data engineering and communication technology: proceedings of 3rd ICDECT-2K19. Springer Singapore; 2020. p. 903–12.
[11] Bhat R, Chawande S, Chadda S. Prediction of test for heart disease diagnosis using artificial neural network. Indian J Appl Res 2019;9.
[12] Abushariah MA, Alqudah AA, Adwan OY, Yousef RM. Automatic heart disease diagnosis system based on artificial neural network (ANN) and adaptive neuro-fuzzy inference systems (ANFIS) approaches. J Software Eng Appl 2014;7(12):1055.
[13] Hasan TT, Jasim MH, Hashim IA. Heart disease diagnosis system based on multilayer perceptron neural network and support vector machine. Int J Curr Eng Technol 2017;77(55):2277–4106.
[14] Sapra L, Sandhu JK, Goyal N. Intelligent method for detection of coronary artery disease with ensemble approach. In: Advances in communication and computational technology: select proceedings of ICACCT 2019. Springer Singapore; 2021. p. 1033–42.
[15] Chen AH, Huang SY, Hong PS, Cheng CH, Lin EJ. HDPS: heart disease prediction system. In: 2011 computing in cardiology. IEEE; 2011, September. p. 557–60.
[16] Haq AU, Li JP, Memon MH, Nazir S, Sun R. A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Inf Syst 2018;2018:1–21.
[17] Singh A, Kumar R. Heart disease prediction using machine learning algorithms. In: 2020 international conference on electrical and electronics engineering (ICE3). IEEE; 2020, February. p. 452–7.
[18] Sharma S, Parmar M. Heart diseases prediction using deep learning neural network model. Int J Innovative Technol Explor Eng 2020;9(3):2244–8.
[19] Muhammad Y, Tahir M, Hayat M, Chong KT. Early and accurate detection and diagnosis of heart disease using intelligent computational model. Sci Rep 2020;10(1):19747.
[20] Dutta A, Batabyal T, Basu M, Acton ST. An efficient convolutional neural network for coronary heart disease prediction. Expert Syst Appl 2020;159:113408.
[21] Shorewala V. Early detection of coronary heart disease using ensemble techniques. Inform Med Unlocked 2021;26:100655.

[22] Sekar J, Aruchamy P, Sulaima Lebbe Abdul H, Mohammed AS, Khamuruddeen S. An efficient clinical support system for heart disease prediction using TANFIS classifier. Comput Intell 2022;38(2):610–40.

[23] Rajendran R, Karthi A. Heart disease prediction using entropy based feature engineering and ensembling of machine learning classifiers. Expert Syst Appl 2022; 207:117882.

[24] Chang V, Bhavani VR, Xu AQ, Hossain MA. An artificial intelligence model for heart disease detection using machine learning algorithms. Healthcare Analytics 2022;2:100016.

[25] Mohapatra S, Maneesha S, Mohanty S, Patra PK, Bhoi SK, Sahoo KS, Gandomi AH. A stacking classifiers model for detecting heart irregularities and predicting Cardiovascular Disease. Healthcare Analytics 2023;3:100133.

[26] Shehab M, Abualigah L, Shambour Q, Abu-Hashem MA, Shambour MKY, Alsalibi AI, Gandomi AH. Machine learning in medical applications: a review of state-of-the-art methods. Comput Biol Med 2022;145:105458.

[27] Diwan S, Thakur GS, Sahu SK, Sahu M, Swamy NK. Predicting heart diseases through feature selection and ensemble classifiers. In: Journal of physics: conference series. vol. 2273. IOP Publishing; 2022, May, 012027. 1.

[28] Chaurasia V, Chaurasia A. Novel method of characterization of heart disease prediction using sequential feature selection-based ensemble technique. Biomed Mater Dev 2023:1–10.

[29] Ansari GA, Bhat SS, Ansari MD, Ahmad S, Nazeer J, Eljialy AEM. Performance evaluation of machine learning techniques (MLT) for heart disease prediction. Comput Math Methods Med 2023;2023.

[30] Saketha Rama BV, Suryanarayana G, Ansari MD, Begum R. An empirical comparison of classification machine learning models using medical datasets. In: International conference on data science, machine learning and applications. Singapore: Springer Nature Singapore; 2022, December. p. 321–32.

ARTICLES FOR FACULTY MEMBERS

AN IMPROVED MLP-RFE MODEL FOR ENHANCED ACCURACY IN HEART DISEASE PREDICTION USING DEEP LEARNING

Effective heart disease prediction using novel MLP-EBMDA approach / Deepika, D., & Balaji, N.

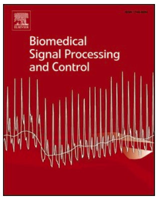
Biomedical Signal Processing and Control
Volume 72 (2022) 103318 Pages 1-9
https://doi.org/10.1016/j.identj.2025.109308
(Database: ScienceDirect)

Biomedical Signal Processing and Control 72 (2022) 103318

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



Effective heart disease prediction using novel MLP-EBMDA approach

Deepika D[b,*], Balaji N.[a]

[a] Information Technology, Karpagam College of Engineering, Coimbatore, India
[b] Research Scholar, Anna University, Chennai, India

ARTICLE INFO

Keywords:
Heart disease prediction
MLP-EBMDA
Optimized unsupervised method

ABSTRACT

Heart disease prediction is more important to prevent the death rate. The death rate increases due to lack of initial detection of heart disease in humans. To predict heart disease in an effective way by using feature selection and classification approach. Thus, an optimized unsupervised technique for feature selection and novel Multi-Layer Perceptron for Enhanced Brownian Motion based on Dragonfly Algorithm (MLP-EBMDA) for classification of heart disease has been proposed in this study. The dataset of the heart disease is obtained as input and pre-processing is performed. Features are selected through the optimized unsupervised technique. Based on the selected features, classification of heart disease is performed using the novel hybrid MLP-EBMDA approach. The analysis of the proposed system with various existing systems in terms of accuracy has explored 94.28%. The analysis of the proposed system in terms of precision has showed 96%, in terms of recall the results of the proposed system has been found to be 96%, in terms of F1-score the results of the proposed system has been found to be 96%. Thus the overall performance analysis of the proposed methodology exhibited efficient results in predicting the heart disease than the various state-of-the-art methods.

1. Introduction

Heart disease has become the major reason for death rate globally. Heart disease prediction is a serious problem. Predicting heart disease is a hard task due to various risk factors that involves high cholesterol and blood pressure, diabetes, unusual pulse rate and several other factors. ML (Machine Learning) techniques have been found to be efficient in predicting heart disease by making decisions from huge data generated from the industry of healthcare. The model for heart disease prediction is presented with varied feature combinations and various classification methods used by conventional methods.

Further, [1] recommended a prediction model for heart disease that comprises of SMOTE-ENN (Synthetic Minority Over-Sampling Technique-Edited Nearest Neighbour), DBSCAN (Density-Based Spatial Clustering of Applications with Noise) and XGBoost to manage the distribution of training data, to identify the outliers as well as to eradicate it and for heart disease prediction. Comparative analysis of the proposed and existing system has been carried out. It has been found that the proposed model performed effectively than the traditional models in terms of accuracy. Several data samples and its comparison is yet to be considered with the extensive medical datasets and the parameters of the model. Moreover this study [2] introduced an innovative method that targets in discovering important features by employing machine learning methods leading to enhanced accuracy in cardiovascular disease prediction. The proposed model has been presented with various known classification methods and varied feature combinations. Thus the proposed model exhibited effective performance in terms of accuracy. Various machine learning methods can be introduced in near future for predicting the disease in a better way.

Optimized unsupervised feature selection technique has been utilized to select the features for heart disease prediction. Further, all the unsupervised and supervised methods have been explored for heart disease prediction via various existing techniques [3].

Few machine learning and data mining methods have been utilized to anticipate the heart disease that involves decision tree, ANN (Artificial Neural Network), KNN (K-Nearest Neighbour), fuzzy logic, logistic regression and SVM (Support Vector Machine). It affords instinct ideas related to the existing systems along with the complete summarization of the existing methods. Classification has been performed by multilayer perceptron incorporated with enhanced Brownian motion on the basis of dragonfly algorithm. Finally performances of the proposed system are analyzed to determine the efficiency in predicting the heart

---

Abbreviation: MLP-EBMDA, Multi-Layer Perceptron for Enhanced Brownian Motion based on Dragonfly Algorithm.

[*] Corresponding author.

E-mail address: deepikaphd11@gmail.com (D. Deepika).

https://doi.org/10.1016/j.bspc.2021.103318
Received 25 February 2021; Received in revised form 30 October 2021; Accepted 2 November 2021

disease.

This study propose an effective system for heart disease prediction on the basis of techniques of machine learning. Prediction of Heart disease in many cases is based on the complex association of pathological and clinical data. Due to this difficulty, there arises an important interest among researchers and clinical professionals for the prediction of heart disease in an accurate and efficient manner. This helps in heart disease prediction in the initial stage of the disease occurrence. Thus the accuracy in predicting the heart disease as normal or abnormal is enhanced via the effective feature selection and classification techniques on the basis of the proposed approaches.

The major contribution of this study involves the following:

- To enhance the accuracy of heart disease prediction at the initial stage by utilizing effective feature selection and classification techniques.
- To execute the optimized unsupervised feature selection technique for effective feature selection.
- To implement the hybrid multilayer perceptron incorporated with enhanced Brownian motion depended on dragonfly algorithm (MLP-EBMDA) for effective heart disease classification and prediction.
- The performance analysis of the proposed system is implemented to assess the efficiency in predicting the heart disease.

The upcoming Section 2 describes the reviews of various existing works related to the prediction of heart disease using various techniques. Section 3 explores the proposed work that involves the use of optimized unsupervised method for feature selection and hybrid multilayer perceptron integrated with MLP-EBMDA for classification of the heart disease. Results of the proposed system is explored in Section 4. The final conclusion of the entire proposed system is given in Section 5.

2. Review of existing works

The following section represents the literature review of existing works that has utilized various techniques for prediction of heart disease.

2.1. Classification models for heart disease prediction

This study [4] developed a recognition system for heart disease by utilizing HOBDBNN (Higher Order Boltzmann Deep Belief Neural Network) on the basis of medical IoT device. The efficiency of the proposed model has been ensured via efficient use of activation and learning function. The proposed system is assessed in terms of efficiency by utilizing a simulation tool (MATLAB) to guarantee maximum accuracy for recognition and minimum loss value. Thus the proposed system has to be enhanced further in near future by utilizing efficient processes for feature selection and optimization techniques to handle the progress and database of the disease.

Furthermore [5] proposed a clustering algorithm in terms of data mining for prediction of heart disease. The existing algorithms in the clustering methods have been compared to accomplish the maximum accuracy in prediction of heart disease. The dataset has been pre-processed by various unsupervised and supervised algorithms to improve the solution's accuracy. Based on the results, Simple k-means, filtered cluster and make density based cluster have been found to be the efficient algorithms and also serves as a theoretical tool for doctors to efficiently predict the critical cases.

Thus this paper [6] introduced a diagnostic system for heart disease. It has utilized DNN (Deep Neural Network) for the purpose of classification and X^2 Statistical model for the purpose of refinement of the features. The proposed system has been evaluated by utilizing six varied metrics that includes sensitivity, accuracy, AUC, specificity, ROC and MCC (Matthews Correlation Coefficient). The proposed method has also been compared with the existing techniques. Thus it has explored that the proposed system can enhance the decision making quality at the time of heart disease diagnosis. There is a drawback in this study which has been found that the individual hidden layer's optimal width in DNN and ANN (Artificial Neural Network) model is examined by utilizing grid search algorithm. Fast and sophisticated algorithms such as genetic algorithm has yet to be used.

This study aimed [7] for structuring two varied classification models based on ANFIS (Adaptive Neuro Fuzzy Inference System) to predict heart disease. It has revealed that the classifiers learnt the way of dataset classification. The performance of the proposed method has been assessed for classification in terms of testing, accuracy and training. Thus the proposed model can be improved further by employing several derivation in jacobian matrix computation so as to improve the result's convergence speed. In addition [8] introduced a healthcare system for prediction of heart disease by utilizing feature fusion and ensemble deep learning methods. The proposed model has been trained for heart disease prediction. The proposed methodology has been assessed with the data pertaining to heart disease and this data has been compared with conventional classifiers on the basis of feature selection, weighting methods and feature fusion. The proposed system exhibited high accuracy than the traditional methods. Data mining methods is yet to be utilized to improve the feature fusion performance to afford a refined dataset for diagnosis of heart disease.

2.2. Feature selection for heart disease prediction

This study [9] aimed to develop an optimization function based on SVM (Support Vector Machine). GA (Genetic Algorithm) utilizes this objective function for choosing the important features to predict the heart disease. The proposed results have been compared with various existing methods. Thus the comparative analysis revealed that the proposed system explored effective results than the state-of-the-art methods. Similarly, this study [10] intended to anticipate the heart disease that guides doctors in heart disease diagnosis at initial stages. The research has been carried out on publicly available datasets of the heart disease. The analysis revealed that the proposed system has performed better than the existing systems. Thus the proposed system can be tested and developed by utilizing streaming data in near future.

Additionally, [11] proposed a PSO (Particle Swarm Optimization) on the basis of attribute selection with enhanced Fuzzy ANN classifier for predicting the risk of heart disease. It has been found that the proposed system revealed efficient results in terms of accuracy for all the patients. The proposed system outperformed other methods. Further supplementary algorithms can also be employed for attribute selection and regarded as future work. Moreover [12] recommended an improved data mining algorithm for employing it in healthcare. Prognosis Prediction using RNN (PP-RNN) utilized various RNNs for learning the sequences of the patient's diagnosis code so as to predict the heart disease occurrence. Thus it has revealed that the proposed system exhibited effective results in terms of accuracy. This field has to be improved further in terms of the patient's lifetime as well as the treatments by consistently managing and examining the data of the health sector. This prediction can be improved by utilizing hybrid models.

In addition, this study [13] aimed to anticipate the various types of diseases by utilizing GWO + RNN (Grey Wolf Optimization based on RNN). The results revealed that the proposed system exhibited good results than the state-of-the-art methods. The proposed system showed better prediction results. The productivity of various classification of medical disease has to be enhanced further by using hybrid technique. Further [14] proposed a hybrid technique for heart disease diagnosis. The proposed system has the capability to enhance the neural network performance by improving its preliminary weights by utilizing genetic algorithm that recommends neural network with better weights. Thus through this proposed system, the rate of sensitivity, accuracy and specificity has been accomplished. This technique can be employed as an

Table 1
Comparative Analysis of various existing works in heart disease prediction.

S. no	References	Objectives	Dataset used	Results Obtained	Advantages/Disadvantages
1.	[6]	The study intended to introduce a diagnostic system for heart disease. It has utilized DNN (Deep Neural Network) for the purpose of classification and X^2 Statistical model for the purpose of refinement of the features.	Cleveland heart disease dataset is used.	93.3% accuracy	The results has explored that the proposed system can enhance the decision making quality at the time of heart disease diagnosis with an accuracy 93.3%. There is a drawback in this study which has been found that the individual hidden layer's optimal width in DNN and ANN (Artificial Neural Network) model is examined by utilizing grid search algorithm. Fast and sophisticated algorithms such as genetic algorithm has yet to be used.
2.	[7]	This study aimed for structuring two varied classification models based on ANFIS (Adaptive Neuro Fuzzy Inference System) to predict heart disease.	Statlog-Cleveland heart disease dataset is used.	accuracy rate of 75% -training data and 76.6% – testing data	The proposed model can be improved further in terms of accuracy as it shows accuracy rate of 75% in training data and 76.6% in testing data. Several derivation must also be employed in jacobian matrix computation so as to improve the result's convergence speed.
3.	[10]	The paper intended to anticipate the heart disease that guides doctors in heart disease diagnosis at initial stages through the use of hybrid fuzzy logic classifier and genetic algorithm.	UCI Heart disease datasets are used.	90% accuracy – Cleveland dataset, 91% accuracy -Hungarian dataset and 89% accuracy – Switzerland dataset.	Proposed system showed 90% accuracy in Cleveland dataset, 91% accuracy in Hungarian dataset and 89% accuracy in Switzerland dataset. However accuracy has to be further enhanced.
4.	[11]	This study proposed a PSO (Particle Swarm Optimization) on the basis of attribute selection with enhanced Fuzzy ANN classifier for predicting the risk of heart disease.	Dataset is taken from heart based health centres.	88.82% & 88.05% accuracy for male and female	88.82% accuracy has been attained to predict males affected with the disease and 88.05% accuracy for predicting females. Accuracy must be improved to enhance the prediction rate.
5.	[23]	Machine learning algorithms such as Random Forest (RF), ANN (Artificial Neural Network), K-NN (K-Nearest Neighbour), NB (Naïve Bayes), DT (Decision Tree), LR (Logistic Regression) and SVM (Support Vector Machine) have been employed to predict heart disease.	Cleveland heart disease dataset is used.	89% accuracy	LR performed better with accuracy 89%. However, accuracy has to be further enhanced.
6.	[2]	Aimed to introduce an innovative method that targets in discovering important features by employing machine learning methods leading to enhanced accuracy in cardiovascular disease prediction.	Cleveland heart disease dataset is used.	88.7% accuracy	The proposed model exhibited effective performance in terms of accuracy with 88.7% when employing RF with Linear Model (LM). Enhancing accuracy rate will assist in efficient heart disease prediction.

alternative for genetic algorithm as it can improve the proposed method's performance.

2.3. Prediction of heart disease by various other techniques

Diseases related to heart or CVDs (Cardiovascular diseases) have become the major reason for large counts of death over last few years. It has also evolved as the major disease that threatens the life of people in the entire universe. Thus in this study [15], examined the data mining applications in medical domain along with few methods utilized in prediction of the disease. It has been found that the outcomes may differ for varied disease diagnosis on the basis of methods and tools utilized. Thus data mining affords efficient outcome in diagnosis of disease when suitable methods and tools are employed.

On the other hand, algorithms of machine learning [16] has been utilized for the heart disease prediction. Deep learning has become an evolving domain of AI (Artificial Intelligence) and explored few favourable outcomes in supplementary medical diagnosis fields with maximum accuracy. Few deep learning techniques have been discussed that can be employed for prediction of heart disease with machine learning algorithms. Comparative analysis has been performed to determine the efficient algorithm for dataset pertaining to medicine. The work of short-term medical dataset has to be forwarded in near future where dataset differs with time. Dataset also has to be retrained.

Similarly [17] proposed a framework for diagnosis of Heart Failure (HF) by utilizing GRU (Gated Recurrent Units) that are methods of deep learning. Comparative analysis has been performed to assess the proposed system's performance. It has been found that compared to various existing systems the proposed system exhibited efficient performance in HF diagnosis prediction. Applications of healthcare has to be extended further by integrating the knowledge of experts into the proposed framework.

Moreover [18] introduced a robust and effective architecture for prediction of HF by utilizing a neural network. Word vectors and one-hot encoding has been used for diagnosis events modelling and HF prediction by utilizing the fundamental principles of LSTM (Long Short-Term Memory Network) model. The proposed system has been evaluated to determine the performance of the system. Thus the assessment on the basis of dataset in real-world explore the efficiency and favourable use of the proposed system in HF risk prediction. The applications of healthcare has to be extended further by integrating the knowledge of the expert into the proposed system. Thus varied heart disease prediction models has also been investigated in this study [19]. It also chose the significant features for heart disease by utilizing genetic algorithm. The performance of the proposed model has been found to be effective than the state-of-the-art methods. The results revealed that the performance of the proposed model has been improved when genetic algorithm has been integrated with the prediction models.

The malfunctioning of heart due to some blockages or narrowing of arteries supplying blood to heart is known as Coronary heart disease. This may be the major cause for heart attacks. There are many forms and causes for heart diseases. Such heart diseases can be predicted beforehand and preventing major damages happening. This study is referring a multilayer perceptron method for preventing those heart diseases. Taking reference from Framingham Heart study Dataset preprocessing of data using a multilayer perceptron with a deep learning approach has improved the quality of data. This study is having high percentile of detecting Coronary Heart Disease at early stage. The outcome of this

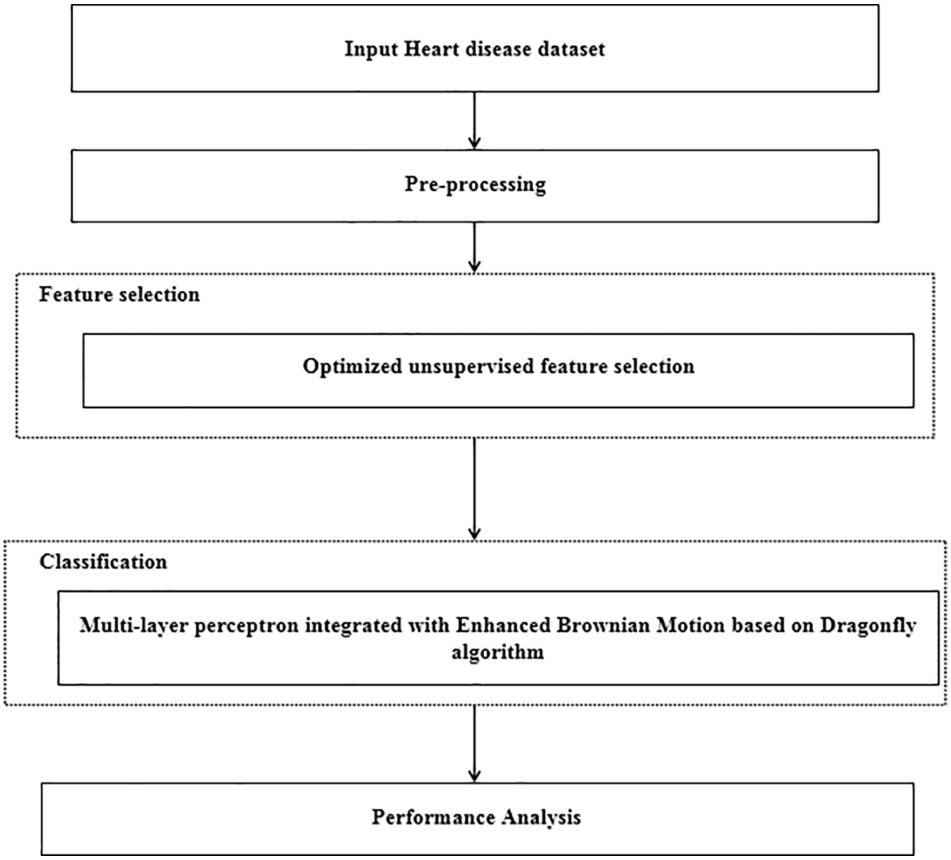


Fig. 1. An overview of the proposed system.

study was accurate up to 96.50% [20]. Siding to the above said work another study using multi-layer perceptron has been analyzed. This work has given Heart disease prediction model (HDPM) for detecting the issue in clinics called as Clinical decision support system (CDSS). The model was implemented in publicly available two data sets and the accuracies were 95.90% and 98.40% in predicting the faults. The model is designed in such a way that heart diseases CDSS helps the physicians to find out the anomalies soon and avoid un necessary complications [1].

Early detection of heart disease can make the success rate to increase in saving a particular patient. This study is introducing a fitness oriented Dragon fly algorithm for selected features. A deep learning is executed to detect the characters of a particular feature selected. The accuracy rate for this algorithm was 6.7% more when compared with particle swarm optimization, 6.96% better than Gray wolf optimization [21].The above methods used for comparison were used to detect heart disease eventually. Similar to the above discussed work, this work is also based on dragon fly algorithm used in detecting heart disease. This algorithm was an effective Swarm optimization algorithms compared with others. They were actively involved in medical field to detect the abnormalities and anomalies found in heart. The outcome was quite impressive helping technicians to find the damages accurately [22]. The comparative analysis of various existing works in predicting the heart disease is shown in the below Table 1.

3. Proposed work

3.1. Novel Multi-Layer Perceptron integrated with Enhanced Brownian Motion based on Dragonfly Algorithm (MLP-EBMDA)

In the proposed work, dataset for heart disease is taken as input. Then pre-processing is performed followed by the feature selection using the optimized unsupervised technique. Based on the selected features, classification was done using the MLP-EBMDA to predict the heart disease in humans. Finally the performance of the proposed methodology has been analysed in terms of various parameters. An overview of the proposed system is shown in Fig. 1.

3.2. Optimized unsupervised feature selection

A feature's information gain is generally inferred as the information quantity that the feature affords to the classification technique. Thus feature A's information gain is relative to S (Set of patterns) and is given as,

$$InfoGain((K|P) \equiv A(K) - \sum_{u \in Values(P)} A_u A(K_u) \tag{1}$$

Here values (P) denotes the set of entire possible values and feature P takes these values, A_u represents the K pattern's probability that belongs to u value in accordance with feature P, K_u is a subset of pattern K in which feature P consists of value u and $A(K)$ performs the entropy calculation of the K (set of patterns).

The variable Y's entropy is described in the below equation (2).

$$K(Y) = \sum_{u \in Values(Y)} -log_2(A_u) \tag{2}$$

Information gain inclines to support features with huge possible values. On the other hand, these features lack the power of prediction upon hidden patterns. Hence symmetrical uncertainity and gain ratio is presented to solve the issue.

A feature's gain ratio represents uniform and broad pattern division by the features and is represented in the below equation.

$$GR(K,P) \equiv \frac{InfoGain(K/P)}{A(P)} \tag{3}$$

Here $InfoGain(K/P)$ explores the feature P's information gain and $A(P)$ computes the feature P's entropy.

Symmetrical uncertainity inhibits the feature selection with many values, it also limits the values in a specific range of (0,1). It is given in below equation (4).

$$SyU(K,P) \equiv 2\left[\frac{InfoGain(K/P)}{A(K)+A(P)}\right] \tag{4}$$

Here $A(K)$ is the pattern K's entropy. $SyU = 1$ represents the set of pattern K and specified feature P are entirely correlated and $SyU = 0$ represents that (K) and P are independent.

Gini Index (GI) is a technique of an impurity split and it is appropriate for consistent numeric values. The GI of the set of pattern K is given in equation (5).

$$GI(K,P) \equiv Gini(K) - \sum_{u\in Values(P)} A_u Gini(K_u) \tag{5}$$

where $Gini(K) = 1 - \sum_{u\in Values(K)} (A_u)^2$

Fisher Score (FS) chooses features that involves the patterns corresponding to similar class are close to one another and patterns belonging to dissimilar classes are far from one another. The FS of the P feature is computed by the following equation.

$$FS(K,P) = \frac{\sum_{u\in Values(K)} n_u(\overline{P}_u - \overline{P})^2}{\sum_{u\in Values(K)} n_u(\sigma_u(P))^2} \tag{6}$$

Here $\overline{P}$ represents the mean of entire patterns that are related to feature P, n_u denotes the number of patterns that consists of class label u and $\overline{P}_u$ and $\sigma_u(P)$ are the mean and variance of feature P on class u.

Team Variance is the basic univariate assessment of features. It represents the features with huge variance values that consists of useful information. It is given in the following equation.

$$T(K,P) = \frac{1}{|K|}\sum_{i=1}^{|S|}(P(i) - \overline{P})^2 \tag{7}$$

Here $P(i)$ represents the feature P's value for the pattern i and $|S|$ indicates the total count of patterns.

3.3. *Classification approach – MLP-EBMDA*

Based on the selected features, classification was done using the MLP-EBMDA to predict the heart disease in humans.

MLP are capable of generalisation, that is, they classify an unknown pattern with other known patterns that share the same distinguishing features. This means noisy or incomplete inputs will be classified because of their similarity with pure and complete inputs. Multi-Layer Perceptron (MLP) contains one or more hidden layers (apart from one input and one output layer) so it can also learn non – linear functions.

Dragon Fly algorithm is a technique, which is a Metaheuristic algorithm based on swarm intelligence, is inspired by the static and dynamic behaviours of dragonflies in nature. There are two main stages of optimization: exploration and exploitation. These two phases were modelled by dragonflies, either dynamically or statically searching for food or avoiding the enemy

Brownian motion method is that its isotropic approach (completely independent of direction) increases the discovery capability. On the contrary, the sizes of the steps, both controllable size, and time-based random form, prevent the outgoing of the search space, providing continuity of motion. That why we are using in the Dragon Fly algorithm.

Thus, we researched and concluded our algorithm can works efficiently well and we have compared to other algorithm by providing better results.

3.3.1. *Dragonfly algorithm-DA*

The usual operation of dragonflies follows the standard of segregation from the present neighbour, arrangement towards good food, deviating from the enemy, coherence to the neighbour and desirability towards the food in search of optimal result. The task of exploitation and exploration to attain the global optima in dragonfly algorithm (DA) has been via the composition of dragonfly's dynamic and static behaviour.

Mathematical model for DA is explained below:

The dragonfly's population matrix with size N is given by

$$Y_j^E = (y_1, y_2, \cdots y_N \tag{8}$$

Here $j = 1, 2, 3\ldots.\ N$, Y_j^E denotes the position of j $^{\text{th}}$ dragonfly in the E search space and N is the count of search agent.

Diverse components that comprises of segregation, alignment and cohesion can be computed for $\overline{N}$ neighbour individual. The segregation parameter is computed by Eq. (9).

$$K_j = -\sum_{i=1}^{\overline{N}} Y - Y_i \tag{9}$$

The individual cohesion in the swarms is represented by the following equation.

$$B_j = \frac{\sum_{i=1}^{\overline{N}} Y_i}{\overline{N}} - Y \tag{10}$$

The alignment is represented in swarms as follows:

$$X_j = \frac{\sum_{i=1}^{\overline{N}} Q_i}{\overline{N}} \tag{11}$$

Here the terms j and i are utilized for the present individual as well as neighbour individual. Q_i Indicates the velocity of the neighbouring individual of i.

The desirability towards the source of food H_j and deviation from enemies is computed from Eqs. (12) and (13).

$$H_j = Y^+ - Y \tag{12}$$

$$I_j = Y^- + Y \tag{13}$$

Here Y^+ and Y^- represents the position of food and enemy. Y Representsthe dragonfly's current position.

For population with a one minimum neighbour dragonfly, the equation of dragonfly's position with respect to step vector is given by equation (14).

$$\Delta Y_{a+1} = \left(kK_j + xX_j + bB_j + hH_j + iI_j\right) + w\Delta Y_a \tag{14}$$

Here x, k, h, i and b are alignment, segregation, food desirability factor, enemy deviation factor and cohesion factor. w Denotes the inertia's weight and a denotes the iteration number. The dragonfly's position is given in equation (15).

$$Y_{a+1} = Y_a + \Delta Y_{a+1} \tag{15}$$

The dragonfly's position is updated by utilizing levy flight function that is given by equation (16).

$$Y_{a+1} = Y_a + lev(c)Y_a \tag{16}$$

Hence the dragonfly's fitness is assessed on the basis of position vector in a simplified form and sustained till a particular condition is met.

3.3.2. *MLP (Multi-Layer Perceptron)*

MLP is a kind of FNNs (Feed-forward Neural Networks) that passes the information in a single direction via Neural Networks (NNs) and its corresponding neurons are organized in several layers in parallel. The first layer in several parallel layers is the input layer, the last layer is called the output layer and the intermediate layers are called as hidden layers. When FNNs consists of a single hidden layer then it is called as MLP.

Based on the weights, inputs and biases, the MLP's outputs are computed in the below steps:

Initially the weighted total (input) is calculated by the following equation:

$$a_i = \sum_{j=1}^{n} \left(Y_{ji}, Z_j\right) - Y_{a+1}, \quad \mathrm{i} = 1, 2\cdots\mathrm{c} \tag{17}$$

Here n denotes the number of input nodes, Y_{ji} represents the association weight's accuracy from the j $^{\mathrm{th}}$ node thereby corresponding to the input layer to the i $^{\mathrm{th}}$ node corresponding to the hidden layer, Z_j representsthe input j, c denotes the count of hidden nodes and Y_{a+1} is the threshold value of the hidden node i.

The individual hidden node's output is calculated by the following equation:

$$P_i = sigM(a_i) = 1/(1 + exp(- a_i)), \quad \mathrm{i} = 1, 2\cdots\mathrm{c} \tag{18}$$

The overall outputs are categorised on the basis of hidden node's evaluated outputs.

$$r_s = \sum_{i=1}^{s} (Y_{is}, P_i) - D'_s, \quad \mathrm{s} = 1, 2\cdots\mathrm{n} \tag{19}$$

$$H_s = sigM(r_s) = 1/(1 + exp(- r_s)) \quad \mathrm{k} = 1, 2\cdots\mathrm{n} \tag{20}$$

Thus the output of MLP has been found via the biases and weights. The DA is used for the parameters of MLP as a trainer.

4. Results and discussion

The results of the proposed system to predict the heart disease as normal and abnormal is discussed in this section. Moreover the comparative analysis is also performed.

4.1. Dataset description

The dataset of the heart disease [24] is taken into account for this study. The dataset has been partitioned into two sections. The first part is utilized for training and building the model. The second part is utilized for testing to determine the model's accuracy. UCI Cleveland Dataset. Dataset consist of 305 Data and 14 features. It's a labelled dataset.

4.2. Performance metrics

The proposed system is evaluated in terms of various performance metrics such as accuracy, precision, recall and F1-score.

(i) *Accuracy*

Accuracy is the prediction fraction that has been obtained for the model correctly and it can be written by the following equation.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{21}$$

Here *TP* represents True Positive which are predicted correctly and are abnormal, *TN* represents the True Negative that are predicted correctly and normal. *FP* Represents the False Positive that are predicted incorrectly but are normal and *FN* represents the False Negative that are predicted incorrectly and are abnormal.

(ii) *Precision*

It is defined as the tuple's proportion which predict the abnormal patient have heart disease exactly that is calculated by the following equation.

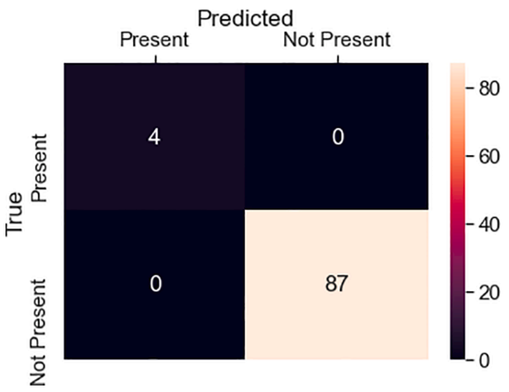


Fig. 2. Confusion matrix.

Table 2
Comparative analysis of the proposed and traditional studies [25].

Classifiers	Sensitivity	Specificity	Accuracy	F-score	Kappa
OANN(DBMRITLBO-ANN)	97.33	93.23	95.41	95.74	90.76
TLBO-ANN	97.22	84.91	90.75	90.91	81.59
J48	78.94	78.03	78.55	80.59	56.64
Random tree	74.41	71.11	72.94	75.3	45.38
RBF Network	85.19	81.56	83.49	84.66	66.81
NB Tree	80.24	77.94	79.21	80.97	58.06
Random forest	81.98	82.44	82.18	83.93	63.95
Proposed MLP-EBMDA	**98.92**	**96.47**	**97.63**	**96.45**	**96.75**

$$Precision = \frac{TP}{TP + FP} \tag{22}$$

(iii) *Recall*

It is the ratio of tuples that correctly predict the abnormal patients by exploring that the particular person have heart disease. It is given by:

$$Recall = \frac{TP}{TP + FN} \tag{23}$$

(iv) *F1-Score*

The harmonic mean of recall and precision is defined as F1-score. It is computed by the following equation.

$$F1 - Score = \frac{2{*}precision{*}recall}{precision + recall} \tag{24}$$

4.3. Experimental results

The performance of the heart disease prediction system is exhibited by comparing with the various existing works.

The above Fig. 2 shows 87 as true negative value and thus the study is correctly classified abnormality. And true positive also showing that 4 classes are correctly classified as normal. Further, the false positive and false negative is zero and thus the no incorrect classifications are present.

4.3.1. Comparative analysis

The proposed system is compared with three conventional studies for assessing the efficiency of the introduced system. Accordingly, the study [25] has been considered for analysis. Here, existing studies like Optimal

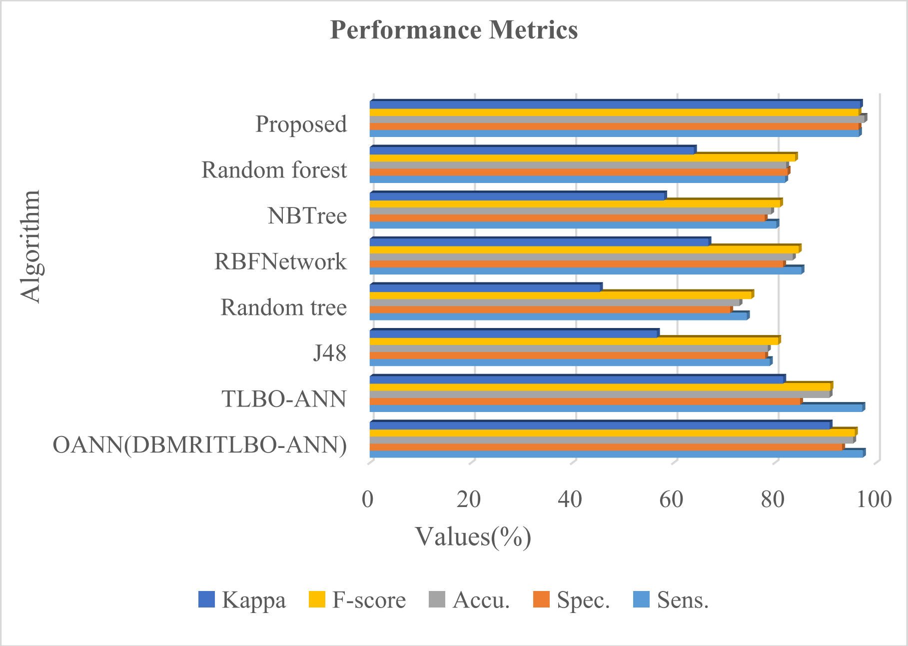


Fig. 3. Analysis of the traditional [25] and proposed system.

Table 3
Analysis of the traditional [26] and proposed system.

Algorithm	Accuracy (%)	Precision (%)	Sensitivity (%)	F1 Score (%)
KNN	60	61	59	58
LR	78	79	78	78
LDA	78	80	79	79
SVM	79	80	79	79
CART	68	69	68	68
GB	81	79	84	81
RF	83	81	87	84
Existing method	93	96	91	93
Proposed MLP-EBMDA	**97**	**95**	**98.92**	**96**

Artificial Neural Network (Distance Based Misclassified Instance Removal-Teaching and Learning Based Optimization-ANN), Teaching and Learning Based Optimization-ANN (TLBO-ANN), Random Tree, Naïve Bayes Tree (NB Tree), J48, Radial Basis Function Network (RBF Network) and Random Forest is considered for comparison.

From Table 2, the results obtained after comparative analysis explore that the traditional OANN (DBMRITLBO-ANN) showed sensitivity of 97.33%, TLBO-ANN showed 97.22% sensitivity, J48 showed 78.94% sensitivity and Random Tree showed 74.41% sensitivity. Similarly, RBF Network showed 85.19%, NB Tree showed 80.24% and RF showed 81.98% as sensitivity rate and the proposed system showed higher sensitivity rate as 98.92%. Nevertheless, accuracy is the significant metric to confirm the efficiency of a method. In that sense, the proposed system shows high accuracy rate of 97.63% than the traditional studies. This proves the efficiency of the proposed system and is graphically shown in Fig. 3.

Subsequently, analysis has been undertaken by using the study [26]

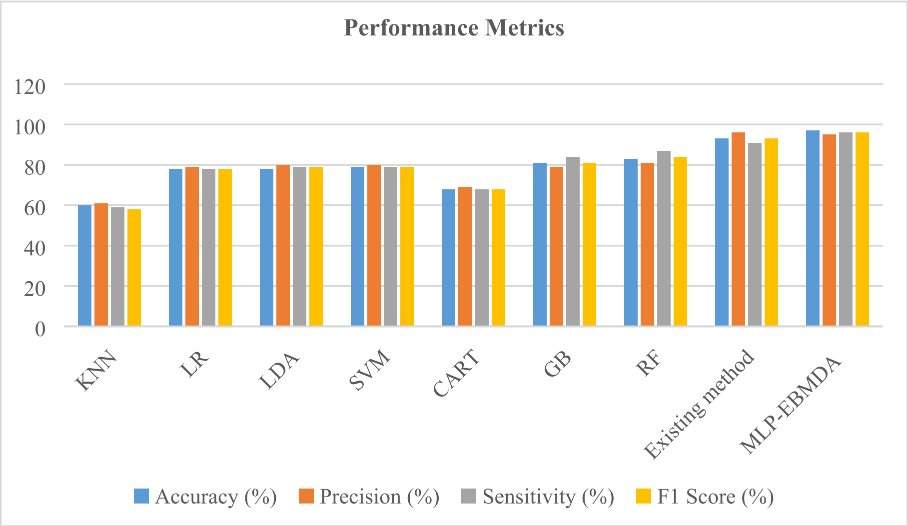


Fig. 4. Comparative analysis of the existing [26] and the proposed MLP-EBMDA.

Table 4
Analysis of the proposed and traditional system [26].

Method	Accuracy (%)	Precision (%)	Sensitivity (%)	F1 Score
Majority vote with NB, BN, RF, and MLP	85.48	NA	NA	NA
L1 Linear SVM + L2 Linear & RBF SVM	92.22	NA	82.92	NA
HRFLM	88.4	90.1	92.8	90
NB and AES	89.77	NA	NA	NA
ANN and Fuzzy_AHP	91	NA	NA	NA
Randomized decision tree ensemble	93	96	91	93
Proposed MLP-EBMDA	**97**	**95**	**98.92**	**96**

*NA-Not Available.

Table 5
Comparative analysis in terms of accuracy, sensitivity and specificity [27]

Method	Specificity	Sensitivity	Accuracy
KNN + SOM + PCA + Fuzzy SVM	0.9435	0.9606	0.9686
Hot-Deck + SOM + PCA + Fuzzy SVM	0.9334	0.954	0.9568
SOM + PCA + SVM	0.9143	0.9412	0.9449
PCA + SVM	0.8571	0.902	0.8976
SOM + SVM	0.8131	0.8693	0.8661
SVM	0.7642	0.8377	0.8268
NB	0.6909	0.7778	0.7677
DT	0.6727	0.7616	0.7441
NN	0.633	0.7351	0.7087
Proposed MLP-EBMDA	**0.9647**	**98.92**	**0.9763**

for comparison. Accordingly, K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Linear Discriminant Analysis (LDA), Classification and Regression Tree (CART) and Gradient Boosting (GB) has been the existing studies used to compare with the proposed system. The analysis has been undertaken with respect to accuracy, sensitivity, F1-score and precision and the obtained results are shown in Table 3.

The results showed that the accuracy of existing KNN has been found to be 60%, LR and LDA showed 78%, RF showed 83%, SVM showed 79%. However, the proposed system showed high accuracy of 97% than other methods. In addition, the proposed system shows high sensitivity, F1-score and precision rate than the existing studies. Though, the existing studies used various methods for classifying the heart disease dataset, the proposed system shows high accuracy rate confirming its efficiency in predicting heart disease. It is graphically shown in Fig. 4.

In addition, analysis has been undertaken by considering the existing study [26] for analysis. Here, various traditional methods have been considered such as Majority vote with Naïve Bayes (NB), Bayes Net (BN), Random Forest (RF) and Multi-Layer Perceptron (MLP), Hybrid Random Forest with Linear Model (HRFLM), L1 Linear SVM + L2 Linear and RBF SVM, NB and Advanced Encryption Standard (AES), ANN and Fuzzy Analytical Hierarchy Process (AHP) and Randomized decision tree ensemble. The results obtained by the comparative analysis are tabulated in Table 4.

From the results, it is clearly seen that most of the methods have not been evaluated with respect to precision, sensitivity and F1-score. Accuracy has been taken into account by all the methods. On the other hand, the proposed system has considered all the metrics for analysis. Outcomes reveal that the accuracy of proposed MLP-EBMDA is higher than traditional studies. Comparison has also been carried out with various other traditional methods [27]. In this analysis, several existing methods have been taken for comparison namely KNN + Self-Organizing Map (SOM) + Principal Component Analysis (PCA) + Fuzzy SVM, PCA + SVM, SVM, NB, Hot-Deck + SOM + PCA + Fuzzy SVM, SOM + PCA + SVM, SOM + SVM, DT and NN. It is shown in Table 5.

From the analytical results, it is found that the existing KNN + SOM + PCA + Fuzzy SVM showed specificity rate 0.9435%, sensitivity rate as 0.9606% and accuracy as 0.9686%. similarly, the other considered existing methods like Fuzzy SVM, PCA + SVM, SVM, NB, Hot-Deck + SOM + PCA + Fuzzy SVM, SOM + PCA + SVM, SOM + SVM, DT and NN showed better outcomes in terms of the three metrics. However, the proposed system shows outstanding performance than existing methods with 0.9647% as specificity rate, 0.9892% as sensitivity rate and 0.9686% as accuracy rate. It is graphically shown in Fig. 5.

From all the three comparative analysis, it is found that the proposed system is better than existing studies for heart disease. Typically, MLPs

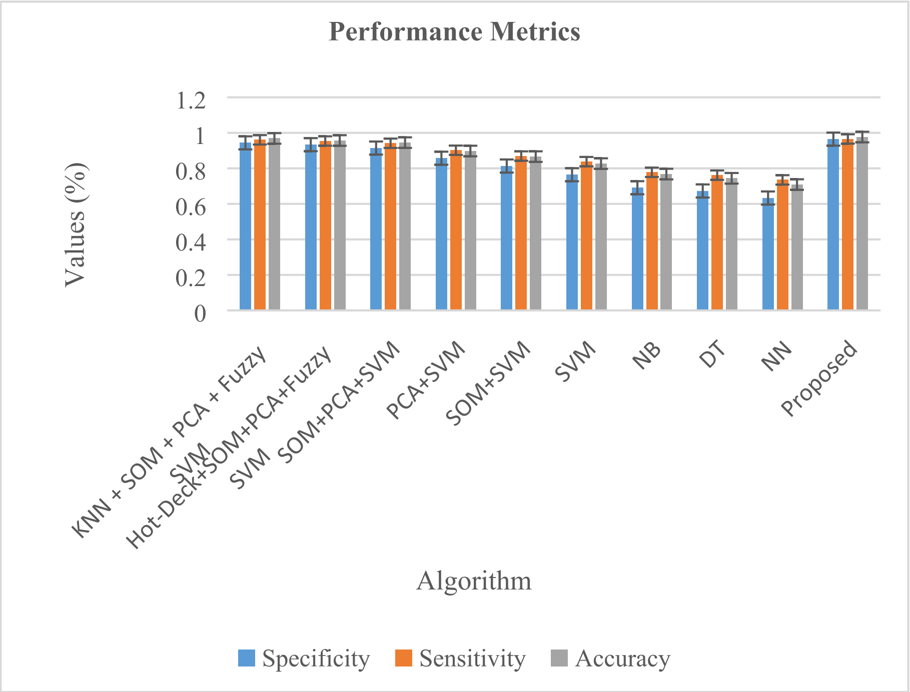


Fig. 5. Analysis of the existing [27] and proposed methods.

find value in research due to their ability for solving issues stochastically that often permits suitable solutions to extreme issues like fitness calculation. It is also capable of adaptive learning. On the other hand, Dragon fly algorithm is capable of optimizing several diverse issues in various areas. This is also easy for implementation. It possess few parameters to tune which is an added advantage. Moreover, the algorithm's convergence time is rational. This algorithm is robust than other algorithms and can be easily hybrid with different algorithms. Considering all these merits, this study proposed Enhanced Brownian motion depended on dragonfly algorithm (MLP-EBMDA). Hence, these advantages have made the proposed system to perform better than traditional studies which proves the suitability of the introduced system in heart disease prediction.

5. Conclusion

In this study, a Multi-Layer Perceptron integrated with enhanced Brownian motion on the basis of Dragonfly Algorithm (MLP-EBMDA) has been used for classification and optimized unsupervised method for feature selection has been proposed for cardiac disease prediction. The suggested system has been analyzed in terms of certain specific parameters such as precision, recall, F1-Score, kappa and accuracy. The proposed system has been compared with the existing system in terms of the mentioned parameters and the taken testing data to determine the efficiency in heart disease prediction. The analytical results explored that the proposed system has shown effective results than the traditional methods in terms of accuracy for predicting the heart disease. The proposed system revealed prediction accuracy at the rate of 94.28% and sensitivity as 98.92%, thus resulting in better prediction of heart disease as normal or abnormal.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] N.L. Fitriyani, M. Syafrudin, G. Alfian, J. Rhee, HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System, IEEE Access 8 (2020) 133034–133050, https://doi.org/10.1109/ACCESS.2020.3010511.

[2] S. Mohan, C. Thirumalai, G. Srivastava, Effective heart disease prediction using hybrid machine learning techniques, IEEE Access 7 (2019) 81542–81554, https://doi.org/10.1109/ACCESS.2019.2923707.

[3] R. Singh, A Review on Heart Disease Prediction using Unsupervised and Supervised Learning, Neural Netw., vol. 99, p. 100.

[4] Z. Al-Makhadmeh, A. Tolba, Utilizing IoT wearable medical device for heart disease prediction using higher order Boltzmann model: A classification approach, Measurement 147 (2019) 106815, https://doi.org/10.1016/j.measurement.2019.07.043.

[5] M. Mirmozaffari, A. Alinezhad, A. Gilanpour, Heart disease prediction with data mining clustering algorithms, Int. J. Comput. Commun. Instrum. Eng. 4 (2017) 16-19. 10.15242/ijccie.dir1116009.

[6] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, J.A. Khan, An Automated Diagnostic System for Heart Disease Prediction Based on χ^2 Statistical Model and Optimally Configured Deep Neural Network, IEEE Access 7 (2019) 34938–34945, https://doi.org/10.1109/ACCESS.2019.2904800.

[7] A.M. Sagir, S. Sathasivam, A Novel Adaptive Neuro Fuzzy Inference System Based Classification Model for Heart Disease Prediction, Pertanika J. Sci. Technol. 25 (2017).

[8] F. Ali, S. El-Sappagh, S.M.R. Islam, D. Kwak, A. Ali, M. Imran, K.-S. Kwak, A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion, Inf. Fusion 63 (2020) 208–222, https://doi.org/10.1016/j.inffus.2020.06.008.

[9] C.B. Gokulnath, S.P. Shantharajah, An optimized feature selection based on genetic approach and support vector machine for heart disease, Cluster Comput. 22 (S6) (2019) 14777–14787, https://doi.org/10.1007/s10586-018-2416-4.

[10] G.T. Reddy, M.P.K. Reddy, K. Lakshmanna, D.S. Rajput, R. Kaluri, G. Srivastava, Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis, Evolut. Intell. 13 (2) (2020) 185–196.

[11] B. Narasimhan, A. Malathi, Altered particle swarm optimization based attribute selection strategy with improved fuzzy Artificial Neural Network classifier for coronary artery heart disease risk prediction, Int. J. Adv. Res. Ideas Innov. Technol. 5 (2019) 1196–1203.

[12] N.S.R. Pillai, K.K. Bee, J. Kiruthika, Prediction of heart disease using RNN algorithm, 2019.

[13] S.B. Babu, A. Suneetha, G.C. Babu, Y.J.N. Kumar, G. Karuna, Medical disease prediction using grey wolf optimization and auto encoder based recurrent neural network, Period. Eng. Natural Sci. 6 (2018) 229–240, https://doi.org/10.21533/pen.v6i1.286.

[14] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, A.A. Yarifard, Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm, Comput. Methods Programs Biomed. 141 (2017) 19–26.

[15] V. Kirubha, S.M. Priya, Survey on data mining algorithms in disease prediction, Int. J. Comput. Trends Tech. 38 (2016) 24–128, https://doi.org/10.14445/22312803/ijctt-v38p122.

[16] H. Sharma, M. Rizvi, Prediction of heart disease using machine learning algorithms: A survey, Int. J. Recent Innov. Trends Comput. Commun. 5 (2017) 99–104.

[17] E. Choi, A. Schuetz, W.F. Stewart, J. Sun, Using recurrent neural network models for early detection of heart failure onset, J. Am. Med. Inform. Assoc. 24 (2017) 361–370, https://doi.org/10.1093/jamia/ocw112.

[18] B. Jin, C. Che, Z. Liu, S. Zhang, X. Yin, X. Wei, Predicting the risk of heart failure with EHR sequential data modeling, IEEE Access 6 (2018) 9256–9261, https://doi.org/10.1109/access.2017.2789324.

[19] T. Salem, Study and analysis of prediction model for heart disease: an optimization approach using genetic algorithm, Int. J. Pure Appl. Math. 119 (2018) 5323–5336.

[20] N. Masih, H. Naz, S. Ahuja, Multilayer perceptron based deep neural network for early detection of coronary heart disease, Health Technol. 11 (1) (2021) 127–138.

[21] S. Koppu, P.K.R. Maddikunta, G. Srivastava, Deep learning disease prediction model for use with intelligent robots, Comput. Electr. Eng. 87 (2020), 106765.

[22] M. Alshinwan, L. Abualigah, M. Shehab, M. Abd Elaziz, A.M. Khasawneh, H. Alabool, et al., Dragonfly algorithm: a comprehensive survey of its results, variants, and applications, Multimedia Tools Appl. (2021) 1–38.

[23] A.U. Haq, J.P. Li, M.H. Memon, S. Nazir, R. Sun, A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms, Mobile Inf. Syst. 2018 (2018) 1–21, https://doi.org/10.1155/2018/3860146.

[24] https://www.kaggle.com/ronitf/heart-disease-uci.

[25] R. Thanga Selvi, I. Muthulakshmi, An optimal artificial neural network based big data application for heart disease diagnosis and classification model. J Ambient Intell Hum Comput, Springer, Berlin, 2020.

[26] I.D. Mienye, Y. Sun, Z. Wang, An improved ensemble learning approach for the prediction of heart disease risk, Inf. Med. Unlocked 20 (2020) 100402, https://doi.org/10.1016/j.imu.2020.100402.

[27] M. Nilashi, H. Ahmadi, A.A. Manaf, T.A. Rashid, S. Samad, L. Shahmoradi, N. Aljojo, E. Akbari, Coronary heart disease diagnosis through self-organizing map and fuzzy support vector machine with incremental updates, Int. J. Fuzzy Syst. 22 (4) (2020) 1376–1388, https://doi.org/10.1007/s40815-020-00828-7.

D. Deepika received her B.E Degree in Computer Science and Engineering from Anna University, Chennai, India in 2011 and received her M.E from Anna University, Chennai, India in 2013. From 2014 through 2020, she served as an Assistant Professor with K.L.N College of Information Technology, Sivagangai, TamilNadu. She is currently pursuing her Ph.D. in Anna University, Chennai, India. Her research interests include Artificial intelligence, Machine Learning, Deep Learning.

Dr. N. Balaji currently serves as a Professor in the Department of Information Technology, Karpagam College of Engineering. He did his Doctorate Degree in Computer Science and Engineering from Anna University, Chennai. He served as Faculty with Thiagarajar College of Engineering, Madurai and also acted as Head for Information Technology Department. He served for K.L.N.Group of Institutions for more than 15 years including Principal for K.L.N.College of Information Technology. He have 25 years of Teaching Experience and also have 26 Journals/Conference Publications and has authored one Book. Eight candidates had completed their PhD under his able guidance.

ARTICLES FOR FACULTY MEMBERS

AN IMPROVED MLP-RFE MODEL FOR ENHANCED ACCURACY IN HEART DISEASE PREDICTION USING DEEP LEARNING

Heart-related Clinical Biomarker Classification through Machine Learning Algorithms / Safdar, S., Rauf, A., Tariq, A. B., Qazi, S., & Bakhtiar, S. M.

Current Trends in OMICS
Volume 5 Issue 1 (2025) Pages 40 - 66
https://doi.org/10.32350/cto.51.03
(Database: Current Trends in OMICS ,CTO)

Current Trends in OMICS (CTO)
Volume 5 Issue 1, Spring 2025
ISSN$_{(P)}$: 2790-8283, ISSN$_{(E)}$: 2790-8291
Homepage: https://journals.umt.edu.pk/index.php/cto

Article QR

Title:

Heart-related Clinical Biomarker Classification through Machine Learning Algorithms

Author (s): Sahar Safdar[1], Abdur Rauf[2], Areeba Bint Tariq[1], Shawana Qazi[1], and Syeda Marriam Bakhtiar[1]

Affiliation (s): [1]Capital University of Science and Technology, Islamabad, Pakistan
[2]National University of Science & Technology, Islamabad, Pakistan

DOI: https://doi.org/.10.32350/cto.51.03

History: Received: February 05, 2025, Revised: March 20, 2025, Accepted: April 27, 2025, Published: June 10, 2025

Citation: Safdar S, Rauf A, Tariq AB, et al. Heart-related clinical biomarker classification through machine learning algorithms. *Curr Trend OMICS*. 2025;5(1):40–66. https://doi.org/.10.32350/cto.51.03

Conflict of Interest: Author(s) declared no conflict of interest

A publication of
The Department of Life Sciences, School of Science
University of Management and Technology, Lahore, Pakistan

Heart-related Clinical Biomarker Classification through Machine Learning Algorithms

Sahar Safdar[1], Abdur Rauf[2], Areeba Bint Tariq[1], Shawana Qazi[1], and Syeda Marriam Bakhtiar[1]*

[1]Department of Bioinformatics & Biosciences, Genetic & Molecular Epidemiology Research Group, Capital University of Science and Technology, Islamabad, Pakistan
[2]National University of Science & Technology, Islamabad, Pakistan

ABSTRACT

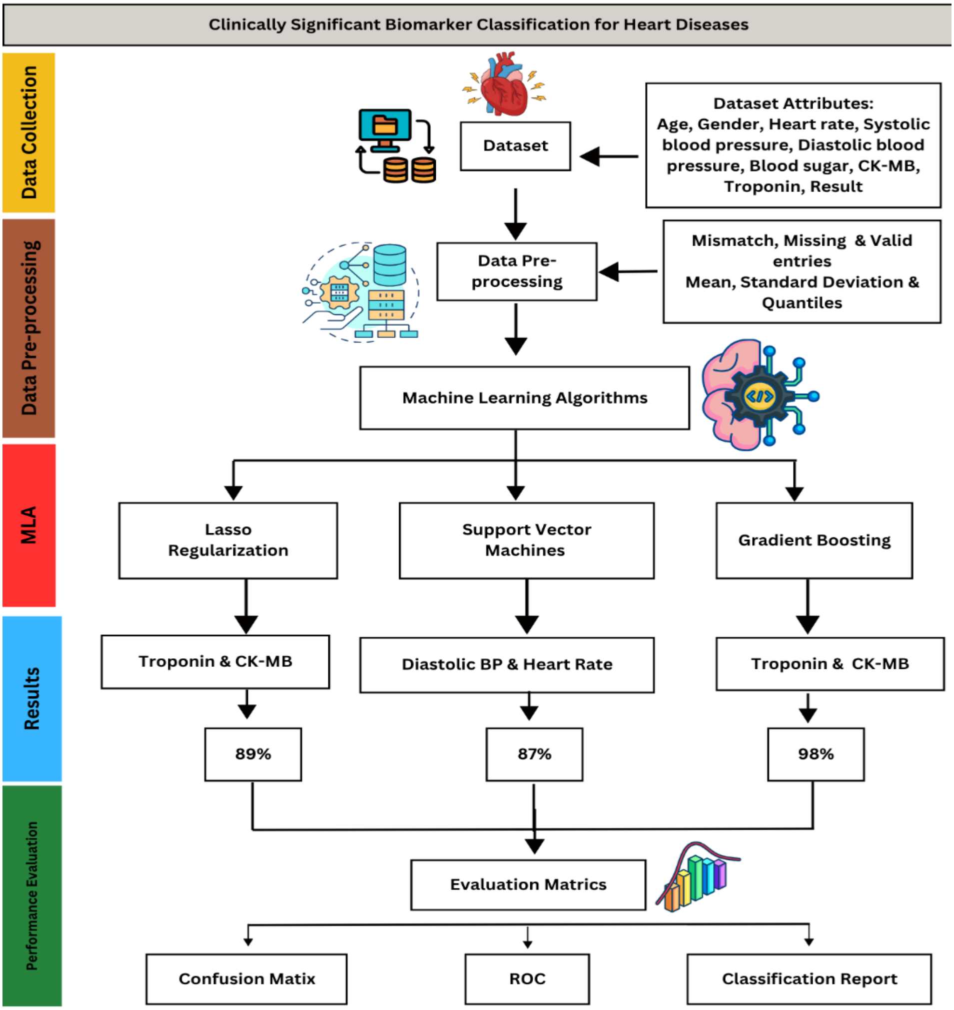
Heart diseases continue to be a major global cause of morbidity and death. Their timely and accurate diagnosis for improved patient outcomes is direly needed. Clinical biomarkers for the timely diagnosis of heart diseases are known but underutilized due to the use of conventional analytical methods that lower the efficiency to handle large datasets. Furthermore, conventional methods also fail to incorporate demographic biomarkers such as age and hemodynamic biomarkers such as heart rate and diastolic blood pressure. This significantly influences heart diseases. Using cutting edge machine learning (ML) techniques including Lasso regularization, support vector machine (SVM), and gradient boosting (GB), this study investigated the importance of clinical biomarkers for heart disease prediction. Troponin and Creatinine Kinase - MB (CK-MB) were found to be the most significant predictors among the examined characteristics in every model, underscoring their crucial importance in the diagnosis of myocardial ischemia and damage. Diastolic blood pressure was also found to be an adequate predictor, highlighting its role in increasing cardiovascular risk because of autonomous dysfunction. While SVM and GB performed strongly in managing intricate data relationships, Lasso regularization successfully decreased feature redundancy. The results support the use of clinically applicable biomarkers in conjunction with machine learning to improve the accuracy of diagnosis and also opens the door to the personalized treatment of heart disease. Validating these findings in a variety of populations and adding more biomarkers for a thorough risk assessment should be the main goals of future studies.

---

*Corresponding Author: marriam@cust.edu.pk

Keyword: CK-MB, gradient boosting (GB) classifier, heart diseases, Lasso regularization, machine learning (ML) algorithms, support vector machine (SVM), troponin

GRAPHICAL ABSTRACT



1. INTRODUCTION

For the past few decades, heart diseases (HDs), often referred to as cardiovascular diseases (CVDs), have been the leading cause of mortality all around the globe. These diseases encompass a variety of conditions that affect the heart. Every year, over 17.9 million deaths are reported due to

CVDs, as estimated by WHO (World Health Organization) [1]. The European Society of Cardiology estimates that 3.6 million people are diagnosed with heart diseases each year, out of an overall population of 26 million. Common symptoms of heart diseases include swollen feet, muscular weakness, and shortness of breath [2].

Clinical biomarkers are essential for diagnosing and managing heart diseases. They provide valuable insights into the physiological and pathological processes associated with cardiovascular risks. These biomarkers include troponin, CK-MB, and blood pressure, all of which are widely used in clinical practices due to their high specificity and sensitivity [3]. The integration of these biomarkers in diagnostics significantly enhances clinical decision-making by enabling early detection, reducing uncertainty, and increasing precision. The important biomarkers associated with heart diseases are shown below in Table 1.

Table 1. Clinical Biomarkers and their Significance in Heart Diseases

Biomarkers	Significance
Age	Aging causes arterial stiffening, endothelial dysfunction, and increased prevalence of some conditions, such as hypertension [4].
Gender	Men have a higher risk of developing heart diseases at a younger age than women, which could be elevated due to hormonal changes [5].
Systolic Blood Pressure	Its elevated level predicts a strong cardiovascular event [6].
Diastolic Blood Pressure	It indicates elevated peripheral resistance, especially in younger individuals [6].
CK-MB	It indicates myocardial necrosis and also detects reinfarction [7].
Troponin	It specifies myocardial injury and also assesses the severity of cardiac damage [8].

In the current global scenario, diagnosing CVDs at an early stage through their symptoms remains significantly difficult. If not detected timely, CVDs may cause mortality [9]. Without access to modern technologies and trained medical personnel, diagnosing and treating heart diseases is highly challenging. This is due to the fact that the assessment of the patient's medical background and physical examination, including detecting CVDs is based on critical symptoms [10]. However, the findings of this diagnostic procedure do not accurately identify the heart patient. Additionally, the analysis is costly and computationally challenging^2. Clinically useful methods to predict and detect CVDs would involve machine learning and data mining [11].

The expanding availability of healthcare records and the rapid advancement of analytical tools is revolutionizing the healthcare industry. Within artificial intelligence, machine learning (ML) is a developing field. Its main goal is to build systems, give them the ability to learn, and use that learning to generate predictions. To create models, it looks for hidden trends in the input dataset and produces highly accurate predictions for fresh datasets. After cleaning the dataset, any values that are missing are filled in. The new input is used to predict the risk associated with heart diseases, followed by an accuracy assessment [1].

Recent research confirms that ML algorithms can be used in clinical contexts. Internal validation requires the integration of biomarkers with demographic information in order to obtain high accuracy in predicting cardiovascular events via gradient boosting (GB) [12]. A recent study revealed that support vector machine (SVM) is easily adapted to different cardiac situations and is also useful in predicting heart failure. In the current study, CVDs are predicted using various ML methods, such as SVM, random forest (RF), and logistic regression (LR). These methods were chosen because they work effectively in solving non-linear classification problems. Based on the results, SVM performed better than both LR and RF, averaging an area under the ROC curve value of 78.84%, while giving 78.41% for LR and 77.50% for RF. The above results indicate that SVM is more suitable as a method to predict CVDs, as compared to other methods [13].

The suggested hybrid framework bridges the gap between clinical practice and ML by utilizing interpretability and feature selection. The model identifies the strongest predictive variables, while keeping things transparent for physicians by fusing ML with healthcare. This is crucial in the medical field, as doctors need precise and useful information to determine that AI-driven predictions meet the accepted diagnostic standards. By ensuring the model's accuracy and interpretability, this method faciliates real-time diagnostic decision-making and individualized treatment planning [14].

Although considerable progress has been made, almost all efforts focus on either biological validation or computational techniques, rarely bridging the gap between the two. Datasets often lack adequate demographic and heamodynamic variety to draw generalizable conclusions. Hence, this study employs the use of ML models to classify clinical biomarkers comprising demographic and heamodynamic factors for early and timely diagnosis of heart diseases. By bridging this gap in clinical diagnostics and incorporating ML, the study anticipates significant advancement in the cardiology precision medicine.

2. METHOD

ML is gaining popularity in the field of cardiovascular medicine. Finding the most practical ML algorithm for CVD datasets is still difficult, even with the abundance of available techniques [15].

The approach employed in this research aims to achieve this challenging objective by applying a wide range of ML algorithms to a well-selected dataset. For this purpose, the proposed methodology comprises data collection and preprocessing, application of ML algorithms, and performance evaluation, as shown in Fig. 1 below.

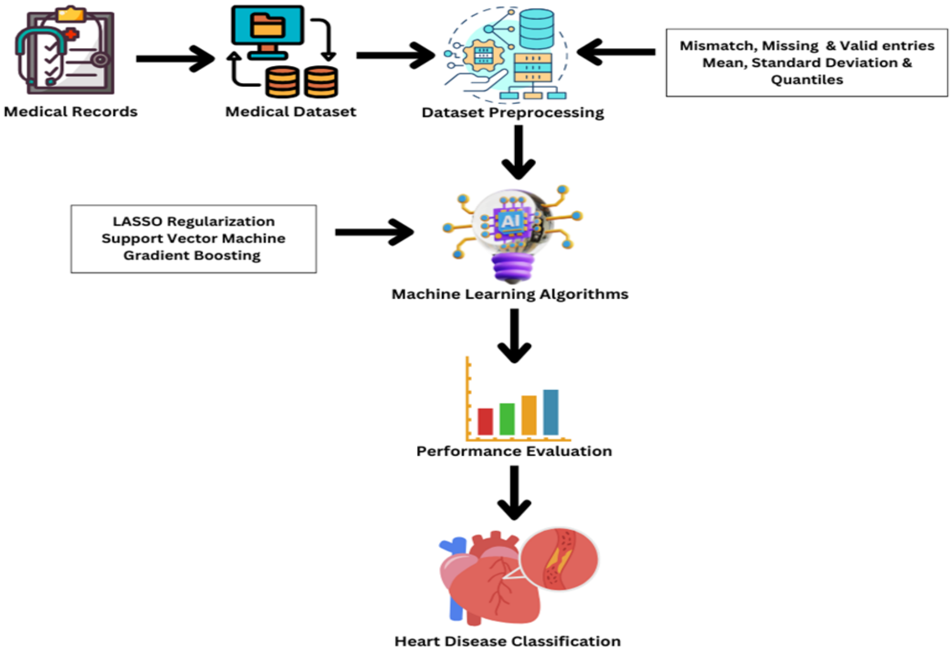


Figure 1. Heart Disease Classification Framework

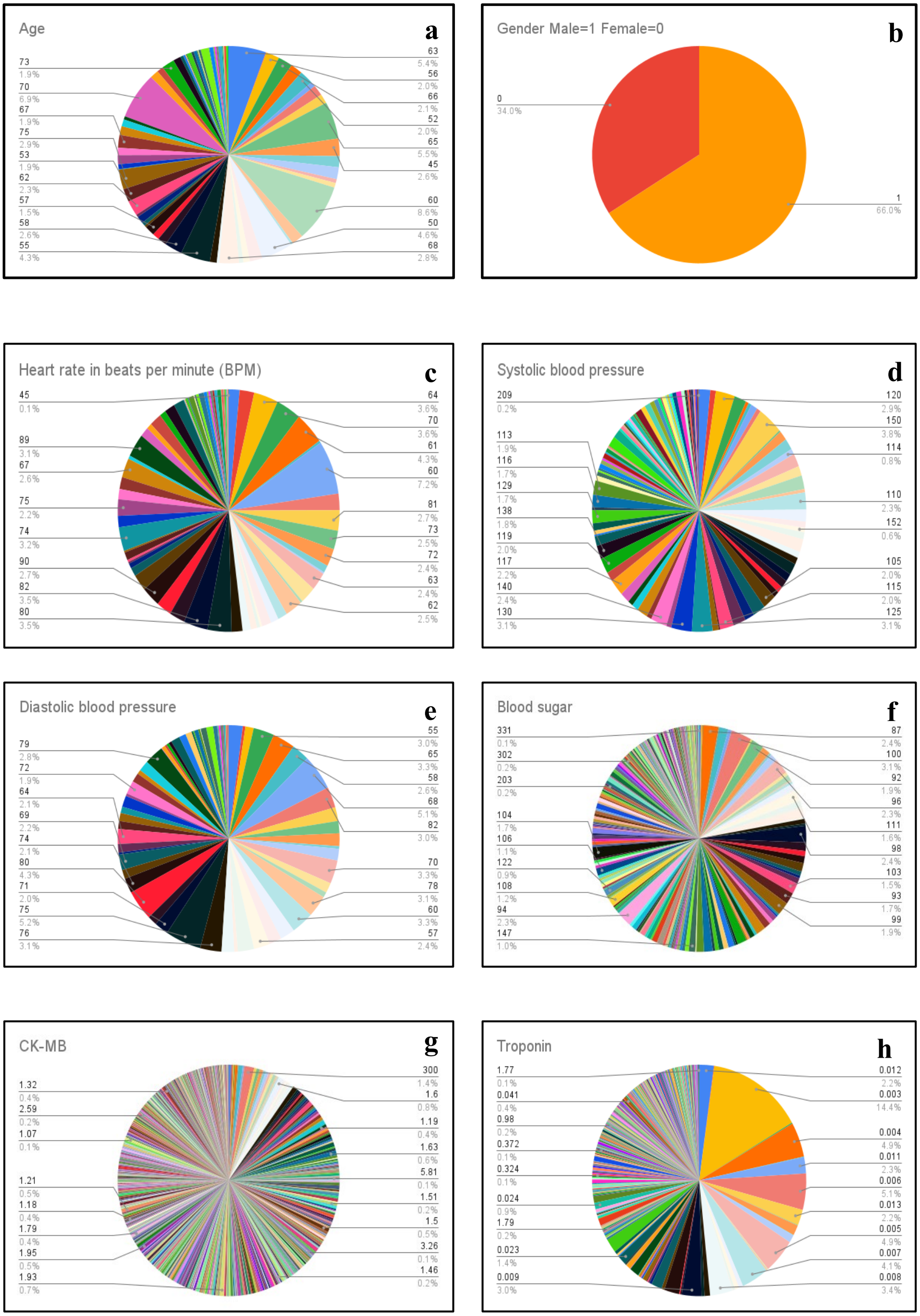
2.1. Dataset and Preprocessing

Dataset related to heart diseases was extracted from the online repository [16] and organized for the purpose of this research. The properties of a dataset are its features that are essential to analyze and predict the concerned problem. Many factors are considered to predict diseases, such as gender, chest discomfort, and blood pressure. Conversely, one method for dataset attribute selection is to use the matrix of correlation. The dataset used in the study comprised 1,319 entries with the following 9 attributes, as shown in Table 2.

Table 2. Description and Classification of Attributes Used in the Dataset for Heart Diseases [16]

Attribute	Description	Data Type
Age	Patient's age	Integer
Gender	Patient's gender (1 = Male, 0 = Female)	Integer
Heart Rate	Heart rate of the patient (beats per minute)	Integer
Systolic Blood Pressure	Patient's systolic blood pressure (mmHg)	Integer
Diastolic Blood Pressure	Patient's diastolic blood pressure (mmHg)	Integer
Blood Sugar	Patient's blood sugar level (mg/dL)	Float
CK-MB	Creatine Kinase-MB level	Float
Troponin	Troponin level	Float
Result	Diagnosis result (positive or negative)	Categorical (Object)

The graphical distribution of the individual attributes of the dataset reveals crucial trends, patterns, and variability. The representation provides a robust foundation for the application of ML models to predict heart diseases effectively. The pie charts shown in Fig. 2 allow an intuitive understanding of individual variables.


Age
a
73
1.9%
70
6.9%
67
1.9%
75
2.9%
53
1.9%
62
2.3%
57
1.5%
58
2.6%
55
4.3%
63
5.4%
56
2.0%
66
2.1%
52
2.0%
65
5.5%
45
2.6%
60
8.6%
50
4.6%
68
2.8%
Gender Male=1 Female=0
b
0
34.0%
1
66.0%
Heart rate in beats per minute (BPM)
c
45
0.1%
89
3.1%
67
2.6%
75
2.2%
74
3.2%
90
2.7%
82
3.5%
80
3.5%
64
3.6%
70
3.6%
61
4.3%
60
7.2%
81
2.7%
73
2.5%
72
2.4%
63
2.4%
62
2.5%
Systolic blood pressure
d
209
0.2%
113
1.9%
116
1.7%
129
1.7%
138
1.8%
119
2.0%
117
2.2%
140
2.4%
130
3.1%
120
2.9%
150
3.8%
114
0.8%
110
2.3%
152
0.6%
105
2.0%
115
2.0%
125
3.1%
Diastolic blood pressure
e
79
2.8%
72
1.9%
64
2.1%
69
2.2%
74
2.1%
80
4.3%
71
2.0%
75
5.2%
76
3.1%
55
3.0%
65
3.3%
58
2.6%
68
5.1%
82
3.0%
70
3.3%
78
3.1%
60
3.3%
57
2.4%
Blood sugar
f
331
0.1%
302
0.2%
203
0.2%
104
1.7%
106
1.1%
122
0.9%
108
1.2%
94
2.3%
147
1.0%
87
2.4%
100
3.1%
92
1.9%
96
2.3%
111
1.6%
98
2.4%
103
1.5%
93
1.7%
99
1.9%
CK-MB
g
1.32
0.4%
2.59
0.2%
1.07
0.1%
1.21
0.5%
1.18
0.4%
1.79
0.4%
1.95
0.5%
1.93
0.7%
300
1.4%
1.6
0.8%
1.19
0.4%
1.63
0.6%
5.81
0.1%
1.51
0.2%
1.5
0.5%
3.26
0.1%
1.46
0.2%
Troponin
h
1.77
0.1%
0.041
0.4%
0.98
0.2%
0.372
0.1%
0.324
0.1%
0.024
0.9%
1.79
0.2%
0.023
1.4%
0.009
3.0%
0.012
2.2%
0.003
14.4%
0.004
4.9%
0.011
2.3%
0.006
5.1%
0.013
2.2%
0.005
4.9%
0.007
4.1%
0.008
3.4%

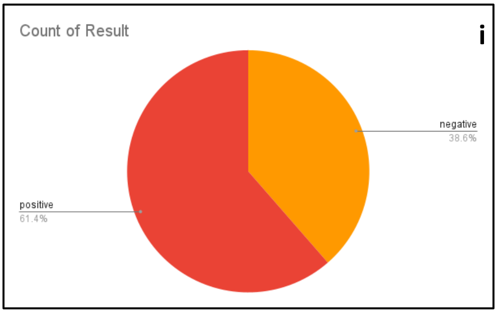


Figure 2. Distribution of Attributes in the Dataset (a) Age (b) Gender (c) Heart Rate (d) Systolic Blood Pressure (e) Diastolic Blood Pressure (f) Blood Sugar (g) CK-MB (h) Troponin (i) Results

Initially, the preprocessing of dataset involves removing noises or missing values to obtain accurate and authentic results. This was carried out by the addition of new variables with a strong predictive value and the removal of any features deemed unnecessary for analysis. The missing values were imputed, features were chosen and scaled, and classes were balanced to improve the accuracy of results. It was ensured that each feature had a standard deviation of one and a mean of zero.

Table 3. Data Quality of Dataset

Data Quality	Count	Percentage
Valid	1319	100%
Mismatched	0	0%
Missing	0	0%

As shown in Table 3, dataset was preprocessed with 1319 valid entries. There were no mismatched or missing values. According to the table, the mean value is 56.2, while the standard deviation is 13.6. The distribution of data is shown by quantiles; the minimum quantile is 14 and the maximum quantile is 103. This characterizes the data as having central tendency and variability. Data distribution and the quantiles of dataset are shown below in Table 4.

Table 4. Data Distribution and Quantiles of Dataset

Data Distribution	Value
Mean	56.2
Standard Deviation	13.6

Data Distribution	Value
Quantiles	
Minimum	14
25%	47
50% (Median)	58
75%	65
Maximum	103

2.2. Machine Learning Algorithms

The use of ML is on the increase in the field of cardiovascular medicine. Despite the availability of ML techniques, it still remains difficult to find the most practical ML algorithm for cardiovascular disease datasets [15]. The models used in this study are discussed below.

2.2.1. LASSO Regularization. Regression models are often applied in statistical analyses. They are commonly used to calculate the probability (often the expected risk) of some future occurrences or expected results [17].

LASSO regression is a common variable screening technique. The coefficient is always compressed by the addition of a penalty term to the model estimation term. The goal for it is to be simple enough, but still be able to solve the problems of overfitting and multiple correlations [18]. LASSO regression strives to find the variables and hence matches regression coefficients that lead to the lowest prediction error in a model. To do this, a constraint is applied to the model parameters, so that the sum of regression coefficients' absolute values is less than the specified value (λ) [17].

The λ with the smallest deviation is what the cross-validation curve shows. This value results in the best fitting effect of the LASSO regression model [18].

$$w = \arg\min_{w} \sum_{i=1}^{N} (y_i - w^T x_i) = (X^T X)^{-1} X^T y \tag{1}$$

Equation (1) eleborates the least squares method for linear regression, where w is model parameter, x_i is the input feature, y_i is label, N is the training sample number, X is input data matrix, y is the target outputs, and

$\underset{w}{\arg\min}$ is the minimum sum of squared errors for w. Whereas, $(X^T X)^{-1} X^T y$ is the optimal analytical solution for w [18].

2.2.2. Support Vector Machine. SVM is a very powerful and frequently used ML technique for classification and regression. Its job is to separate different classes in a dataset by attempting to find a hyperplane maximizing the margin between the classes. Shipping means that it looks for the best boundary with which to distinguish between different categories of data points. SVM identifies the subsets of data points, the so-called support vectors, during the training process. As these points define the decision boundary, they are critical. In fact, only a small fraction of the whole dataset is recycled as support vectors which makes SVM efficient [19].

To measure how closely predictions and actual outcomes match, mean absolute error or MAE is used, as shown below in equation (2).

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \tilde{y}_i| \tag{2}$$

The definition of an estimator with respect to the determined parameter θ' is the root squared of the mean square error or RMSE. It may be computed using equation (3) which predicts that lower the RMSE, the better the performace of the model [20].

$$RMSE = \sum_{i=1}^{n} \frac{\left(y_{i-\tilde{y}_i}\right)^2}{n} \tag{3}$$

In SVM, average distance can be calculated as in equation (4) which tells us how well the classes are separate [20].

$$\text{aver}D_K = \frac{\sum_{jk=1}^{m}}{\text{m}} \tag{4}$$

Recursive Feature Elimination (RFE) method is particlulary used with SVM, where some features might be irrelvent or redundent which hurts the accuracy. It is an iterative method which trains the model, as well as measures and ranks the features’ importance based on the magnitude of model coefficient w. For linear SVM, coefficient W_j directly indicates the feature importance, as shown in equation (5) below.

$$\text{Feature Importance} = |\, w_j \,| \tag{5}$$

2.2.3. Gradient Boosting Classifier. Gradient boosting (GB) is one of the most powerful ML techniques used for regression and classification tasks. The method ensembles learning, also known as boosting, by combining multiple weak learners (models that perform slightly better than random guess) into a strong learner [21, 22].

It operates in a step-by-step fashion, whereby the successor reduces the error by learning from the mistakes of the predecessor. Weights are updated at each stage of this process. The iterations keep going until the loss function is minimized. The additive model, the loss function, and weak learners are the three primary components of GB. The loss function is used to calculate the accuracy of heart disease prediction using the existing data. The additive model adds one weak learner at a time, working sequentially and iteratively. Weak learners are the ones that categorize the data poorly [21].

The aim of GB is to estimate a target function $F^*(x)$, which associates a value y with each input instance x. To minimize a loss function $L(y,F(x))$, it measures how well do predictions by the model fit to the real values of y. Moreover, it includes ways to prevent overfitting. This applies shrinkage (decreasing the contribution of each model) and reducing model complexity [22].

2.3. Performance Evaluation

2.3.1. Accuracy. Accuracy is the ratio of accurately predicted observation to the total observation [1]. Two performance metrics, namely accuracy and false alarm rate computed as equation 5, are used to assess the performance [23].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{5}$$

Here, true positive is TP, true negative is TN, false negative is FN, and false positive is FP. False alarm rate (FAR) can be calculated as shown in equation 6 below [23].

$$FAR = \frac{FP}{FP + TN} \tag{6}$$

2.3.2. Precision. It is the ratio of true positive observations to the sum of true positives and false positive observations. It is computed below as shown in equation 7.

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

Here, true positive is TP while false positive is FP.

2.3.3. Recall. It is the ratio of true positive predictions to the sum of true positive and false negative predictions. By definition, it is calculated as

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

2.3.4. F1 Score. It is the average of recall and precision. So, it can be calculated as [1]

$$F1\, Score = \frac{2(recall \times precision)}{recall + precision} \tag{9}$$

3. RESULTS

The dataset was split into training (70%), validation (15%), and testing (15%) subsets. This sort of splitting prevents overfitting and ensures fair evaluation. Out of 1319 samples in the dataset, 923 samples were trained, while 198 samples were validated and tested on models. The process of determining which hyperparameter combinations enable the model to optimize its accuracy is known as hyperparameter tuning [24]. The parameters employed for classification are shown below in Table 5.

Table 5. Hypertuning Parameters

Support Vector Machine	
Different Regularization Strengths C	[0.01, 0.1, 1, 10]
Kernal	Linear
Gamma	Scale
Top Feature Selection	5
Gradient Boosting	
n_estimators	10, 50, 100
learning_rate	0.1
max_depth	3
Subsample	1.0

Lasso Regularization	
Alpha	1.0
max_iter	1000
Tolerance	1e-4

3.1. LASSO Regularization

Lasso (L1) regularization was applied to logistic regression in order to predict the most significant feature by shrinking the irrelevant features towards zero, as shown in Table 6.

Table 6. Significant Clinical Biomarkers and their Respective Coefficient Value Predicted by Lasso Regularization

Clinical Biomarker	Coefficient Value
Troponin	78.9295
CK-MB	19.0294
Age	0.5598
Gender	0.1626
Heart Rate	0.0059
Systolic Blood Pressure	-0.0754
Diastolic Blood Pressure	0.0322
Blood Sugar	-0.1193

According to the table, troponin has the highest coefficient valuemaking it the most influential biomarker, probably because it is linked to heart-related diseases. CK-MB also plays a significant role but much lower than troponin. Even though it has less of an impact, age still influences prediction. Lasso shrinks the value of less important features to zero which helps in feature selection.

Through Lasso regularization, the bar graph classifies the significant biomarkers for heart disease prediction, showing their respective coefficients on y-axis and biomarkers as features on x-axis. The significance of these characteristics in predicting the target variable is shown by their respective Lasso regression coefficient. The stronger the correlation with the result, the higher the coefficient.

3.2. Classical Gradient Boosting

The total number of samples were 1319. The samples used for training, validation, and testing were 923, 198 and 193, respectively. Validation accuracy, checked with 50, 100, and 200 number of estimators was 0.9697, where test samples were evaluated through confusion matrix. The most effective biomarkers found by using the GB model, ordered according to how significant they were in predicting the results, are shown in Table 7 below.

Table 7. Top Biomarkers Identified via GB

Rank	Feature	Importance
7	Troponin	0.609001
6	CK-MB	0.376480
3	Systolic blood pressure	0.005626
5	Blood sugar	0.004981
1	Gender	0.001960

Based on the above table, troponin stands out as the most significant predictor with the highest significance score (~0.6) in the graph, which displays feature importance ratings determined via GB. With a significant contribution (~0.3) CK-MB comes next, emphasizing its value in the diagnosis of cardiac disorders. On the other hand, characteristics like blood sugar, systolic and diastolic blood pressure, age, gender, and heart rate remain insignificant and play only a small part in the model's predictions. According to these findings, troponin and CK-MB are the most important biomarkers for predicting cardiac-related outcomes, whereas other characteristics have a minimal bearing on the final prognosis.

3.3. Support Vector Machine

SVM was used to classify biomarkers through bar graph, as shown in Fig. 5, using recursive feature elimination (RFE). Based on the impact of biomarkers on classification, the model prioritizes diastolic blood pressure, heart rate, and blood sugar as key predictors, signifying their higher contribution to the prediction of heart diseases, as shown in Figure 5. It shows how efficiently these features separate heart disease patients from normal individuals.

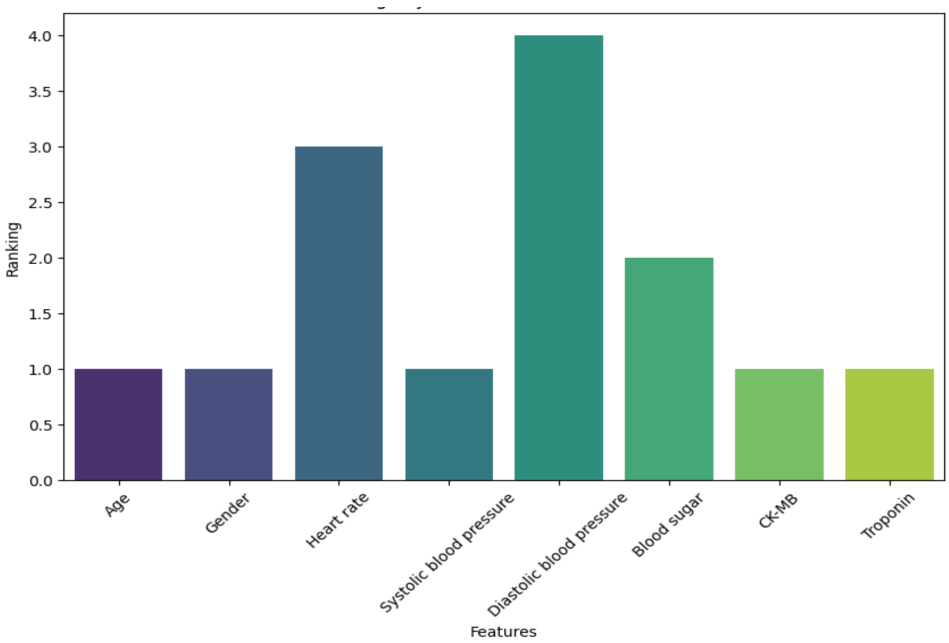


Figure 5. Biomarker Identification via SVM

After training the model, hypertuning parameters with different values of regularization were set. The best performance was achieved at C = 10, with a validation accuracy of 86.87%, showing the model's robustness in classifying heart diseases, as shown in Table 9.

Table 8. Grid Search for the Regularization Parameter

Regularization Parameter Values	Accuracy
0.01	67.68%
0.1	73.23%
1	76.77%
10	86.87%

3.4. Evaluation Matrics

These algorithms were evaluated by using confusion matrix. It is a valuable tool to measure the performance of classification models. It provides the breakdown of correct and incorrect predictions across various classes.

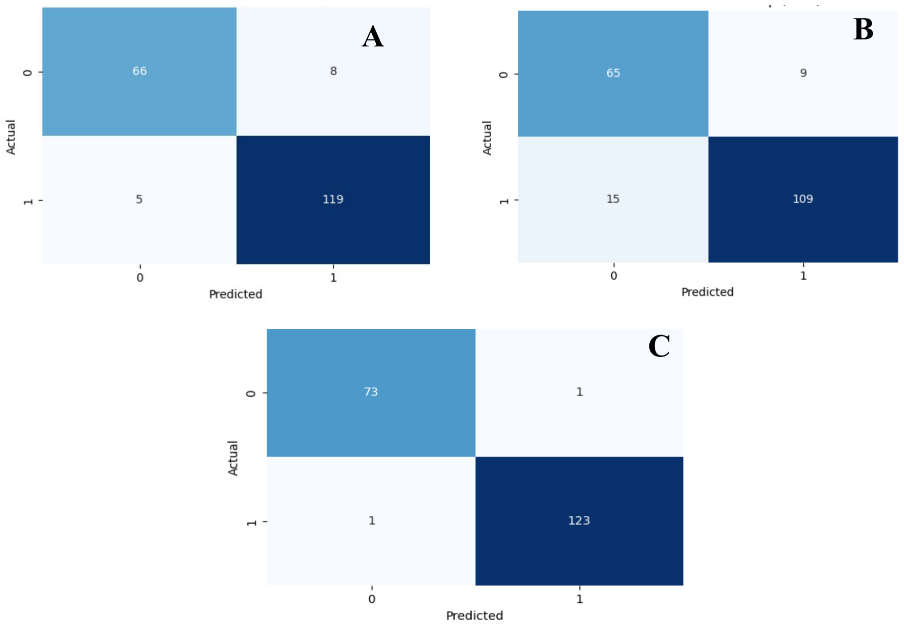


Figure 3. Confusion Matrix for Performance Evaluation (a) LASSO Regularization (b) SVM (c) GB

The confusion matrix of Lasso regularization, shown in Fig. 5 (a), revealed that 66 healthy individuals were classified correctly. On the other hand, 119 individuals with heart disease were identifed correctly. The model is robust as the rate of false negative and false positive remains low, that is, 8 and 5 misclassifications, respectively.

SVM classified 65 cases accurately as healthy individuals and 109 as diseased. On the contrary, it mistakenly identified 9 healthy individuals as diseased and failed to detect 15 cases of affected patients, as shown in Fig. 5 (b). The GB classifier obtained a high accuracy of 95% according to the confusion matrix, as shown in Fig. 5 (c). Moreover, it accurately detected 161 positive cases (true positives) and 97 negative cases (true negatives). Nevertheless, it incorrectly identified two positive cases as negative (false negatives) and four negative cases as positive (false positives). These findings show that the GB model performs an effective task of classifying attributes into appropriate classes.

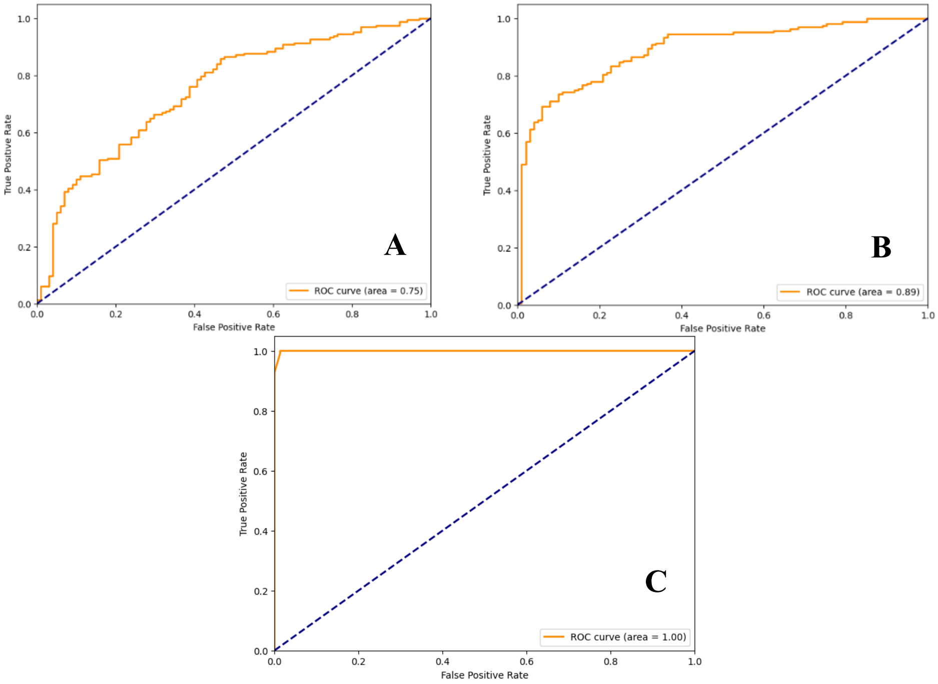


Figure 4. Receiver Operating Characteristics (ROC) Curve (a) LASSO Regularization (b) SVM (c) GB

These models were also evaluated based on the receiver operating characteristics (ROC) curve. It shows the graphical representation of the diagnostic ability of a classification model. As the classification threshold changes, this plot illustrates the trade-off between the true positive rate (sensitivity) and the false positive rate (specificity). The performance of LASSO regularization is shown in Fig. 6 (a) graphically by using the ROC curve. The model can differentiate between positive and negative classes with 75% probability, as indicated by its modest performance area under graph (AUC) of 0.75.

The ROC curve of SVM, shown in Fig. 6 (b), has an AUC of 0.96 which is close to 1. This shows the excellent performance of the model in distinguishing between positive and negative classes. The model achieves a high true positive rate (sensitivity) and a low false positives rate (specificity) as the curve moves upwards and comes close to the top-left corner of the plot. With an AUC of 1.0, the ROC curve shows that GB, as shown in Fig. 6 (c), performs exceptionally well and comes close to being a perfect classifier. This model shows specificity through a true positive rate nearly equal to 100% and a false positive rate close to 0%.

Table 9. Classification Report

Model	Class	Precision	Recall	F1-score	Support	Accuracy
Lasso Regularization	0	0.93	0.94	0.93	74	0.89
	1	0.89	0.96	0.93	124	
Support Vector Machine	0	0.81	0.88	0.84	74	0.87
	1	0.92	0.88	0.90	124	
Gradient Boosting	0	0.98	0.96	0.97	74	0.98
	1	0.98	0.99	0.98	124	

Comprehensive performance evaluation was conducted to assess the effectiveness of the models for feature selection. Metrics including accuracy, precision, recall, and F1-score were used to assess each model’s performance. In Lasso regularization, hyperparameter tunning was performed by using different values of alpha to optimize regularization strength. The validated accuracy of 89.39% is shown by 0.001. It ensures a fine balance between feature selection and predictive performance. In the final model evaluation, the test dataset showed 93.43% accuracy, ensuring a high reliability of the model in biomarker classification for heart diseases. After the training model, hypertuning parameter with different values of the regularization parameter were performed for SVM. The best performance was achieved at C = 10, with a validation accuracy of 86.87%, which shows its robustness in classifying heart diseases. This model achieves a precision of 92% and a recall of 88%. GB correctly classified the major cases with an overall accuracy of 98%. With a precision rate of 98%, it actually classified negative cases with 98% recall.

The classification report is displayed in Table 9. Performance metrics revealed that GB outperforms both Lasso regularization and SVM with the highest accuracy of 98% and balanced precision, recall, and F1-scores across both classes, making it a reliable model for classification.

4. DISCUSSION

A key element of preventive healthcare is the early detection of cardiac disease, since it enables efficient interventions that significantly improve patient diagnosis. Numerous conditions, such as hypertension, coronary artery disease, valvular heart disorders, and cardiomyopathy contribute to heart illness, which is categorized as a complicated clinical syndrome. The heart's ability to sufficiently pump blood to meet the body's needs is severely hampered by the reduced cardiac function caused by the aforementioned disorders [25].

The global disease burden of ischemic heart disease (IHD) is huge, that is, approximately 125 million individuals are affected world wide, presenting 1.72% of the global population. There were about 9 million deaths alone in 2017 due to IHD, making it a major cause of mortality. By 2030, an increase is expected in its prevalence, which is forecasted to be 1,845 per 100,000 persons due to aging and worsening lifestyle risk factors [26]. This escalating burden signifies the critical need for effective and early diagnostics.

This study found that troponin and CK-MB are consistently significant biomarkers to predict heart disease using both GB and Lasso regularization. Diastolic blood pressure and heart rate are also significant determinants of heart disease, as suggested by SVM. The difference is created by clinical biomarker identification through important feature selection across these algorithms due to their underlying feature selection mechanisms and functional framework.

Lasso regularization linearly builds the correlation between the predictor and outcomes and shrinks the value of non-important features to zero, while the dominant becomes more prominent [27]. In GB, the relationship is built non-linearly among features and weak points are boosted to equally distribute their potential and rank them according to their importance [12]. On the other hand, SVM makes feature classification by using hyperplane with nonlinear interaction [28]. Acute cardiac biomarkers, such as troponin and CK-MB, because of their strong correlation with heart disease, are classified by Lasso regularization and GB. While, hemodynamic parameters including diastolic blood pressure and heart rate, which contribute to the risk of long-term heart disease, are prioritized by SVM.

In this study, GB outperformed others with 98% accuracy. This was also demonstrated by the study of Ren et al., which predicted coronary heart index (CHI) as a non-invasive, effective, and trustworthy diagnostic tool based on an XGB-based model. The application of this novel technique may result in more accurate and successful treatment plans, which would ultimately enhance patient outcomes [29]. The Lasso regression model works well with low-dimensional data, is simple to use, and is highly effective with linear data. Nevertheless, when variables have intricate relationships, it fails to achieve high accuracy [30]. The random forest model can suffer from over-fitting if it detects noise in the training data, but it scales well with big datasets, achieves high accuracy with many decision trees, and is robust to noise. In contrast, the SVM model is more robust than the Lasso regression model, performs well in classifying semi-structured or unstructured data, and has a lower risk of overfitting. However, it is not appropriate for large datasets with many features or datasets with missing values, which results in poor generalization on new data and lack of interpretability [31].

Heart diseases comprise complex pathophysiological processes which vary greatly among individuals. Hence, it is difficult to collect all features. Consequently, the attributes employed for the ML model might not be applicable to all individuals. ML algorithms depend on multiple factors. In addition to clinical characteristics, the ultimate performance of a model is determined by its complexity, hyperparameters, and method variations.

One very specific biomarker of myocardial damage is troponin, which is released into the bloodstream during cardiac injury. Studies have repeatedly demonstrated its use as a diagnostic and prognostic tool in heart disease [32]. Myocardial CK-MB or creatine kinase, for many years, has been considered an isoenzyme of creatine kinase and used to identify myocardial necrosis. Though less specific than troponin, CK-MB remains a useful marker and, in combination with other clinical indicators, may still be useful [33]. Although these biomarkers are used commonly in clinics, research on how they could be incorporated into ML models remains scarce. The performance of various ML models varies greatly. The ideal ML model should be determined based on the the quantity and quality of the data, as well as the optimized algorithms with the best hyperparameters.

Troponin is an important component of contractile apparatus in cardiac myocytes and results in the formation of troponin-tropomyosin complex. This complex is significant for the interaction between actin and myosin during muscle contraction through calcium signaling. Ischemia and necrosis lead to myocardial injury that affects cardiac sarcolemma. This causes the release of cardiac specific Troponin T and Troponin I into the bloodstream.

This is a direct indicator of cardiomyocyte damage. Troponin is widely used as a gold standard biomarker for the diagnosis of acute myocardial infarction, since elevated levels of Troponin I and Troponin T have high sensitivity and specify to cardiomyocyte damage. During myocardial injury, troponin is present in 99% high concentration for days, acting as a diagnostic window [8]. In addition to its diagnostic role, troponin level is also associated with cardiac damage and highlights risk stratification and long-term prognosis in patients with heart disease [34].

Creatine Kinase-MB (CK-MB) is an isoenzyme of creatine kinase that acts as a valuable biomarker for myocardial injury, particularly working with troponin. CK-MB level rises in blood rapidly after myocardial damage. It peaks within 18 to 24 hours and normalizes in 2 to 3 days, making it useful in the diagnosis of recurrent ischemic event and reinfarction. It is also used in monitoring perioperative cardiac infarction and differentiating cardiac injury from skeletal muscle injury [33].

Diastolic blood pressure (DBP) and heart rate are chronic risk markers that emphasize long-term progression of cardiovascular events due to autonomous cardiac dysfunction. They are associated with coronary perfusion impairment, vascular stiffness, left ventricular hypertrophy, inflammation, and oxidative stress [34].

This study signifies the combined importance of troponin, CK-MB, and age in ML-based predictive models. Advances with computational methods enable the integration of various biomarkers with demographic and clinical data that helps to improve the reliability and accuracy of prediction.

Troponin and CK-MB are biological indicators of myocardial pathophysiology. Their elevated levels signal critical cellular processes, such as apoptosis, necrosis, and inflammation, which are the hallmark of heart diseases. The findings highlight the importance of such models being integrated within diagnostic procedures and show that clinical biomarkers may be combined with machine learning for the prediction of heart disease. The goal of future research should be to validate these results in broader and more varied groups. Finally, predictive accuracy of these models may also be improved by examining imaging data and other clinical characteristics.

4.1. Conclusion

This investigation underscores the significance of troponin and CK-MB as powerful diagnostic markers for heart conditions. A variety of ML approaches were used to consistently find these markers, namely Lasso regularization, SVM, and GB. Their use in clinical cardiology confirmed their biological value as biomarkers of myocardial damage and ischemia. Numerous algorithms showed that machine learning and clinical biomarkers can be joined to increase diagnostic precision. Future studies should attempt to improve their predictive performance by incorporating other biomarkers and validate these findings in larger, more diverse populations. Utilizing cutting edge computational techniques, this strategy offers a mechanism to more efficiently diagnose and treat heart diseases, ultimately improving patient outcomes.

CONFLICT OF INTEREST

The authors of the manuscript have no financial or non-financial conflict of interest in the subject matter or materials discussed in this manuscript.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed during this study.

FUNDING DETAILS

The authors did not receive any funding for this research.

REFERENCES

1. Shah D, Patel S, Bharti SK. Heart disease prediction using machine learning techniques. *SN Comput Sci*. 2020;1:e345. https://doi.org/10.1007/s42979-020-00365-y

2. Jevin AJ, Jayant H, Sanjay R, Hemasai V, Venkatasrinivas PV. Heart disease identification method using machine learning classification in e-healthcare. *Int J Adv Res Arts Sci Eng Manag*. 2023;10(3):2322–2327.

3. Ruetzler K, Smilowitz NR, Berger JS, et al. Diagnosis and management of patients with myocardial injury after noncardiac surgery: a scientific statement from the American Heart Association. *Circulation*. 2021;144:e287–e305. https://doi.org/10.1161/CIR.0000000000001024

4. Lakatta EG, Levy D. Arterial and cardiac aging: major shareholders in cardiovascular disease enterprises: part II: the aging heart in health: links to heart disease. *Circulation*. 2003;107(2):346–354. https://doi.org/10.1161/01.CIR.0000048893.62841.F7

5. Gheisari F, Emami M, Shahraki HR, Samipour S, Nematollahi P. The role of gender in the importance of risk factors for coronary artery disease. *Cardiol Res Pract*. 2020;2020:e527820. https://doi.org/10.1155/2020/6527820

6. Carey RM, Whelton PK. Prevention, detection, evaluation, and management of high blood pressure in adults: synopsis of the 2017 American College of Cardiology/American Heart Association hypertension guideline. *Ann Intern Med*. 2018;168(5):351–358. https://doi.org/10.7326/M17-3203

7. Yancy CW, Jessup M, Bozkurt B, et al. 2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines and the Heart Failure Society of America. *J Am Coll Cardiol*. 2017;70(6):776–803. https://doi.org/10.1016/j.jacc.2017.04.025

8. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth universal definition of myocardial infarction (2018). *J Am Coll Cardiol*. 2018;72(18):2231–2264. https://doi.org/10.1016/j.jacc.2018.08.1038

9. Rani P, Kumar R, Ahmed NMOS, Jain A. A decision support system for heart disease prediction based upon machine learning. *J Reliab Intell Environ*. 2021;7:263–275. https://doi.org/10.1007/s40860-021-00133-6

10. Saboor A, Usman M, Ali S, Samad A, Abrar MF, Ullah N. A method for improving prediction of human heart disease using machine learning algorithms. *Mob Inf Syst*. 2022;2022:e410169. https://doi.org/10.1155/2022/1410169

11. Ali MM, Paul BK, Ahmed K, Bui FM, Quinn JMW, Moni MA. Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison. *Comput Biol Med*. 2021;136:e104672. https://doi.org/10.1016/j.compbiomed.2021.104672

12. Feng M, Wang X, Zhao Z, Jiang C, Xiong J, Zhang N. Enhanced heart attack prediction using extreme gradient boosting. *J Theo Pract Eng Sci*. 2024;4(4):9–16. https://doi.org/10.53469/jtpes.2024.04(04).02

13. Wang Y. Identification of cardiovascular diseases based on machine learning. Paper presented at: International Symposium on Artificial Intelligence for Medicine Sciences, October 13–15, 2022; Amsterdam, Netherlands. https://doi.org/10.1145/3570773.3570855

14. Kumar A, Dhanka S, Sharma A, et al. A hybrid framework for heart disease prediction using classical and quantum-inspired machine learning techniques. *Sci Rep*. 2025;15:e25040. https://doi.org/10.1038/s41598-025-09957-1

15. Nadakinamani RG, Reyana A, Kautish S, et al. Clinical data analysis for prediction of cardiovascular disease using machine learning techniques. *Comput Intell Neurosci*. 2022;2023:e815067. https://doi.org/10.1155/2022/2973324

16. Rashid TA, Hassan B. Heart attack dataset. Menedleey Data Web site. Accessed April 29, 2022. http://dx.doi.org/10.17632/WMHCTCRT5V.1

17. Cook JA, Ranstam J. Overfitting. *Br J Surg*. 2016;103(13):e1814. https://doi.org/10.1002/bjs.10244

18. Wang S, Chen Y, Cui Z, Lin L, Zong Y. Diabetes risk analysis based on machine learning LASSO regression model. *J Theory Pract Eng Sci*. 2024;4(01):58–64. https://doi.org/10.53469/jtpes.2024.04(01).08

19. Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A. A comprehensive survey on support vector machine classification: applications, challenges and trends. *Neurocomputing*. 2020;408:189–215. https://doi.org/10.1016/j.neucom.2019.10.118

20. Rauf A, Ammar M, Azhar M, Noor N, Bakhtiar SM. Unveiling the Secrets of obesity with machine learning (ml) techniques/algorithms. *Curr Trend OMICS.* 2024;4(1):55–76. https://doi.org/10.32350/cto.41.04

21. Sultan SQ, Javaid N, Alrajeh N, Aslam M. Machine learning-based stacking ensemble model for prediction of heart disease with

explainable AI and K-Fold cross-validation: a symmetric approach. *Symmetry*. 2025;17(2):e185. https://doi.org/10.3390/sym17020185

22. Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. *Artif Intell Rev*. 2021;54:1937–1967. https://doi.org/10.1007/s10462-020-09896-5

23. Primartha R, Tama BA. Anomaly detection using random forest: a performance revisited. Paper presented at: 2017 International Conference on Data and Software Engineering (ICoDSE); November 1–2, 2017; Palembang, Indonesia.

24. Sharma NK, Chauhan AS, Fatima S, Saxena S. Enhancing heart disease diagnosis: leveraging classification and ensemble machine learning techniques in healthcare decision-making. *J Integr Sci Technol*. 2024;13(1):e1016. https://doi.org/10.62110/sciencein.jist.2025.v13.1016

25. Alsabhan W, Alfadhly A. Effectiveness of machine learning models in diagnosis of heart disease: a comparative study. *Sci Rep*. 2025;15:e24568. https://doi.org/10.1038/s41598-025-09423-y

26. Khera AV, Emdin CA, Drake I, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N Engl J Med*. 2016;375(24):2349–2358. https://doi.org/10.1056/nejmoa1605086

27. Albahr A, Albahar M, Thanoon M, Binsawad M. Computational learning model for prediction of heart disease using machine learning based on a new regularizer. *Comput Intell Neurosci*. 2021;2021:e628335. https://doi.org/10.1155/2021/8628335

28. Hoque R, Billah M, Debnath A, Hossain SMS, Sharif NB. Heart disease prediction using SVM. *Int J Sci Res Arch*. 2024;11(2):412–420. https://doi.org/10.30574/ijsra.2024.11.2.0435

29. Ren W, Zhang Z, Wang Y, et al. Coronary health index based on immunoglobulin light chains to assess coronary heart disease risk with machine learning: a diagnostic trial. *J Transl Med*. 2025;23:e22. https://doi.org/10.1186/s12967-024-06043-4

30. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC*

Med Informat Dec Mak 2019;19:e281. https://doi.org/10.1186/s12911-019-1004-8

31. Kwiendacz H, Huang B, Chen Y, et al. Predicting major adverse cardiac events in diabetes and chronic kidney disease: a machine learning study from the Silesia Diabetes-Heart Project. *Cardiovasc Diabetol*. 2025;24:e76. https://doi.org/10.1186/s12933-025-02615-w
32. Collet JP, Thiele H, Barbato E, et al. 2020 ESC guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. *Eur Heart J.* 2021;42(14):1289–1367. https://doi.org/10.1093/eurheartj/ehab285
33. Boulet J, Sridhar VS, Bouabdallaoui N, Tardif JC, White M. Inflammation in heart failure: pathophysiology and therapeutic strategies. *Inflamm Res*. 2024;73:709–723. https://doi.org/10.1007/s00011-023-01845-6
34. Potter JM, Hickman PE, Cullen L. Troponins in myocardial infarction and injury. *Aust Prescr*. 2022;45(2):53–57. https://doi.org/10.18773/austprescr.2022.006

ARTICLES FOR FACULTY MEMBERS

AN IMPROVED MLP-RFE MODEL FOR ENHANCED ACCURACY IN HEART DISEASE PREDICTION USING DEEP LEARNING

Machine learning and Mendelian randomization identify key lifestyle factors in coronary heart disease: A NHANES-Based study / Cui, Y. yang, Zhang, Y., Zeng, L., Li, S., Mei, X., Yang, X., Zhou, P., Xiong, L., Huang, Y., Luo, J., Wu, F., & Yue, R.

International Journal of Cardiology: Cardiovascular Risk and Prevention
Volume 27 (2025) 200536 Pages 1–11
https://doi.org/10.1016/j.ijcrp.2025.
(Database: Science Direct)

International Journal of Cardiology Cardiovascular Risk and Prevention 27 (2025) 200536

Contents lists available at ScienceDirect

International Journal of Cardiology Cardiovascular Risk and Prevention

journal homepage: www.journals.elsevier.com/international-journal-of-cardiology-cardiovascular-risk-and-prevention

Machine learning and Mendelian randomization identify key lifestyle factors in coronary heart disease: A NHANES-Based study

Yang yang Cui [a,1], Yonghong Zhang [a,1], Lang Zeng [a,1], Shikang Li [a], Xue Mei [c], Xiangmei Yang [d], Peng Zhou [b], Lijuan Xiong [b], Yijuan Huang [a], Jing Luo [a], Fenglin Wu [a], Rongchuan Yue [a,b,*]

[a] *Department of Cardiology, Affiliated Hospital of North Sichuan Medical College, Nanchong, 637000, China*
[b] *Department of Cardiology, People's Hospital of Guang 'an District, Guang 'an, 638550, China*
[c] *School of Pharmacy, Institute of Material Medica, North Sichuan Medical College, Sichuan, Nanchong, 637000, China*
[d] *Department of Emergency, Affiliated Hospital of North Sichuan Medical College, Nanchong, 637000, China*

ARTICLE INFO

Handling Editor: Dr D Levy

Keywords:
NHANES
Machine learning
Mendelian randomization
Coronary heart disease
Causal inference

ABSTRACT

Objective: This study aims to bridge the gap between predictive modeling and causal inference by utilizing lifestyle data from the National Health and Nutrition Examination Survey (NHANES) database to compare the predictive performance of multiple machine learning models for coronary heart disease (CHD). By incorporating Mendelian randomization, the study seeks to validate and identify the lifestyle variables with both predictive power and causal impact on CHD.

Methods: We extracted variables related to demographic characteristics and lifestyle from the NHANES database (2013–2018; n = 29,400). Recursive feature elimination (RFE) was employed to rank variable importance and determine the optimal feature subset. Subsequently, eight machine learning models-including Support Vector Machine (SVM), Neural Network (NN), Naive Bayes (NB), Extreme Gradient Boosting (XGBoost), Multilayer Perceptron (MLP), Generalized Linear Model (GLM), Adaptive Boosting (AdaBoost), and Decision Tree (DT)-were developed for CHD prediction. Model performance was evaluated using metrics such as accuracy, precision, sensitivity, specificity, recall, F1-score, and the Receiver Operating Characteristic (ROC) curve, with variable contributions visualized using Shapley Additive Explanations (SHAP). Additionally, Mendelian randomization (MR) was applied to distinguish associative from causal relationships, validating top predictors via Genome-Wide Association Study (GWAS)-derived genetic instruments.

Results: RFE identified age, sex, fasting blood glucose, body mass index (BMI), total cholesterol (TC) intake, sleep duration, diastolic blood pressure, and smoking as the most significant predictors of CHD. Among the models, SVM outperformed DT, AdaBoost, XGBoost, NN, MLP, NB, and GLM. The SVM model achieved the highest performance with an accuracy of 83.4 % and an AUC value of 0.909, demonstrating clinically actionable predictive power. MR confirmed causal associations for five variables: BMI (OR: 1.01, $P < 0.001$), TC (OR: 1.01, $P < 0.001$), insomnia (OR: 1.03, $P < 0.001$), diastolic blood pressure (OR: 1.20, $P < 0.001$), and smoking (OR: 1.03, $P < 0.001$), while fasting glucose showed null causality ($P > 0.05$).

Conclusion: The SVM machine learning model, based on NHANES data, enables faster and more efficient prediction of CHD. The study identified age, sex, BMI, TC intake, sleep duration, diastolic blood pressure, and smoking as the lifestyle variables with the greatest impact on CHD. This dual approach advances precision prevention by combining predictive accuracy with genetic evidence.

1. Introduction

Coronary heart disease (CHD), primarily driven by coronary atherosclerosis, remains a leading global cause of mortality despite advances in diagnosis and treatment [1]. Its rising prevalence is closely linked to population aging and unfavorable shifts in lifestyle and dietary patterns [2]. There is an urgent need for innovative strategies to effectively mitigate its prevalence, particularly through enhanced public

---

* Corresponding author. Department of Cardiology, Affiliated Hospital of North Sichuan Medical College, Nanchong, 637000, China.
E-mail address: yyc@nsmc.edu.cn (R. Yue).

[1] Yang yang Cui, Yonghong Zhang and Lang Zeng contributed equally to this work.

https://doi.org/10.1016/j.ijcrp.2025.200536
Received 11 August 2025; Received in revised form 15 October 2025; Accepted 23 October 2025

health campaigns and the promotion of healthier lifestyles. A decade-long cohort study utilizing the UK Biobank database revealed a significant association between frequent night shift work and an increased risk of CHD, with a hazard ratio (HR) of 1.22 (95 % confidence interval: 1.11–1.35) [3]. Sedentary behavior, obesity, and smoking have also been shown to significantly increase CHD incidence, while higher red meat consumption has been positively correlated with disease risk [4,5]. Although these findings underscore the multifaceted role of lifestyle—encompassing diet, physical activity, sleep, and substance use—in CHD development, observational evidence is inherently limited by confounding and reverse causation. Establishing causal relationships between specific lifestyle factors and CHD risk therefore requires more robust analytical approaches.

Machine learning is increasingly used in medicine for high-dimensional prediction and feature selection [6,7]. In cardiovascular research, gradient-boosted trees and multi-layer perceptrons reached 95 % accuracy for CHD prediction, echoing the 95 % performance recently reported for deep-learning systems that use high-risk treadmill ECG features to detect obstructive CAD and for networks that forecast short-term mortality in acute pulmonary embolism [8–10]. These studies show that machine learning can both quantify CHD risk and flag its principal drivers. However, population-based models that predict CHD from lifestyle data rarely include sleep metrics and remain genetically untested; moreover, their associations are confounded and causally unverified [11,12]. Integrating prediction with Mendelian randomization addresses these gaps.

We integrate machine learning with two-sample MR to unify prediction and causal testing. In 20,000+ NHANES adults we train and tune eight lifestyle-only models for incident CHD, select the best on nested cross-validation, then verify the top predictors by MR with publicly available GWAS summary statistics (Fig. 1). This yields an accurate, biomarker-free screening tool and a genetically supported shortlist of modifiable targets for precision prevention.

2. Study population and methods

2.1. Machine learning

2.1.1. Data source and participants

We used the 2013–2018 NHANES cycles (n = 29,400), a cross-sectional, nationally representative survey conducted by the CDC. NHANES records self-reported physician diagnoses of CHD, lifestyle, diet, and laboratory data; all participants gave informed consent and the NCHS Research Ethics Review Board approved the protocol [13].

2.1.2. Data preprocessing

The inclusion criteria were as follows: (1) participants aged ≥18 years; (2) participants who completed the questionnaire regarding whether a doctor had informed them of having coronary heart disease (CHD), with complete questionnaire content and data. The exclusion criteria were: (1) participants aged <18 years; (2) participants with a CHD status of 9 (indicating an uncertain CHD status) based on the NHANES questionnaire [14]. Ultimately, 16,997 participants were included in the analysis. We incorporated the latest cardiovascular health metrics updated by the American Heart Association—Life's Essential 8 (LE8) [15]—and included 21 variables from four domains:

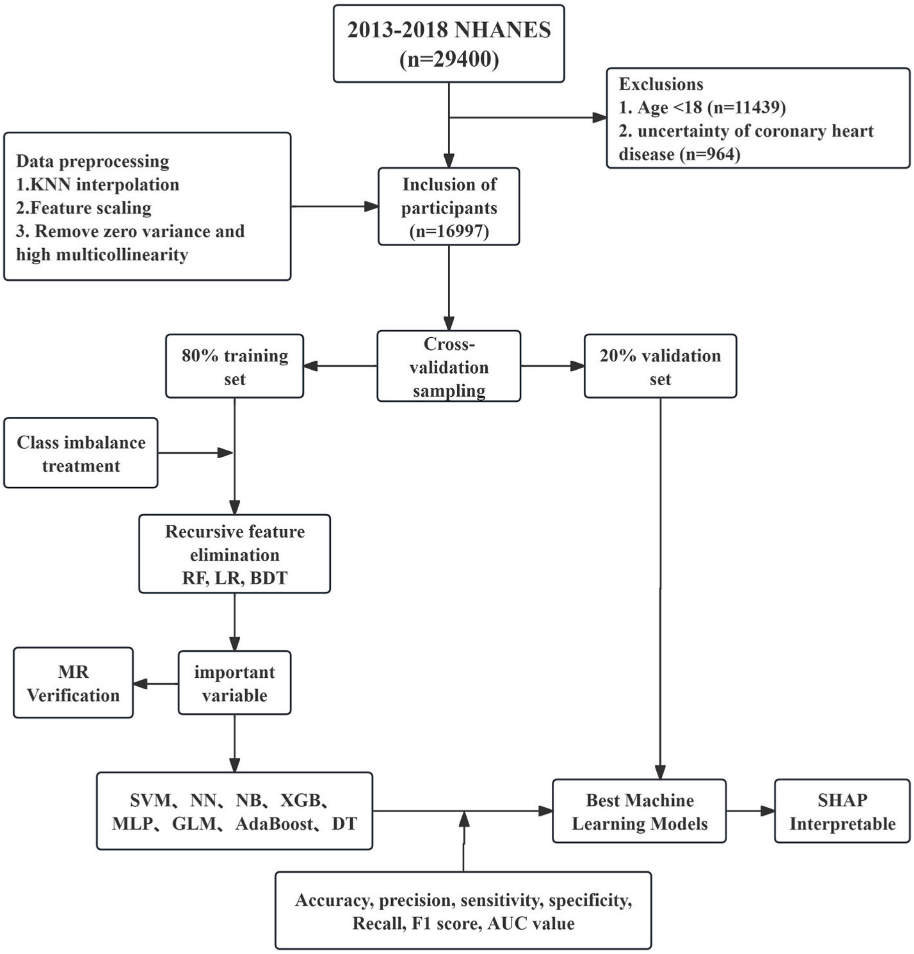


Fig. 1. Study design flow The flow shows the process of screening and machine learning the NHANES database data, including the validation of Mendelian randomization.

(1) Demographic data (age, gender, education level, race, marital status); (2) Questionnaire data (physical activity, sleep status, smoking, and alcohol consumption); (3) Examination data (body mass index [BMI], poverty index, fasting blood glucose, systolic and diastolic blood pressure); and (4) Dietary data (energy, protein, carbohydrates, total sugar, dietary fiber, total cholesterol [TC], and total fat). The diagnosis of CHD was based on participants' self-reported questionnaire data indicating that a doctor had informed them of having CHD. All analyses were performed without sample weights because the aim was to examine biological associations rather than produce nationally representative estimates.

We first conducted a statistical analysis of missing values and chose the K-Nearest Neighbors (KNN) imputation method (with $K = 5$) to complete the dataset. This method is more accurate than simple imputation techniques (e.g., mean or median imputation), especially for small to medium-sized datasets [16]. To meet the data format requirements of machine learning algorithms, we normalized the data to a range of (0, 1) using feature scaling [17]. We removed variables with zero variance and those exhibiting high multicollinearity to enhance model performance, interpretability, and reduce the risk of overfitting [18]. The dataset was divided using K-fold cross-validation with $K = 5$, resulting in 4 training sets and 1 test set. The final training set contained 13,598 samples, and the validation set contained 3399 samples [19]. To address the significant disparity between positive (CHD) and negative (non-CHD) outcomes, we employed the Random Over-Sampling Examples (ROSE) method to balance the outcome variable, thereby reducing the numerical gap [20–22]. We utilized three common recursive feature elimination (RFE) algorithms-Random Forest (RF), Logistic Regression (LR), and Bagging Decision Trees (BDT)-to rank the importance of variables. These algorithms evaluate feature importance by iteratively building models and eliminating less important features, ultimately identifying the optimal feature subset [23].

2.1.3. Machine learning framework

The dataset was divided into 80 % for training ($n = 13{,}598$) and 20 % for testing ($n = 3399$). We employed eight different machine learning models to identify CHD caused by lifestyle exposures: Support Vector Machine (SVM), Neural Network(NN), Naive Bayes (NB), XGBoost, Multilayer Perceptron (MLP), Generalized Linear Model (GLM), Adaptive Boosting (AdaBoost), and Decision Tree (DT). Each machine learning algorithm has its unique characteristics, as described below: SVM is a maximum-margin classifier whose computational complexity grows polynomially with both the number of features and samples [24]. NN is a multilayered, non-linear model that captures high-dimensional interactions by mimicking neuronal signal transmission [25]. NB is a Bayes-theorem-based probabilistic classifier that assumes conditional independence of features; it is computationally light and scales well to high-dimensional data [26]. XGBoost is a gradient boosting algorithm that can handle imbalanced data by adjusting weights. It is widely used for classification problems [27]. MLP is a feedforward artificial neural network that effectively captures nonlinear relationships, making it an ideal choice for classifying complex and multifactorial diseases [28]. GLM is a flexible extension of linear regression that models non-normal dependent variables by linking predictors to outcomes. It can represent binary and count outcomes through different conditional distributions and functional relationships [29]. AdaBoost is a boosting method that typically achieves high training accuracy. However, it may reduce classification accuracy for imbalanced data and increase computational complexity [30]. DT is a A nonparametric supervised learning method that is easy to understand and interpret, supports visual analysis, but is prone to overfitting [31].

2.1.4. Model evaluation

To evaluate and compare the predictive performance of the machine learning models, we employed various evaluation metrics, including: Accuracy and 95 % Confidence Interval (95 % CI), Precision, Sensitivity, Specificity, Recall, F1 Score, ROC Curve and AUC [32]. In addition to the aforementioned evaluation metrics, we introduced a confusion matrix to provide a more intuitive visualization of the predictions made by the machine learning models.

2.1.5. Model optimization

The best-performing model was re-trained with an exhaustive grid-search of candidate hyper-parameters; each configuration was evaluated by 10-fold cross-validation to stabilise performance estimates. Calibration plots (predicted vs. observed probability) and decision-curve analysis were then generated to confirm that the final model supplies both well-calibrated probabilities and net clinical benefit across the range of threshold probabilities encountered in practice.

2.1.6. Statistical methods

In this study, data selection, the establishment of machine learning models, model optimization, and the creation of key visualizations were conducted using R version 4.4.2. A seed number of 1234 was set for all analyses to ensure reproducibility.

2.2. Mendelian randomization

2.2.1. Concept introduction

Large-scale genome-wide association studies conducted over the past decade have identified numerous genetic variations associated with cardiac metabolic traits and risk factors. These findings have enabled the application of Mendelian Randomization (MR), a method that uses genetic variations as instrumental variables for causal inference between exposure and outcome, thereby determining whether the presumed risk factors might have a causal impact on disease [33]. Based on the selection of variables that significantly influence CHD from the optimal machine learning models, this study attempts to utilize Mendelian Randomization to validate the causal relationships, thereby enhancing its persuasive power and clinical applicability.

2.2.2. GWAS data collection and processing

Samples were obtained by accessing the IEU OpenGWAS project (mrcieu.ac.uk), collecting Genome-Wide Association Study (GWAS) data for the important variables identified through machine learning. The data for the exposure variables were preprocessed first, applying a threshold of $P < 10^{-8}$ to filter single nucleotide polymorphisms (SNPs) significantly associated with the exposure variables. To ensure independence among the SNPs and to avoid the confounding effects of linkage disequilibrium, a linkage disequilibrium coefficient ($r^2 = 0.001$) was established, limiting the correlation among adjacent SNPs. Additionally, to restrict the analysis scope, a region width of 10,000 kb was set, including only those SNPs within a certain distance from the target SNP, thereby excluding the influence of pleiotropy on the results to obtain instrumental variables related to the exposure variables [34]. After merging the exposure and outcome factors, a threshold of $P < 10^{-8}$ was again applied to exclude SNPs that were directly associated with the outcome variable, CHD. The final dataset thus formed constitutes the requisite instrumental variables for further analysis and research in this study [35](Supplementary Table 1).

2.2.3. Mendelian randomization analysis

This study employed five regression models for Mendelian Randomization: MR-Egger regression, weighted median estimation (WME), inverse variance weighting (IVW), a simple model, and a weighted model. SNPs were used as instrumental variables to validate the causal relationship between the exposure variables and the outcomes. MR-Egger regression is a causal inference method that estimates the causal effect based on instrumental variables while accounting for potential instrument bias. The WME is another non-parametric approach that can reduce potential biases by excluding data based on the median. IVW estimates the overall effect by averaging the weights based on the variance of each study. The simple model is a linear regression method that establishes relationships by either weighting or not weighting the data, aiming to estimate the genetic correlation effect. The weighted model provides more accurate estimates by considering the sample size and error variance of each study.

2.2.4. *Sensitivity analysis*

Heterogeneity across SNPs was quantified with Cochran's Q ($P < 0.05$ indicates significant heterogeneity); when present, random-effects IVW was prioritised [36]. Directional horizontal pleiotropy was examined with the MR-Egger intercept [37]. Robustness was further assessed by leave-one-out analysis: exclusion of any single SNP should not materially alter the causal estimate [38].

2.2.5. *Statistical methods*

In this study, data import, instrumental variable selection, Mendelian Randomization analysis, and sensitivity analysis were all conducted using the TwoSampleMR package within R version 4.4.2. Key parameters and assumptions were adjusted to explore variations in the results. Additionally, a significance level of $\alpha = 0.05$ was adopted throughout the analysis.

3. Results

3.1. *Machine learning component*

3.1.1. *Baseline characteristics of NHANES participants*

As is shown in Table 1, the characteristics of study participants with and without CHD from the NHANES conducted between 2013 and 2018 were summarized. A total of 16,997 participants were included in the analysis, of which 48.1 % were male, with an average age of 50.0 years. Among the participants, 741 individuals were diagnosed with CHD (4.5 %). Notably, CHD patients were predominantly male, older, and primarily non-Hispanic white compared to controls (all $P < 0.05$).

3.1.2. *Preliminary selection of important variables*

The results of the recursive feature elimination algorithm for the ranking of important variables are presented in Supplementary Fig. 1. After synthesizing the top five important variables from three different algorithms, eight significant variables were identified for further analysis in the machine learning models. These variables include age, sex, fasting blood glucose, BMI, TC, sleep duration, diastolic blood pressure, and smoking status.

3.1.3. *Establishment and validation of machine learning models*

After the preliminary selection of the eight important variables, eight different machine learning models-SVM, NN, NB, XGBoost, MLP, GLM, AdaBoost, and DT-were developed. The evaluation metrics for each model were computed, and confusion matrices were generated (Supplementary Fig. 2). A comparison of the various parameters is provided in Table 2, revealing that the SVM model exhibited the best AUC performance at 0.909 (Fig. 2A) and an average accuracy rate of 83.4 % (82.8 %–84.1 %), indicating that it performs well in identifying CHD. When the validation dataset was input into the SVM model, the AUC remained high at 0.835, suggesting that even with unfamiliar data, the SVM model is capable of maintaining strong recognition performance, further confirming its stability and generalizability.

3.1.4. *Evaluation and optimization of the best model*

The optimal machine learning model for predicting CHD was identified as the SVM. In machine learning models, a higher Youden's J statistic indicates a stronger overall classification capability [39]. The Youden's J statistic for the SVM model was found to be 0.517, with the best cut-off value achieving a sensitivity of 79.7 % and specificity of 87.4 % (Fig. 2B). These results demonstrate the model's solid performance in distinguishing positive cases (those with CHD) from negative

Table 1
Baseline table of participants.

	[ALL]	0	1	p. overall
	N=16,997	*N=16,256*	*N=741*	
Age	50.0 (17.7)	49.1 (17.5)	68.7 (10.7)	<0.001
Gender:				<0.001
Male	8169 (48.1 %)	7678 (47.2 %)	491 (66.3 %)	
Female	8828 (51.9 %)	8578 (52.8 %)	250 (33.7 %)	
Edu:				0.003
<High school	1612 (9.50 %)	1519 (9.36 %)	93 (12.6 %)	
Completed high school	2095 (12.3 %)	1991 (12.3 %)	104 (14.1 %)	
>High school	13,267 (78.2 %)	12,724 (78.4 %)	543 (73.4 %)	
Race:				<0.001
Mexican American	2486 (14.6 %)	2419 (14.9 %)	67 (9.04 %)	
Other Hispanic	1789 (10.5 %)	1726 (10.6 %)	63 (8.50 %)	
Non-Hispanic White	6242 (36.7 %)	5819 (35.8 %)	423 (57.1 %)	
Non-Hispanic Black	3664 (21.6 %)	3552 (21.9 %)	112 (15.1 %)	
Other Race	2816 (16.6 %)	2740 (16.9 %)	76 (10.3 %)	
Marital:				0.078
Married/Living with partner	10,053 (59.2 %)	9638 (59.3 %)	415 (56.0 %)	
Widowed/Divorced/Separated/Never married	6933 (40.8 %)	6607 (40.7 %)	326 (44.0 %)	
PIR	2.49 (1.62)	2.50 (1.63)	2.33 (1.49)	0.003
BMI	29.5 (7.21)	29.5 (7.24)	30.2 (6.66)	0.008
Smoke:				<0.001
No	9775 (57.5 %)	9497 (58.4 %)	278 (37.5 %)	
Yes	7222 (42.5 %)	6759 (41.6 %)	463 (62.5 %)	
Alcohol:				0.930
No	12,957 (90.6 %)	12,385 (90.6 %)	572 (90.8 %)	
Yes	1342 (9.39 %)	1284 (9.39 %)	58 (9.21 %)	
Sleep	7.43 (2.35)	7.42 (2.26)	7.69 (3.81)	0.062
Physical_Activity:				0.011
Inactive	9961 (58.6 %)	9504 (58.5 %)	457 (61.7 %)	
Moderate	3514 (20.7 %)	3349 (20.6 %)	165 (22.3 %)	
Vigorous	728 (4.28 %)	708 (4.36 %)	20 (2.70 %)	
Both moderate and vigorous	2794 (16.4 %)	2695 (16.6 %)	99 (13.4 %)	
Systolic	125 (18.9)	125 (18.6)	133 (22.3)	<0.001
Diastolic	70.6 (13.1)	70.8 (12.9)	65.7 (14.8)	<0.001
FBG	6.20 (2.11)	6.15 (2.04)	7.20 (3.03)	<0.001
Energy	2102 (1007)	2110 (1012)	1922 (857)	<0.001
Protein	81.0 (44.0)	81.3 (44.2)	72.8 (37.0)	<0.001
Carbohydrate	249 (126)	250 (127)	229 (112)	<0.001
Sugars	107 (76.5)	107 (76.6)	101 (72.8)	0.026
Dietary_fiber	17.0 (10.9)	17.0 (10.9)	16.1 (10.6)	0.040

(*continued on next page*)

Table 1 (*continued*)

	[ALL]	0	1	p. overall
	N=16,997	*N=16,256*	*N=741*	
Fat	82.5 (48.3)	82.7 (48.5)	77.4 (44.5)	0.004
Cholesterol	305 (250)	306 (252)	278 (210)	0.001

† "Cholesterol" in NHANES 2013–2018 denotes dietary cholesterol estimated from 24-h dietary recalls; it does not represent serum total cholesterol.

cases (those without CHD). Subsequently, hyperparameter tuning was performed using a grid search algorithm, with the sigma parameter set to (0.01, 0.02, 0.05, 0.1) and the C parameter set to (1, 2, 3, 4, 5). This parametric setup systematically evaluated the performance of the SVM model under different combinations for the classification task, ultimately selecting the optimal parameter combination to enhance model accuracy. Following the optimization, evaluation on the test dataset yielded an AUC of 0.909, while the AUC for the validation dataset improved to 0.855. This result not only reflects a significant enhancement in the model's accuracy and generalizability but also indicates that the optimized model is more effective at differentiating between CHD patients and non-patients.

To further evaluate the performance of the optimal SVM model, we generated a calibration curve, as shown in Fig. 2C. The curve demonstrates a good fit with the diagonal line (dashed), indicating a high consistency between the model's predictions and the actual outcomes. Additionally, we plotted a decision curve to assess the net benefit of the model. As illustrated in Fig. 2D, the prediction curve is above both the All and None lines across most of the range (0, 1), suggesting that the model holds practical utility in clinical predictions.

3.1.5. SHAP interpretability analysis

SHAP was employed to visually depict the impact of significant variables on CHD. Fig. 3 illustrates the influence of each variable on the detection dataset's ability to identify CHD within the SVM model. On the left panel of the figure, the variables are ranked in order of importance based on their SHAP values, from top to bottom. The right panlee of the figure shows that yellow indicates higher parameter values, while purple represents lower values. Furthermore, a waterfall plot (Supplementary Fig. 3) is presented to exemplify a single prediction. In the plot, the estimated model value is 0.823, while the predicted value for the sample is 0.505. The variables contributing most significantly to the model are age, sex, smoking status, BMI, and fasting blood glucose levels.

3.2. Mendelian randomization component

3.2.1. Instrumental variable selection for mendelian randomization

Following the identification of eight significant factors affecting CHD through machine learning, SHAP interpretability analysis was used to rank these factors by importance. Due to the consensus that age and sex are critical factors influencing CHD and the difficulties in obtaining relevant SNPs for these variables from the GWAS data, the remaining six factors were selected for MR analysis. The GWAS ID for the exposure variables, along with the number of SNPs excluded during the selection process due to duplicate naming and palindrome SNPs with intermediate alleles, is summarized in Table 3. The outcome variable was standardized as CHD, with the GWAS ID designated as ukb-d-I9_CHD.

3.2.2. Results and testing of mendelian randomization

The results of the five Mendelian Randomization analyses for fasting blood glucose indicated $P > 0.05$ (Fig. 4), suggesting a lack of statistical significance. Furthermore, both the heterogeneity test and the horizontal pleiotropy test yielded results of $P > 0.05$ (Table 3), indicating that there is no evidence of heterogeneity or horizontal pleiotropy in the analysis.

The results of the five Mendelian Randomization analyses for diastolic blood pressure revealed $P < 0.05$ (Fig. 4), indicating statistical significance. The heterogeneity test yielded $P < 0.001$, suggesting the presence of certain heterogeneity, which prompts us to focus on the random IVW results. The horizontal pleiotropy test indicated $P > 0.05$, demonstrating no evidence of horizontal pleiotropy (Table 3). The scatter plot (Supplementary Fig. 4A) suggests a consistent directional effect among the five methods, indicating that an increase in diastolic blood pressure is associated with an elevated risk of developing CHD. The leave-one-out analysis results (Supplementary Fig. 5A) show all error bars on one side of zero, while the funnel plot (Supplementary Fig. 6A) reflects a relatively even distribution of points around the IVW method, confirming the robustness of the study. These results suggest that diastolic blood pressure exerts a causal effect on CHD, with an odds ratio (OR) of 1.02 (1.01, 1.02).

In the five MR analyses for insomnia, the *P*-values for the IVW and WME methods were less than 0.05, while the *P*-values for the remaining three methods were greater than 0.05, indicating partial statistical significance (Fig. 4). The heterogeneity test yielded a $P < 0.05$, suggesting the presence of some heterogeneity, which indicates a need to focus on the random IVW results. The horizontal pleiotropy test indicated $P > 0.05$, demonstrating no evidence of horizontal pleiotropy (Table 3). In the scatter plot (Supplementary Fig. 4B), the direction of the MR-Egger method was opposite to that of the other four methods; however, it did not achieve statistical significance. In contrast, both the IVW and WME methods showed consistent directionality, suggesting that as the rate of insomnia increases, the risk of developing CHD also rises. The leave-one-out analysis (Supplementary Fig. 5B) revealed that all error bars were confined to one side of zero, and the funnel plot (Supplementary Fig. 6B) illustrated a relatively even distribution of points around the IVW method, reinforcing the robustness of the findings. The results indicate that insomnia has a causal effect on CHD, with OR of 1.03 (1.01, 1.04).

In the analyses for BMI, the *P*-values for MR-Egger, IVW, and WME were less than 0.05, while the *P*-values for the remaining two methods were greater than 0.05, indicating partial statistical significance (Fig. 4). The heterogeneity test yielded a $P < 0.001$, suggesting the presence of significant heterogeneity, which underscores the importance of focusing on the random IVW results. The horizontal pleiotropy test indicated $P > 0.05$, demonstrating no evidence of horizontal pleiotropy (Table 3). In the scatter plot (Supplementary Fig. 4C), the directions of the MR-Egger,

Table 2

Comparison of machine learning performance of different models.

Models	Accuracy(95 % CI)	Precision	Sensitivity	Specificity	Recall	F1	AUC
LR	0.747(0.739,0.754)	0.733	0.782	0.710	0.782	0.757	0.812
DT	0.791(0.784,0.797)	0.770	0.834	0.747	0.834	0.801	0.862
XGBoost	0.775(0.768,0.782)	0.752	0.826	0.723	0.826	0.788	0.855
AdaBoost	0.780(0.773,0.787)	0.758	0.827	0.732	0.827	0.791	0.861
NB	0.754(0.747,0.761)	0.741	0.787	0.720	0.787	0.763	0.827
MLP	0.761(0.753,0.768)	0.732	0.829	0.691	0.829	0.777	0.830
NN	0.770(0.762,0.777)	0.741	0.834	0.704	0.834	0.785	0.846
SVM	0.834(0.828,0.841)	0.808	0.880	0.787	0.880	0.843	0.909

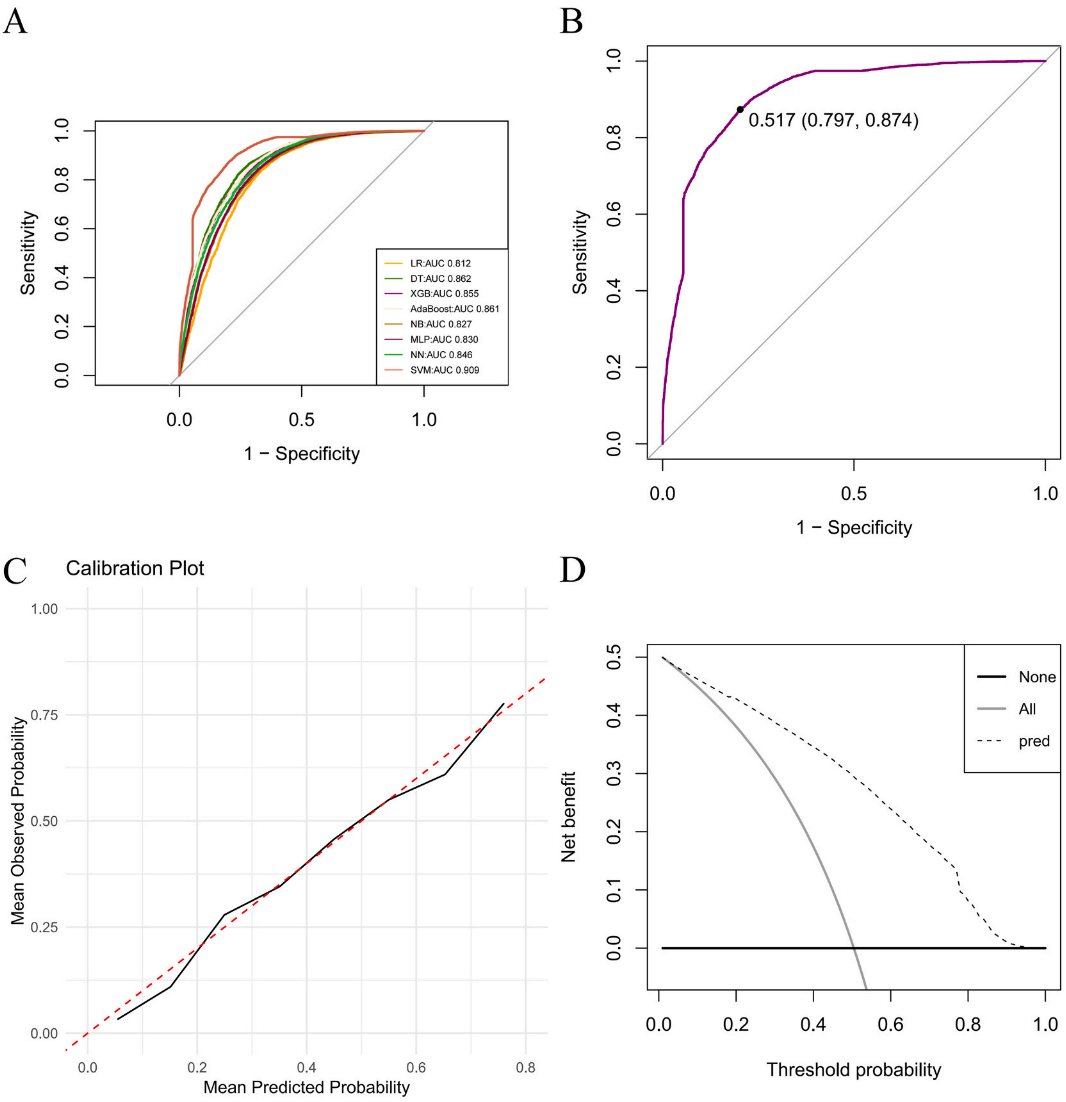


Fig. 2. Plot (A) **Receiver operating characteristic (ROC) curves of different machine learning models** The curves show the predictive performance of logistic regression (LR; AUC = 0.812), decision tree (DT; AUC = 0.862), Extreme Gradient Boosting (XGBoost; AUC = 0.855), Adaptive Boosting (AdaBoost; AUC = 0.861), Naive Bayes (NB; AUC = 0.827), multilayer perceptron (MLP; AUC = 0.830), neural network (NN; AUC = 0.846), and support vector machine (SVM; AUC = 0.909), with sensitivity versus 1-specificity. Different colors represent different machine learning models. Plot (B) **Youden's index analysis of the classifier performance** The ROC curve (SVM) demonstrates the trade-off between sensitivity (0.797) and specificity (0.874), with the optimal cutoff threshold of 0.517 identified by Youden's index. Plot (C) **Calibration curve** The curve illustrates the relationship between predicted values from the model and actual observed outcomes. Plot (D) **Decision curves** The decision curves display the net benefits of using the predictive model at various threshold probabilities. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

IVW, and WME methods were consistent, indicating that as BMI increases, the risk of developing CHD also rises. The leave-one-out analysis (Supplementary Fig. 5C) showed that all error bars were on one side of zero, and the funnel plot (Supplementary Fig. 6C) reflected a relatively even distribution of points around the IVW method, further confirming the robustness of the findings. The results suggest that an increase in BMI has a causal effect on CHD, with an OR of 1.01 (1.01, 1.02).

The results of the five Mendelian Randomization analyses for TC indicated *P*-values less than 0.05, establishing statistical significance (Fig. 4). The heterogeneity test yielded a $P < 0.001$, suggesting significant heterogeneity, which highlights the need to focus on the random IVW results. The horizontal pleiotropy test returned $P > 0.05$, indicating no evidence of horizontal pleiotropy (Table 3). The scatter plot (Supplementary Fig. 4D) reveals consistent directional effects across all five methods, suggesting that an increase in TC is associated with a heightened risk of developing CHD. In the leave-one-out analysis (Supplementary Fig. 5D), all error bars were located on one side of zero, and the funnel plot (Supplementary Fig. 6D) displayed a relatively even distribution of points around the IVW method, further confirming the robustness of the findings. The results indicate that an increase in total cholesterol has a causal effect on CHD, with an OR of 1.20 (1.14, 1.26).

In the MR analyses for smoking, the *P*-values for IVW, WME, and the weighted model were all less than 0.05, indicating statistical significance. In contrast, the *P*-values for the remaining two methods were greater than 0.05, suggesting partial statistical significance (Fig. 4). The heterogeneity test yielded a $P < 0.05$, indicating the presence of significant heterogeneity, which emphasizes the importance of focusing on the random IVW results. The horizontal pleiotropy test indicated $P > 0.05$, demonstrating no evidence of horizontal pleiotropy (Table 3). The scatter plot (Supplementary Fig. 4E) illustrates consistent directional effects across all five methods, indicating that an increase in smoking is associated with an elevated risk of developing CHD. The leave-one-out analysis (Supplementary Fig. 5E) showed that all error bars were positioned on one side of zero, and the funnel plot (Supplementary Fig. 6E)

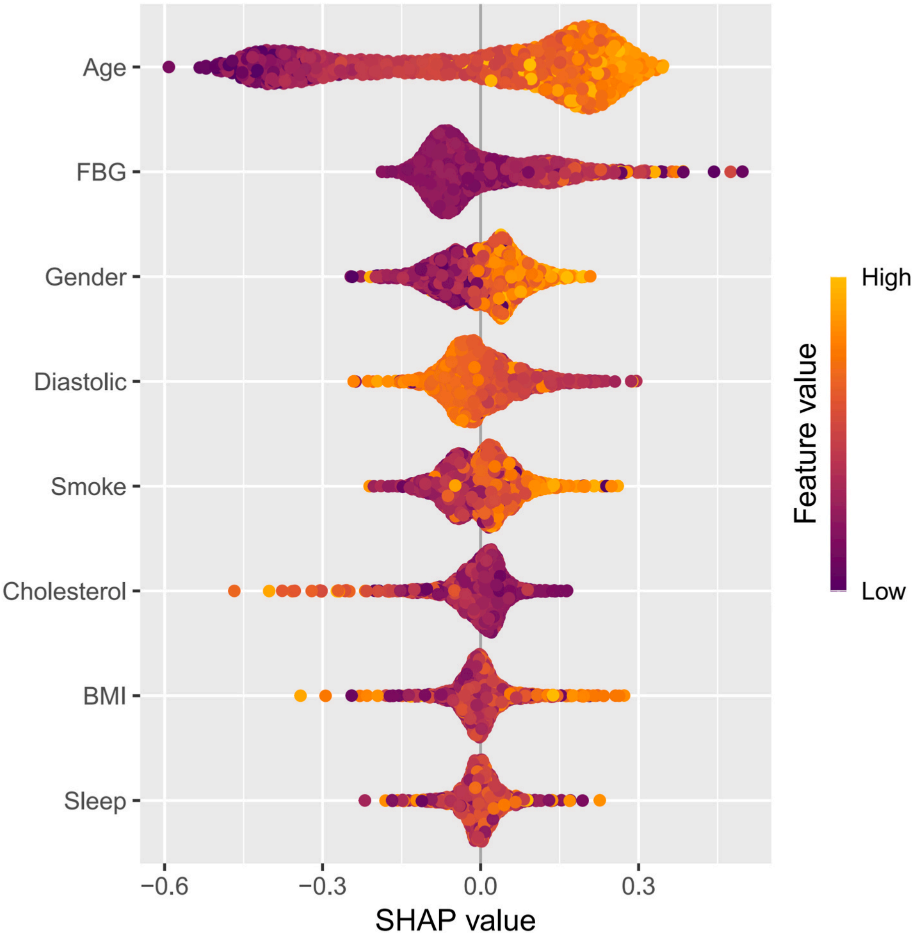


Fig. 3. SHAP Interpretable This figure displays the SHAP values for the features used in the predictive model. On the left side, features are ranked in order of importance from top to bottom based on their SHAP values. The right side of the plot employs a color gradient, where yellow indicates higher parameter values and purple signifies lower parameter values. This visualization helps to elucidate each feature's contribution to the model's predictions and highlights the variables that have the most significant impact on the outcome. Caution: The negative SHAP values for DBP reflect reverse causation—low DBP is driven by advanced CHD or its treatment (see Discussion and Supplementary Table 2) and should not be interpreted as a protective risk factor. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Statistical table of exposure variables.

Variables	GWAS ID	Unit of Measurement	Exclusion of SNPs	Final SNPs	Heterogeneity P-value	Horizontal polytropic p-value
FBG	ebi-a-GCST000568	Per 1 mmol/L	0	14	0.066	0.404
Diastolic	ebi-a-GCST90000059	Per 10 mmHg	4	199	<0.001	0.818
Insomnia	ukb-b-3957	Odds (Ever vs. Never)	3	39	0.007	0.078
BMI	ukb-b-19953	Per 1 SD (kg/m^2)	17	441	<0.001	0.083
TC	ieu-a-301	Per 1 SD (mmol/L)	2	86	<0.001	0.155
Smoke	ebi-a-GCST90029014	Odds (Ever vs. Never)	9	120	0.038	0.333

demonstrated a relatively even distribution of points around the IVW method, reinforcing the robustness of the findings. The results suggest that increased smoking is causally linked to CHD, with an OR of 1.03 (1.02, 1.04).

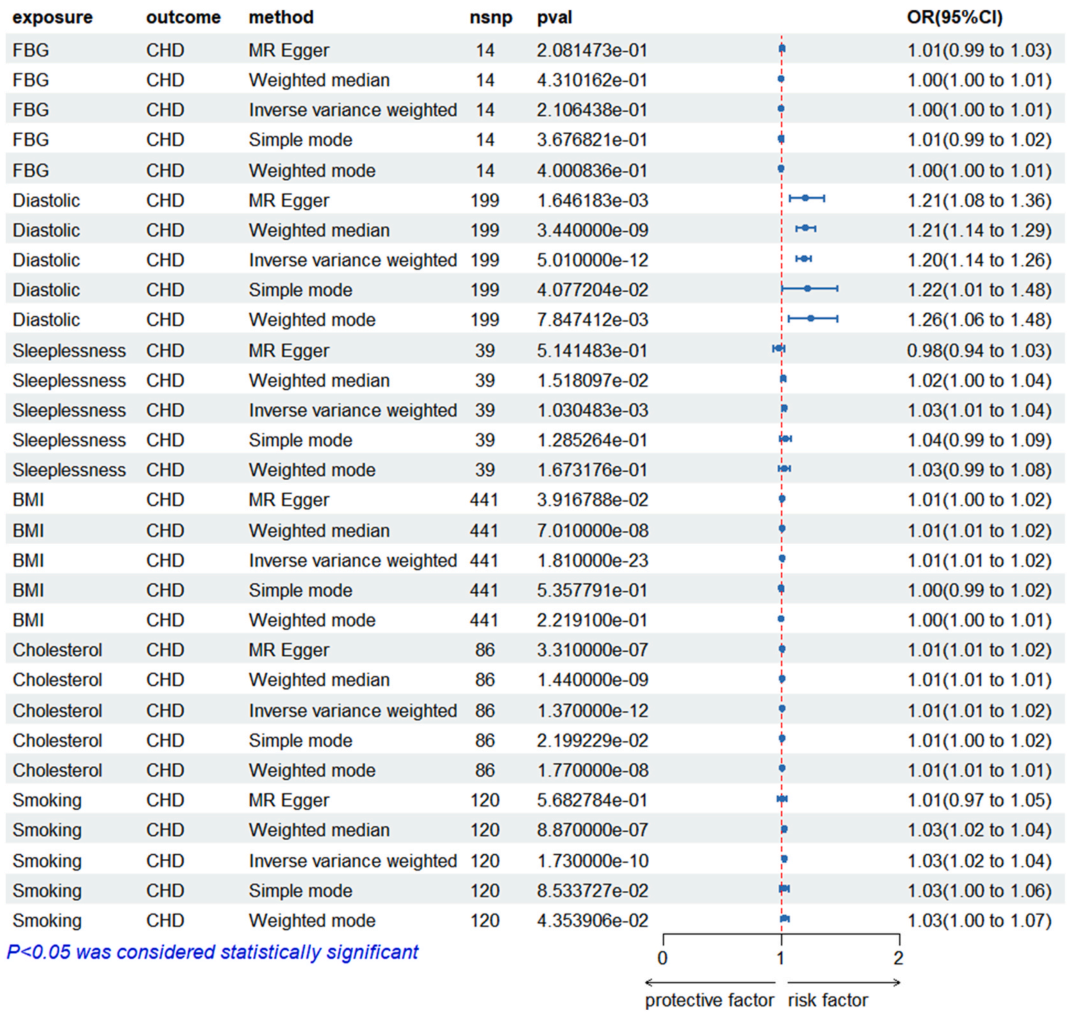
4. Discussion

4.1. Key findings

This study employed eight common machine learning methods, including SVM, NN, NB, XGBoost, MLP, GLM, AdaBoost, and DT, to predict CHD based on lifestyle-related variables. Among them, SVM showed the best performance among the eight machine learning models, achieving an AUC of 0.909 and an accuracy of 83.4 %. The results of MR analyses demonstrated positive associations for five variables: diastolic blood pressure, insomnia, BMI, TC, and smoking. These findings suggest that as the parameters of these five variables increase, the risk developing CHD also rises, indicating that these variables significantly influence the occurrence and progression of the disease. Notably, fasting glucose showed predictive but not causal association (P = 0.21), suggesting its role as a biomarker rather than therapeutic target - a critical distinction enabled by our dual-method design.

4.2. Comparative performance analysis and innovation

Numerous studies have previously attempted to establish machine learning models for predicting CHD. Liu and Zengjing developed a CHD risk prediction model for patients with human immunodeficiency virus



exposure	outcome	method	nsnp	pval	OR(95%CI)
FBG	CHD	MR Egger	14	2.081473e-01	1.01(0.99 to 1.03)
FBG	CHD	Weighted median	14	4.310162e-01	1.00(1.00 to 1.01)
FBG	CHD	Inverse variance weighted	14	2.106438e-01	1.00(1.00 to 1.01)
FBG	CHD	Simple mode	14	3.676821e-01	1.01(0.99 to 1.02)
FBG	CHD	Weighted mode	14	4.000836e-01	1.00(1.00 to 1.01)
Diastolic	CHD	MR Egger	199	1.646183e-03	1.21(1.08 to 1.36)
Diastolic	CHD	Weighted median	199	3.440000e-09	1.21(1.14 to 1.29)
Diastolic	CHD	Inverse variance weighted	199	5.010000e-12	1.20(1.14 to 1.26)
Diastolic	CHD	Simple mode	199	4.077204e-02	1.22(1.01 to 1.48)
Diastolic	CHD	Weighted mode	199	7.847412e-03	1.26(1.06 to 1.48)
Sleeplessness	CHD	MR Egger	39	5.141483e-01	0.98(0.94 to 1.03)
Sleeplessness	CHD	Weighted median	39	1.518097e-02	1.02(1.00 to 1.04)
Sleeplessness	CHD	Inverse variance weighted	39	1.030483e-03	1.03(1.01 to 1.04)
Sleeplessness	CHD	Simple mode	39	1.285264e-01	1.04(0.99 to 1.09)
Sleeplessness	CHD	Weighted mode	39	1.673176e-01	1.03(0.99 to 1.08)
BMI	CHD	MR Egger	441	3.916788e-02	1.01(1.00 to 1.02)
BMI	CHD	Weighted median	441	7.010000e-08	1.01(1.01 to 1.02)
BMI	CHD	Inverse variance weighted	441	1.810000e-23	1.01(1.01 to 1.02)
BMI	CHD	Simple mode	441	5.357791e-01	1.00(0.99 to 1.02)
BMI	CHD	Weighted mode	441	2.219100e-01	1.00(1.00 to 1.01)
Cholesterol	CHD	MR Egger	86	3.310000e-07	1.01(1.01 to 1.02)
Cholesterol	CHD	Weighted median	86	1.440000e-09	1.01(1.01 to 1.01)
Cholesterol	CHD	Inverse variance weighted	86	1.370000e-12	1.01(1.01 to 1.02)
Cholesterol	CHD	Simple mode	86	2.199229e-02	1.01(1.00 to 1.02)
Cholesterol	CHD	Weighted mode	86	1.770000e-08	1.01(1.01 to 1.01)
Smoking	CHD	MR Egger	120	5.682784e-01	1.01(0.97 to 1.05)
Smoking	CHD	Weighted median	120	8.870000e-07	1.03(1.02 to 1.04)
Smoking	CHD	Inverse variance weighted	120	1.730000e-10	1.03(1.02 to 1.04)
Smoking	CHD	Simple mode	120	8.533727e-02	1.03(1.00 to 1.06)
Smoking	CHD	Weighted mode	120	4.353906e-02	1.03(1.00 to 1.07)

P<0.05 was considered statistically significant

Fig. 4. Summary forest plot of the results of Mendelian randomization The analysis was conducted using multiple methods, including MR Egger, weighted median, inverse variance weighted, simple mode, and weighted mode. The p-value threshold for statistical significance was set at $P < 0.05$. The forest plot at the bottom indicates the direction of the effect, with protective factors on the left and risk factors on the right.

using seven machine learning techniques and electronic health record data. Their results indicated that the Light Gradient Boosting Machine (LightGBM) model exhibited superior overall performance, showing enhanced reliability in assessing risk prediction factors for CHD [40]. In another study, Ma and Cai-Yi analyzed the risk of CHD among more than 300,000 diabetic patients in Southwest China using three traditional machine learning models. The random forest model produced an AUC of 0.701 on the test samples, providing personalized guidance for early CHD warning in the diabetic population [41]. Similarly, Xu and Hu utilized machine learning algorithms to predict CHD among diabetic patients, drawing on data from the General Hospital of the PLA. Their study employed five different machine learning algorithms, finding that the established XGBoost model demonstrated significant predictive value for elderly diabetic patients with CHD, achieving an AUC of 0.880 [42]. Compared to previous studies, this research. Unlike previous studies limited to either prediction or correlation analysis, our research advances CHD risk assessment through three key innovations: (1) systematic comparison of eight machine learning algorithms identifying SVM as optimal (AUC = 0.909, accuracy = 83.4 %) using readily obtainable lifestyle variables; (2) integration of MR to distinguish causal risks (smoking, sleep duration) from mere biomarkers (fasting glucose); and (3) development of a clinically actionable tool requiring only non-invasive measurements, bridging a critical gap between risk prediction and causal prevention strategies.

4.3. *Model features and evaluation*

In this study, we selected demographic data and 16 lifestyle-related factors from the NHANES database as the feature variables. These included age, sex, education level, race, marital status, physical activity, sleep status, smoking, alcohol consumption, BMI, poverty index, fasting blood glucose, systolic blood pressure, diastolic blood pressure, energy intake, protein intake, carbohydrate intake, total sugars, dietary fiber, TC, and total fats, totaling 21 variables. In contrast to most previous studies [40–42], which have predominantly considered biochemical indicators as key variables in CHD prediction models, our model focuses on lifestyle-related indicators. These indicators are generally easier to obtain and less invasive, with only fasting blood glucose being a minimally invasive measurement that can be self-collected at home. Therefore, the models we established operate independently of biochemical indicators. Once high-risk individuals are flagged, their modifiable lifestyle targets (sleep, physical activity, diet, alcohol, smoking) can be delivered to patients and clinicians through mobile health (mHealth) apps and telemedicine portals, tools that have already improved heart-rate variability in high-risk diabetics and are poised to expand scalable, guideline-aligned CHD prevention [43,44]. Our goal was to develop a practical, convenient, and highly acceptable screening model, which can facilitate early identification of individuals at risk of CHD based on easily accessible lifestyle-related factors.

As is shown in Fig. 4, the relative importance of various variables identified in the final selection, with age, gender, fasting glucose, diastolic blood pressure, smoking status, TC, BMI, and sleep quality being the most significant factors influencing model design. Given the inherent complexity of machine learning methods and the challenges in intuitively presenting results, we applied SHAP values for interpretability within the SVM model to achieve optimal variable representation and clarity. A positive SHAP value indicates that the associated feature contributes to a higher risk of CHD, whereas a negative SHAP value indicates a lower risk.

The interpretation of SHAP in conjunction with the causal validation through MR reveals results that align closely with earlier studies on coronary heart disease risk factors. Data from 102 prospective studies indicate that for populations without a history of diabetes, fasting glucose levels or impaired fasting glucose do not significantly improve the predictive metrics for vascular diseases when considered alongside several conventional risk factors [45]. A recent meta-analysis assessing various hypertension diagnostic guidelines found that isolated diastolic hypertension is associated with increased cardiovascular risk, and higher diastolic blood pressure thresholds correlate with elevated cardiovascular risk [46]. Although the ML model correctly reproduces the empirical observation that lower DBP is associated with higher contemporaneous probability of CHD (Fig. 3), this association is unlikely to be causal. Because the model was trained on contemporaneous measurements, the captured associations may reflect consequences of already-established disease or its treatment rather than aetiological risk factors. Supplementary Table 2 shows that participants with manifest CHD were more frequently prescribed antihypertensive medications and had lower cardiac output, both of which reduce measured DBP. Consequently, low DBP should be interpreted as a marker of advanced disease or aggressive treatment rather than a causal risk factor for CHD. We attempted prospective validation using the NHANES 2013–2018 linked mortality files; however, the follow-up duration is short and cardiovascular deaths are too few to meet the minimum event-per-variable requirement of Cox regression. Notably, prior NHANES analyses reported a sharp rise in cardiovascular mortality when DBP exceeded 75 mmHg, a direction consistent with our MR estimates and further supporting a positive causal effect of long-term elevated DBP on CHD [47]. Hypertension is widely considered as an independent risk factor for CHD; however, initial discussions often prioritize systolic blood pressure, leading to an underappreciation of diastolic pressure. Our study demonstrates that diastolic pressure plays an indispensable role in the risk associated with coronary heart disease, suggesting that healthcare professionals and the public should equally weigh diastolic pressure alongside systolic pressure when evaluating cardiovascular health. Existing data indicate that insomnia is linked with an increased risk of hypertension, coronary heart disease, and recurrent acute coronary syndrome [48]. Due to societal pressures and health-related issues, insomnia has become a pervasive sleep disorder, significantly impacting the daily lives and mental health of many individuals. Our study further validates the significant role of insomnia in the risk of CHD, indicating that the cardiovascular health risks among individuals with insomnia may be more severe than those in the general population. We emphasize the importance of sleep quality and encourage the public to develop good sleep habits, which could potentially incidence CHD Regarding the well-known cardiovascular risk factor, BMI, a prospective cohort study from the China Biobank indicates that participants who are overweight have a 41 % higher risk of developing CHD compared to those with BMI [49]. With the advancement technology and the rise a smart society, increasing of individuals are opting for modes of, resulting in a general increase in BMI. To mitigate the risk of CHD, recommend that individuals engage in exercise, their BMI, and maintain a healthy. Elevated levels of TC have long been recognized as one of the primary causes of atherosclerosis and cardiovascular diseases. We emphasize that the "cholesterol" entry in Table 1 quantifies dietary cholesterol intake, whereas the MR exposure is based on serum TC. The inverse cross-sectional association between dietary cholesterol and prevalent CHD (Table 1) is therefore not contradictory to the positive genetically-predicted effect of serum cholesterol on incident CHD (Fig. 4). The observed inverse association likely reflects reverse causation: individuals with diagnosed CHD frequently reduce consumption of high-cholesterol foods, attenuating their apparent intake. Circulating cholesterol, in contrast, captures systemic lipid metabolism that causally elevates CHD risk. This distinction underscores the limitation of interpreting dietary variables as proxies for biological exposure and reinforces the need for MR designs that leverage serum lipid biomarkers. Our MR analysis of TC implicates genetically elevated TC as a causal CHD driver; this aligns with evidence that saturated fat and added sugar—rather than dietary cholesterol itself—are the main dietary determinants of TC. Reducing animal-source foods rich in these fats and sugars (e.g., fatty red meat, full-fat dairy, sugar-sweetened items) would therefore be expected to lower serum TC and consequent CHD risk. Smoking has long been recognized as an independent risk factor for cardiovascular diseases, and our study similarly highlights its significant impact on the risk of CHD. Research indicates that the risk of cardiovascular events decreases significantly within five years after quitting smoking, and this risk can diminish to levels comparable to never smokers over several decades [50]. This finding suggests that it is never too late to take action to quit smoking, regardless of when one begins the cessation process. It is crucial for individuals to recognize the harmful effects of smoking on health; distancing oneself from cigarettes can effectively reduce the risk of CHD. By quitting smoking, individuals not only improve their own health status but also contribute to creating a healthier environment for those around them.

4.4. Limitations and future directions

Several limitations are present in our study. First, the diagnosis of CHD was defined based on self-reported data from interview questionnaires in the NHANES database, which may introduce information bias due to cognitive deficits or recall bias. Any inaccuracies in CHD classification could, to some extent, affect the accuracy of machine learning models in identifying CHD. Second, during the inclusion of overall data for machine learning model analysis, there were still some missing metrics for certain participants. Although we employed rigorous statistical methods for imputation, some residual bias may persist. Furthermore, as our study is based on the NHANES database and the GWAS gene repository, the results may only be applicable to the U.S. population and may not be directly generalizable to other global populations. Future research should focus on validating our model across diverse populations, such as those from UK and Asian biobanks, while also incorporating clinical diagnoses, including data from electronic health records, to mitigate self-report bias. Additionally, the use of advanced methodologies like federated learning, deep learning architectures, and multi-omics integration, such as proteomics, has the potential to enhance the model's robustness and biological interpretability. Moreover, implementation studies are crucial to assess the real-world applicability, cost-effectiveness, and development of clinician-friendly interfaces, which are essential for the scalable adoption of the model in preventive healthcare.

5. Conclusion

While previous researchers have developed predictive models for CHD, most of these models have incorporated laboratory metrics from hospital settings and utilized traditional statistical methods. With the continuous advancement and refinement of machine learning in recent years, its applications in healthcare have become increasingly prevalent. This study demonstrates the potential of using lifestyle-related data to predict CHD through machine learning methodologies. We utilized data from 29,400 respondents from the NHANES, incorporating 21 lifestyle variables and applying eight common machine learning models for analysis. Following model training, we evaluated and compared the models using the AUC. The performance ranking of the eight machine learning models was SVM > DT > AdaBoost > XGBoost > NN > MLP > NB > GLM, with the SVM model achieving an accuracy of 83.4 % and an AUC of 0.909. Subsequently, SHAP interpretability and Mendelian randomization validation were performed, and our cumulative results indicate that the SVM model can better predict CHD using lifestyle-related indicators. In daily life, individuals can reduce their risk of developing CHD by minimizing cholesterol intake, paying attention to changes in diastolic blood pressure, managing sleep duration appropriately, maintaining BMI within a normal range, and avoiding smoking or quitting as early as possible. To summarize, by synergizing machine learning and MR, we developed and causally validated a practical CHD prediction tool. Our findings shift prevention paradigms from biomarker-centric approaches to modifiable lifestyle factors, offering a scalable strategy to reduce the global CHD burden.

CRediT authorship contribution statement

Yang yang Cui: Writing – original draft, Methodology, Investigation, Formal analysis. **Yonghong Zhang:** Writing – review & editing, Software, Methodology. **Lang Zeng:** Writing – review & editing, Methodology, Investigation. **Shikang Li:** Validation, Methodology, Investigation. **Xue Mei:** Investigation. **Xiangmei Yang:** Writing – review & editing, Investigation. **Peng Zhou:** Investigation. **Lijuan Xiong:** Investigation. **Yijuan Huang:** Software, Resources. **Jing Luo:** Investigation. **Fenglin Wu:** Supervision, Conceptualization. **Rongchuan Yue:** Writing – review & editing, Writing – original draft, Methodology, Funding acquisition, Conceptualization.

Funding

This study was supported by the Nanchong science and technology plan project (23JCYJPT0059); Research and development program of North Sichuan Medical College (CBY23-TD01); Research and Development Plan project of Clinical School and Affiliated Hospital of North Sichuan Medical University(2024PTZK007); Scientific Research Program of Sichuan Medical Association(2024TG38) and Funding from the Scientific Startup Foundation for Doctors of North Sichuan Medical College to Xue Mei (CBY23-QDA11).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijcrp.2025.200536.

References

[1] M. de Oliveira Laterza Ribeiro, V.M. Correia, L.L. Herling de Oliveira, P.R. Soares, T.L. Scudeler, Evolving diagnostic and management advances in coronary heart disease, Life 13 (4) (2023) 951.

[2] M.M. Tang, S.T. Zhao, R.Q. Li, W. Hou, Therapeutic mechanisms of ginseng in coronary heart disease, Front. Pharmacol. 14 (2023) 1271029.

[3] N. Wang, Y. Sun, H. Zhang, et al., Long-term night shift work is associated with the risk of atrial fibrillation and coronary heart disease, Eur. Heart J. 42 (40) (2021) 4180–4188.

[4] I. Ferreira-González, The epidemiology of coronary heart disease, Rev. Esp. Cardiol. 67 (2014) 139–144.

[5] Laila Al-Shaar, et al., Red meat intake and risk of coronary heart disease among US men: prospective cohort study, BMJ Clin. Res. 371 (2020), https://doi.org/10.1136/bmj.m4141 m4141. 2 Dec.

[6] I. Kononenko, Machine learning for medical diagnosis: history, state of the art and perspective, Artif. Intell. Med. 23 (1) (2001) 89–109, https://doi.org/10.1016/s0933-3657(01)00077-x.

[7] Yifan Qin, et al., Machine learning models for data-driven prediction of diabetes by lifestyle type, Int. J. Environ. Res. Publ. Health 19 (22) (2022) 15027, https://doi.org/10.3390/ijerph192215027.

[8] Ch Anwar Ul Hassan, et al., Effectively predicting the presence of coronary heart disease using machine learning classifiers, Sensors (Basel, Switzerland) 22 (2022), https://doi.org/10.3390/s22197227, 19 7227. 23 Sep.

[9] A. Yilmaz, M.İ. Hayıroğlu, S. Salturk, et al., Machine learning approach on high risk treadmill exercise test to predict obstructive coronary artery disease by using P, QRS, and T waves' features, Curr. Probl. Cardiol. 48 (2) (2023) 101482, https://doi.org/10.1016/j.cpcardiol.2022.101482.

[10] V. Cicek, A.L. Orhan, F. Saylik, et al., Predicting short-term mortality in patients with acute pulmonary embolism with deep learning, Circ. J. 89 (5) (2025) 602–611, https://doi.org/10.1253/circj.CJ-24-0630.

[11] Liying Zhang, et al., Nonlaboratory-based risk assessment model for coronary heart disease screening: model development and validation, Int. J. Med. Inf. 162 (2022) 104746, https://doi.org/10.1016/j.ijmedinf.2022.104746, 18 Mar.

[12] Mohammadreza Naderian, et al., Development and evaluation of a comprehensive prediction model for incident coronary heart disease using genetic, social, and lifestyle-psychological factors: a prospective analysis of the UK biobank, Annals of internal medicine (2024), https://doi.org/10.7326/ANNALS-24-00716, 10.7326/ANNALS-24-00716. 10 Dec.

[13] Ryne Paulose-Ram, et al., The National health and Nutrition Examination Survey (NHANES), 2021-2022: adapting data collection in a COVID-19 environment, American journal of public health 111 (12) (2021) 2149–2156, https://doi.org/10.2105/AJPH.2021.306517.

[14] Xi Li, et al., Development of an interpretable machine learning model associated with heavy metals' exposure to identify coronary heart disease among US adults via SHAP: findings of the US NHANES from 2003 to 2018, Chemosphere 311 (2023) 137039, https://doi.org/10.1016/j.chemosphere.2022.137039. Pt 1.

[15] Donald M. Lloyd-Jones, et al., Life's essential 8: updating and enhancing the American heart association's construct of cardiovascular health: a presidential advisory from the American heart Association, Circulation 146 (5) (2022) e18–e43, https://doi.org/10.1161/CIR.0000000000001078.

[16] Sebastian Daberdaku, et al., A combined interpolation and weighted K-Nearest neighbours approach for the imputation of longitudinal ICU laboratory data, Journal of healthcare informatics research 4 (2) (2020) 174–188, https://doi.org/10.1007/s41666-020-00069-1.

[17] Md Jamal Uddin, Jitang Fan, Interpretable machine learning framework to predict the glass transition temperature of polymers, Polymers 16 (8) (2024) 1049, https://doi.org/10.3390/polym16081049.

[18] Fernando Llanos, et al., The relationship between sentence intelligibility, band importance, and signal covariance, JASA express letters 3 (5) (2023) 055204, https://doi.org/10.1121/10.0019498.

[19] Russell A. Poldrack, et al., Establishment of best practices for evidence for prediction: a review, JAMA Psychiatry 77 (5) (2020) 534–540, https://doi.org/10.1001/jamapsychiatry.2019.3671.

[20] Fahad R. Albogamy, et al., Decision support System for predicting survivability of hepatitis patients, Front. Public Health 10 (2022) 862497, https://doi.org/10.3389/fpubh.2022.862497, 14 Apr.

[21] Gilles Vandewiele, et al., Overly optimistic prediction results on imbalanced data: a case study of flaws and benefits when applying over-sampling, Artif. Intell. Med. 111 (2021) 101987, https://doi.org/10.1016/j.artmed.2020.101987.

[22] Nirajan Budhathoki, et al., Predicting asthma using imbalanced data modeling techniques: evidence from 2019 Michigan BRFSS data, PLoS One 18 (12) (7 Dec. 2023) e0295427, https://doi.org/10.1371/journal.pone.0295427.

[23] Zhaohui Zhang, et al., Identification of potential feature genes in non-alcoholic fatty liver disease using bioinformatics analysis and machine learning strategies, Comput. Biol. Med. 157 (2023) 106724, https://doi.org/10.1016/j.compbiomed.2023.106724.

[24] Chen Ding, et al., Quantum-Inspired support vector machine, IEEE Transact. Neural Networks Learn. Syst. 33 (12) (2022) 7210–7222, https://doi.org/10.1109/TNNLS.2021.3084467.

[25] Andre Esteva, et al., Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (7639) (2017) 115–118, https://doi.org/10.1038/nature21056.

[26] Parastoo Golpour, et al., Comparison of support vector machine, Naïve bayes and logistic regression for assessing the necessity for coronary angiography, Int. J. Environ. Res. Publ. Health 17 (18) (2020) 6449, https://doi.org/10.3390/ijerph17186449, 4 Sep.

[27] Qingqing Li, et al., XGBoost-based and tumor-immune characterized gene signature for the prediction of metastatic status in breast cancer, J. Transl. Med. 20 (1) (2022) 177, https://doi.org/10.1186/s12967-022-03369-9.

[28] Zheyu Zhang, et al., Multilayer perceptron-based prediction of stroke mimics in prehospital triage, Sci. Rep. 12 (1) (26 Oct. 2022) 17994, https://doi.org/10.1038/s41598-022-22919-1.

[29] Connor J. McCabe, et al., Interpreting interaction effects in generalized Linear models of nonlinear probabilities and counts, Multivariate Behav. Res. 57 (2–3) (2022) 243–263, https://doi.org/10.1080/00273171.2020.1868966.

[30] B. Hisham, A. Hamouda, Arabic sign language recognition using Ada-Boosting based on a leap motion controller, Int. j. inf. tecnol. 13 (2021) 1221–1234.

[31] Elric Zweck, et al., Machine learning identifies clinical parameters to predict mortality in patients undergoing transcatheter mitral valve repair, JACC Cardiovasc. Interv. 14 (18) (2021) 2027–2036, https://doi.org/10.1016/j.jcin.2021.06.039.

[32] Jens C. Pruessner, et al., Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change, Psychoneuroendocrinology 28 (7) (2003) 916–931, https://doi.org/10.1016/s0306-4530(02)00108-7.

[33] Susanna C. Larsson, et al., Mendelian randomization for cardiovascular diseases: principles and applications, Eur. Heart J. 44 (47) (2023) 4913–4924, https://doi.org/10.1093/eurheartj/ehad736.

[34] George Davey Smith, Gibran Hemani, Mendelian randomization: genetic anchors for causal inference in epidemiological studies, Hum. Mol. Genet. 23 (R1) (2014) R89–R98, https://doi.org/10.1093/hmg/ddu328.

[35] Fernando Pires Hartwig, et al., Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique, Int. J. Epidemiol. 45 (6) (2016) 1717–1726, https://doi.org/10.1093/ije/dyx028.

[36] Jack Bowden, et al., Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression, Int. J. Epidemiol. 47 (4) (2018) 1264–1278, https://doi.org/10.1093/ije/dyy101.

[37] Biyun Li, et al., The causal relationship between gut microbiota and lymphoma: a two-sample Mendelian randomization study, Front. Immunol. 15 (2024) 1397485, https://doi.org/10.3389/fimmu.2024.1397485, 7 May.

[38] Quentin F. Gronau, Wagenmakers Eric-Jan, Limitations of Bayesian leave-one-out cross-validation for model selection, Computational brain & behavior 2 (1) (2019) 1–11, https://doi.org/10.1007/s42113-018-0011-7.

[39] Chenxue Li, et al., Partial Youden index and its inferences, J. Biopharm. Stat. 29 (2) (2019) 385–399, https://doi.org/10.1080/10543406.2018.1535502.

[40] Zengjing Liu, et al., Predictive model and risk analysis for coronary heart disease in people living with HIV using machine learning, BMC Med. Inf. Decis. Making 24 (2024), https://doi.org/10.1186/s12911-024-02511-5, 1 110. 25 Apr.

[41] Cai-Yi Ma, et al., Predicting coronary heart disease in Chinese diabetics using machine learning, Comput. Biol. Med. 169 (2024) 107952, https://doi.org/10.1016/j.compbiomed.2024.107952.

[42] Hu Xu, et al., Establishment of a diagnostic model of coronary heart disease in elderly patients with diabetes mellitus based on machine learning algorithms, Journal of geriatric cardiology : JGC 19 (6) (2022) 445–455, https://doi.org/10.11909/j.issn.1671-5411.2022.06.006.

[43] M.İ. Hayıroğlu, Telemedicine: current concepts and future perceptions, Anatol. J. Cardiol. 22 (2019) 21–22, https://doi.org/10.14744/AnatolJCardiol.2019.12525. Suppl 2.

[44] M.İ. Hayıroğlu, G. Çinier, G. Yüksel, et al., Effect of a mobile application and smart devices on heart rate variability in diabetic patients with high cardiovascular risk: a sub-study of the LIGHT randomized clinical trial, Kardiol. Pol. 79 (11) (2021) 1239–1244, https://doi.org/10.33963/KP.a2021.0112.

[45] Emerging Risk Factors Collaboration, Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies, Lancet (London, England) 375 (9733) (2010) 2215–2222, https://doi.org/10.1016/S0140-6736(10)60484-9.

[46] Abhimanyu Agarwal, et al., Isolated diastolic hypertension and cardiovascular outcomes across different diagnostic guidelines: a systematic review and meta-analysis, The Egyptian heart journal : (EHJ) : official bulletin of the Egyptian Society of Cardiology 76 (2024), https://doi.org/10.1186/s43044-024-00556-5, 1 127. 14 Sep.

[47] T. Elfassy, C.A. German, P. Muntner, et al., Blood pressure and cardiovascular disease mortality among US adults: a sex-stratified analysis, 1999-2019, Hypertension 80 (7) (2023) 1452–1462.

[48] Sogol Javaheri, Susan Redline, Insomnia and risk of cardiovascular disease, Chest 152 (2) (2017) 435–444, https://doi.org/10.1016/j.chest.2017.01.026.

[49] J.H. Song, et al., Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi 43 (9) (2022) 1357–1363, https://doi.org/10.3760/cma.j.cn112338-20211012-00789.

[50] Nancy A. Rigotti, Mary M. McDermott, Smoking cessation and cardiovascular disease: it's never too early or too late for action, J. Am. Coll. Cardiol. 74 (4) (2019) 508–511, https://doi.org/10.1016/j.jacc.2019.06.003.

ABOUT UMT FACULTY

SDI

Selective Dissemination of Information (SDI) service is a current-awareness service offered by the PSNZ for UMT Faculty Members. The contents selection criteria include current publications (last 5 years), highly cited and most viewed/downloaded documents. The contents with pdf full text from subscribed databases are organized and compiled according to a monthly theme which is determined based on the topics of specified interest.

For more information or further assistance, kindly contact us at 09-6684846/4298 or email to psnz@umt.edu.my/hairol@umt.edu.my

Thank you.

Perpustakaan Sultanah Nur Zahirah
Universiti Malaysia Terengganu
21030 Kuala Nerus, Terengganu.

Tel.   : 09-6684185 (Main Counter)
Fax   : 09-6684179
Email : psnz@umt.edu.my