Advances in Technology Innovation, vol. 10, no. 1, 2025, pp. 58-71

Recursive Feature Elimination and Optimized Hybrid Ensemble Approach for Early Heart Disease Prediction

Jitendra P Chaudhari[1], Kishan P Patel[2,*], Hiren K Mewada[3], Hardikkumar Sudhirbhai Jayswal[4], Yogesh P Kosta[5], Kanchan S Bhagat[6], Shubhangi D Kirange[7]

[1]Charusat Space Research and Technology Center; Department of Electronics and Communication Engineering; Chandubhai S Patel Institute of Technology; Charotar University of Science and Technology, Changa, Gujarat, India

[2]Department of Electrical Engineering; Chandubhai S Patel Institute of Technology, Charotar University of Science and Technology, Changa, Gujarat, India

[3]Electrical Engineering Department, Prince Mohammad Bin Fahd University, Al Khobar, Kingdom of Saudi Arabia

[4]Department of Information Technology, Devang Patel Institute of Advance Technology and Research, Gujarat, India

[5]Provost, UKA Tarsadia University , Gujarat, India

[6]Dept of Computer Engineering, J. T. Mahajan College of Engineering, Jalgaon, Maharashtra, India

[7]Department of Information Technology, Government Polytechnic, Jalgaon, Maharashtra, India

Received 07 June 2024; received in revised form 17 October 2024; accepted 28 October 2024

DOI: https://doi.org/10.46604/aiti.2024.13825

Abstract

Early machine learning prediction improves patient health and prevents heart disease, one of the leading causes of morbidity worldwide. However, challenges such as noise and incomplete data often obscure patterns critical for accurate predictions, and single-classifier models may fail to capture data complexity. This study aims to develop a robust ensemble model leveraging advanced feature selection techniques to enhance prediction accuracy. Various machine-learning algorithms are examined. Recursive feature elimination is applied to remove irrelevant features, improving model performance. The hybrid ensemble method achieves 93.15% accuracy, 93.15% precision, and 92.97% recall, outperforming Principal Component Analysis and symmetrical uncertainty methods. This research sets a benchmark for future studies by leveraging hyperparameter tuning and advanced feature selection to optimize feature reduction and machine learning models.

Keywords: ensemble machine learning, heart disease, hyperparameter tuning, recursive feature elimination

1. Introduction

The WHO identifies heart disease as the leading cause of death, affecting 17.9 million people annually [1]. Hypertension, high cholesterol, overweight, obesity, and hyperglycemia are key risk factors for heart disease. Sleep issues, leg swelling, chronic cough, and increased heart rate are also factors, according to the American Heart Association [2]. The overlap of symptoms with other illnesses makes early diagnosis difficult for doctors. Modern healthcare has shifted toward integrating IoT and AI devices to adapt to changing medical diagnostics. This trend improves practitioners' heart disease diagnosis decisions [3]. Healthcare professionals prefer IoT and AI technologies for more accurate and timely diagnoses.

Healthcare relies on machine learning (ML) to make accurate predictions from large datasets. It helps simplify geometric analyses of extensive medical records [4-6]. Integrating IoT and AI technologies enables early heart disease detection, meeting the need for accurate diagnoses in healthcare. This research focuses on optimizing ML through feature extraction and parameter

* Corresponding author. E-mail address: kishanpatel.ee@charusat.ac.in

English language proofreader: Chih-Wen Teng

tuning. Grid search identifies optimal ML features and hyperparameters to enhance prediction accuracy. An ensemble technique further improves performance by addressing model biases. Various supervised ML classifiers are employed for predicting cardiac disease, utilizing datasets from University Hospital Zurich [7], VA Medical Centre (Long Beach, California), Hungarian Institute of Cardiology (Budapest), and Cleveland Clinical Foundation. The data from the UCI machine repository undergoes pre-processing, including removing missing values and standard scaling. Recursive feature elimination (RFE) is applied to select the six most crucial features, and diverse MLs are trained for classification.

The study progresses through three stages: traditional heart disease prediction, feature elimination using RFE, and hyperparameter tuning via grid search. Key contributions include:

(1) Developing an ensemble model using popular ML algorithms.
(2) Using RFE to identify the six most relevant features, improves the model's performance.
(3) Optimizing hyperparameters to balance model complexity and avoid overfitting.
(4) Evaluating model performance based on recall, precision, accuracy, and F-measure.

2. Related Work

ML for heart disease prediction has been extensively researched to identify early indicators. This effort is crucial as many heart disease risk factors overlap with diabetes, highlighting the importance of early and accurate detection in saving lives. Shah et al. [8] tested ML algorithms like Random Forest (RF), k-nearest neighbors (KNN), Decision Tree (DT), and Naïve Bayes (NB). Their study used 303 instances and 76 attributes from the UCI machine repository, but only 14 were used for their models. Their KNN achieved 90% training accuracy, but testing accuracy was only 78.95%, suggesting overfitting. The heart disease prediction system combines all classification methods into one algorithm [9]. The results indicate that the combined model outperforms individual methods. The lack of detailed performance analysis makes it difficult to assess the actual effectiveness of this hybrid approach.

Ramotra and Mansotra [10] presented an integrated system utilizing a graph-based technique and weighted association rule mining applied to the Andhra Pradesh population. Unfortunately, the study does not specify prediction accuracy levels. Singh and Shrivastava [11] conducted a comprehensive analysis of different heart disease prediction techniques, emphasizing the efficiency of ML in prediction analysis. The absence of comparisons using appropriate datasets limits the generalizability of their findings. Ali et al. [12] used Kaggle datasets to test various ML algorithms, with Multi-layer Perceptron (MLP) and KNN achieving the highest accuracy of 91% and 100%. Nonetheless, further analysis is required to justify the success of KNN over MLP. Salhi et al. [13] focused on data analysis of heart disease, employing correlation matrix-based feature selection and achieving 93% accuracy with neural networks (NN). The authors did not emphasize the importance of features in their approach.

The Multi-Layer Pi-Sigma Neuron Model (MLPSNM) [14] used Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and normalization for feature reduction and a standard Backpropagation (BP) algorithm for classification. PCA is versatile but cannot capture meaningful attributes in non-linear complex data. Muhammad et al. [15] optimized feature spaces using algorithms like Fast Correlation-Based Filter Solution (FCBF), minimal redundancy maximal relevance (mRMR), Relief, and Least Absolute Shrinkage and Selection Operator (LASSO), achieving an accuracy of 94.41%.

However, further exploration of the selection of optimization techniques and a comparative analysis is required. Singh and Kumar [16] identified KNN as the best predictor among support vector machine (SVM), KNN, linear regression, and DT with 87% accuracy. A summary of diverse qualities among ML algorithms for cardiovascular disease prediction is presented in Krittanawong et al. [17], highlighting the promising abilities of SVM and boosting algorithms. The authors commented that an appropriate approach to selecting an ML model is required to interpret the study in the context of clinical practices.

Yadav et al. [18] used Logistic Regression (LR), KNN, and NB, finding that fuzzy KNN produced better results. Akella and Akella [19] employed six ML algorithms, achieving an accuracy of 93% using an artificial neuron Network (ANN) for coronary artery disease detection. They employed 14 features without empirical study on feature selections. García-Ordás et al. [20] explored heart disease prediction using deep learning algorithms, and accuracy is limited to 78.3% with nNN after hyperparameter tuning and feature selection. The authors performed arrhythmia classification with heart rate variability (HRV) in Yaghouby et al. [21] using a small subset of the MIT-BIH dataset. They minimized the features using discriminant analysis techniques, and MLP was used to classify the four classes. Further validation on a large dataset is required. Asl et al. [22] found that feature selection reduced features to 5, with SVM achieving the highest arrhythmia classification accuracy. By focusing on critical elements, feature selection enhances accuracy. This study introduces a new feature selection method and ML optimization to improve classification.

Nagavelli et al. [23] used weighted NB, SVM, and Extreme Gradient Boosting (XGBoost) with synthetic minority over-sampling technique-edited nearest neighbor (SMOTE-ENN) for ischemic heart disease localization and an improved SVM for heart failure detection. They observed XGBoost as the most effective algorithm. The study concludes by suggesting feature enhancements, including dataset updates and integration with hospital databases. Cardiovascular disease causes 32% of global deaths, according to Biswas et al. [24]. They aimed for early-stage identification using ML. Chi-square, ANOVA, and mutual information (SF1, SF2, and SF3) are used to evaluate six ML models. RF is the most promising, with 94.51% accuracy, 94.87% sensitivity, 94.23% specificity, and 0.31 log loss for SF3 feature subsets. The model's performance and selected features suggest it could be used clinically to predict early heart disease at a low cost and in a short time. However, the heart disease dataset was insufficient for developing a more accurate predictive model.

Ahmad and Polat [25] emphasized early detection's importance in fighting heart disease. Their research developed an ML model using Cleveland heart disease data. The jellyfish optimization algorithm reduced the dataset's dimensionality, minimizing overfitting. The SVM classifier with the jellyfish algorithm achieved top performance, with 98.56% sensitivity, 98.37% specificity, 98.47% accuracy, and 94.48% area under the curve. However, they must be implemented into clinical practice to improve patient diagnosis. The research addresses the pressing issue of heart disease, which affects ten billion people annually, as carried out by Saikumar et al. [26].

IoT sensor data and deep learning created an intelligent heart diagnosis app. The DG ConvoNet model, trained on UC Irvine data and tested with Cleveland Clinical Foundation real-time instances, has 96% accuracy. The study used K-means for noise reduction and linear Quadratic Discriminant Analysis for feature extraction achieving 80% sensitivity, 73% specificity, 90% precision, 79% F1-score, and a 75% Receiver Operating Characteristic (ROC) curve area. However, features were chosen randomly, and the paper does not evaluate their impact. Table 1 compares these articles, listing strengths and weaknesses.

Table 1 Overview of studied articles

Ref.	Models	Remarks
Shah et al. [8]	NB, DT, KNN, and RF were tested on the UCI machine repository.	(1) Hyperparameters and features are not optimized. (2) Low testing set accuracy, i.e., 78.95%.
Tarawneh and Embarak [9]	A hybrid approach of NB, KNN, Genetic Algorithm (GA), SVM, and NN was used for classification.	(1) NB and SVM outperform others, so hybridizing other models is inappropriate. (2) Max accuracy is 89.2%.
Ramotra and Mansotra [10]	K-means clustering, PCA, and LR to recognize heart disease.	(1) They separated the datasets into healthy and abnormal using clustering. (2) PCA-based feature reduction regardless of significance.
Ali et al. [12]	MLP, KNN, RF, DT, LR, and AdaboostM1 (AB M1) were applied to Kaggle datasets. 10-fold cross-validation was used in training.	The most predictive features were ranked by importance scores. MLP and KNN failed to generate scores, and without feature ranking, they obtained 91% and 100% accuracy, respectively.

Table 1 Overview of studied articles (continued)

Ref.	Models	Remarks
Salhi et al. [13]	Three ML algorithms, KNN, SVM, and NN, are used on datasets of different sizes (i.e., 600, 800, 1,000, and 1,200). A structured dataset of Algerian hospitals is used.	(1) For 1,200 data records, the maximum accuracy for NN was 93%, SVM was 90%, and KNN was 85.5%. (2) No evaluation was presented for training and testing sets. (3) Precision and F1-scores are not analyzed.
Burse et al. [14]	MLPSNM is proposed for the UCI ML repository with 10-fold cross-validation.	A three-layer network is used and, hence, less complex.
Muhammad et al. [15]	KNN, RF, DT, LR, and ANN were presented. The four distant feature techniques, LASSO, FCBF, Relief, and mRMR, were tested.	(1) The success rate is 85% for ANN and 85.55% for KNN. (2) Relief features were found to be better in comparison to other techniques.
Jabbar et al. [27]	Weighted association rule mining and a graph-based methodology were used.	A subjective rule-based association using age, gender, and BP. Prediction accuracy is not presented.
Proposed approach	An ensemble approach of eight ML is validated using a heart disease prediction dataset from the UCI machine repository.	RFE employs a feature reduction technique. The model succeeded with 93.15% accuracy, with six features out of 14.

This study identified several limitations. First, the models were trained on small, specific datasets, limiting their broader applicability across diverse populations. Second, while effective, the hybrid ensemble approach is complex and resource-intensive, making it unsuitable for real-time or resource-limited settings. Third, the model's performance heavily relies on feature selection, which may not always identify the most relevant features. Lastly, the high accuracy raises concerns about overfitting due to the small dataset and extensive tuning. Future work should focus on larger datasets, simplifying models, and improving feature selection methods.

3. Dataset Description

Medical data collection is challenging due to privacy and security concerns. Common heart disease prediction benchmarks include publicly available datasets like the UCI repository. This study uses the UCI heart disease dataset [7], which has 304 records and 14 features and was reduced to 297 records after removing missing values. The dataset aims to classify heart disease as positive or negative, posing a binary classification challenge. The following sections focus on developing and optimizing ML models for accurate heart disease prediction. Table 2 describes the 14 attributes.

3.1. Data pre-processing

Accurate data pre-processing is essential, as unprocessed data can weaken ML models. In this study, missing records were removed, and features were standardized for comparability and improved performance. A detailed dataset analysis is conducted, with the features described in Table 2. Standardization eliminates the mean and scaled-to-unit variance and aligns features with a standard normal distribution, enhancing the performance of many ML algorithms. This refined dataset will be used to develop and optimize heart disease prediction models.

Table 2 Feature description

#	Feature	Feature description	Non-null count	Data type
0	age	Age of the patient.	304	int64
1	sex	Gender of the patient.	304	int64
2	cp	Categorizes chest pain into four types: 1 for typical angina, 2 for atypical angina, 3 for non-anginal pain, and 4 for asymptomatic.	304	int64
3	trestbps	Denotes blood pressure at rest (mm Hg) at hospital admission.	304	int64

Table 2 Feature description (continued)

#	Feature	Feature description	Non-null count	Data type
4	chol	Defines the serum cholesterol level in mg/dl.	304	int64
5	fbs	Indicates fasting blood sugar level, with 1 for true and 0 for false (if greater than 120 mg/dl).	304	int64
6	restecg	Describes restecg, with 0 for normal, 1 for abnormal ST-T wave, and 2 for left ventricular hypertrophy meeting Estes criteria.	304	int64
7	thalach	Represents the highest heart rate possible in beats per minute (bpm).	304	int64
8	exang	Indicates exercise-induced angina (exang), with 1 for present and 0 for absent.	304	int64
9	oldpeak	Explains the ST depression brought on by exercise in comparison to rest	304	float64
10	slope	Emphasizes the steepest portion of the exercise ST segment, with 1 for upslope, 2 for flat, and 3 for downslope.	304	int64
11	ca	Describes the count of major vessels (0–3) in fluorescence.	304	int64
12	thal	Represents the Thalassemia category, with 3 for normal, 6 for a fixed defect, and 7 for a reversible defect.	304	int64
13	target	Classification, i.e., 0 for no presence of heart disease and 1 for presence.	304	int64

3.2. *Exploratory data analysis*

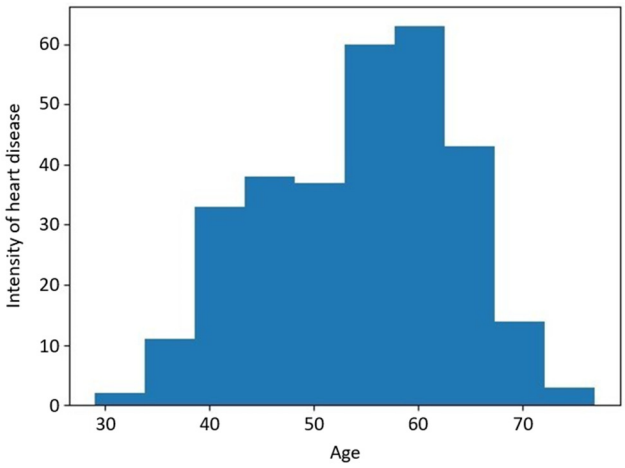

Fig. 1 Age distribution of patients

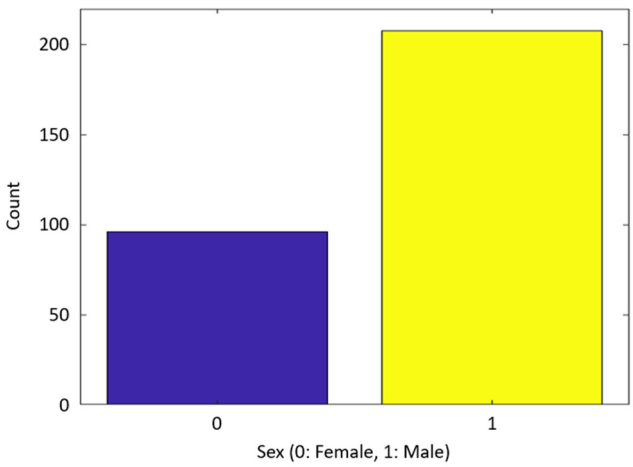

Fig. 2 Sex vs number of records

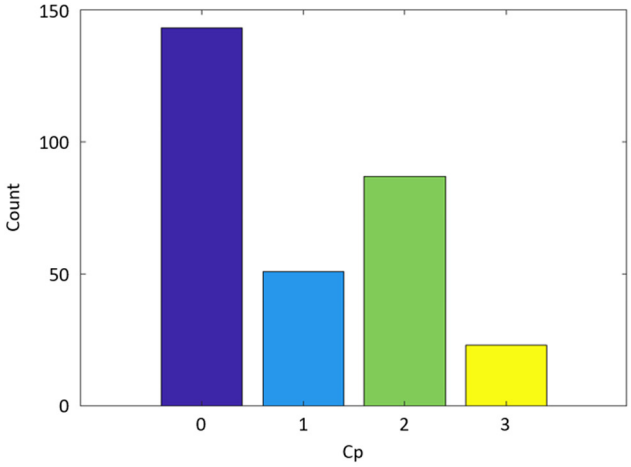

Fig. 3 Exploratory data analysis of chest pain type

To better understand the dataset, exploratory data analysis is conducted, categorizing features into quantitative and categorical groups. Quantitative features take numerical values, representing various measurements. In the heart disease dataset, features such as age, cholesterol, thalach, and ST depression induced by exercise (oldpeak) were identified as quantitative. Categorical features are label values that categorize individuals into groups. Examples from the dataset used here include thalassemia (thal), gender, fbs, cp, exang, slope, ca, and restecg. These categorical features serve as target characteristics for analysis. During feature analysis, key features for heart disease prediction were identified. Fig. 1 shows the age distribution of heart disease patients and healthy individuals. Fig. 2 depicts gender distribution, while Fig. 3 shows chest pain type distribution. Correlation matrices and heatmaps offer insights into variable relationships. Fig. 4 highlights the correlation between attributes in the heart disease dataset.

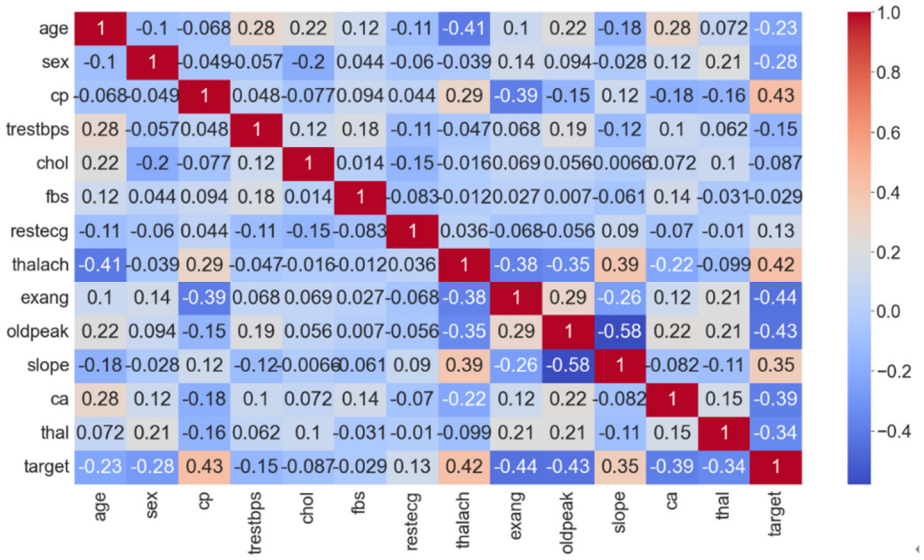

Fig. 4 Correlation between different attributes

4. Methodology

This study uses hyperparameter tuning and feature selection to create a supervised ML algorithm for heart disease detection. To ensure reliable evaluation, 70% of the dataset was used for training and 30% for testing and assessment. Before model development, the following preprocessing steps were applied to the dataset using a standard scaler to ensure consistency in feature scales. The experiments are carried out on a PC with an 11th-generation Intel(R) Core(TM) i5-1135G7 @ 2.40 GHz and 16 GB RAM.

(1) Model development and training: A diverse set of ML algorithms was applied to the training dataset. Hyperparameter tuning was performed to optimize model architecture and improve predictive performance. An ensemble approach combined multiple models to enhance the system's accuracy in predicting heart disease.

(2) Hybrid ensemble classification: A hybrid ensemble classification approach integrated eight ML algorithms, leveraging their unique strengths to create a more robust heart disease prediction model.

(3) Hyperparameter tuning: Hyperparameters were fine-tuned using techniques like grid search to optimize model performance for heart disease prediction.

(4) Classification and evaluation: The final models were evaluated on the test dataset, assessing their generalization ability to new data. Performance metrics such as precision, recall, accuracy, and F-measure were used for performance evaluation.

(5) Heart disease identification strategy: The strategy, visualized in Fig. 5, involves data preprocessing, training multiple ML models with hyperparameter tuning, and evaluating their performance on a test dataset. It combines traditional algorithms with the hybrid ensemble model for reliable heart disease prediction.

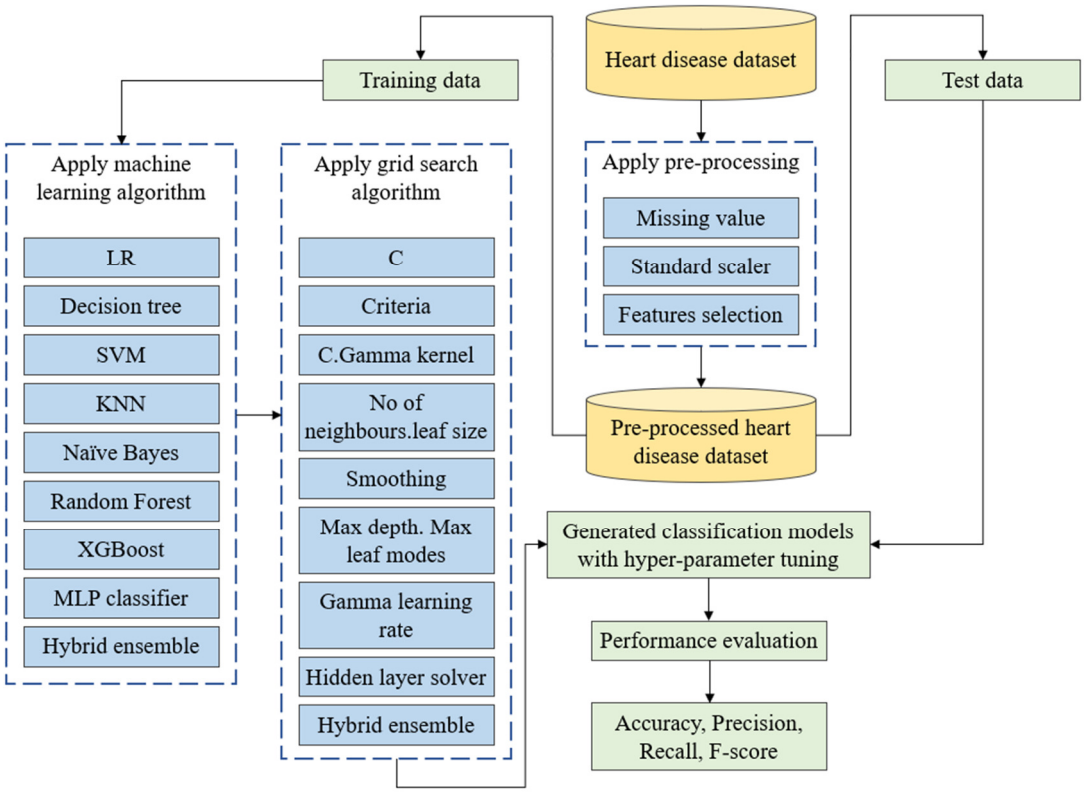

Fig. 5 Methodology for heart disease prediction system

4.1. Feature selection using recursive feature elimination (RFE)

The ML's ability to identify influential parameters is vital. Feature selection enhances algorithm performance, reducing execution time, improving accuracy, and mitigating overfitting. This study used RFE to select features from the heart disease dataset. RFE iteratively removes attributes while evaluating accuracy to determine the most essential features for prediction. It also uses cross-validation to identify the optimal number of features. Table 3 lists the six most relevant features selected by RFE for heart disease prediction. Streamlining the feature set improves model efficiency and interpretability while maintaining or enhancing accuracy. The following sections discuss how these features impact model performance and prediction accuracy.

Table 3 Features selection using RFE

Feature	Feature ranking using RFE	Support
age	7	FALSE
sex	1	TRUE
cp	1	TRUE
trestbps	6	FALSE
chol	8	FALSE
fbs	4	FALSE
restecg	3	FALSE
thalch	5	FALSE
exang	1	TRUE
oldpeak	2	FALSE
slope	1	TRUE
ca	1	TRUE
thal	1	TRUE

4.2. ML algorithms

As listed below, this research adopted eight ML models for analysis and finally proposed an ensemble model.

(1) LR is a supervised classifier where a regression model can be used as a classifier using a decision threshold. It employs the sigmoid function to model the data, and the appropriate threshold selection can lead to high precision and recall. The

sigmoid function is a monotonic continuous function that ranges between 0 and 1. Mathematically, this classification can be expressed as

$$P(y_q) = \frac{1}{1+\exp\left(-(1, x_q)\right)\beta} \tag{1}$$

where $P(y_q)$ gives the probability of y_q to be 1, x_q is the input vector to be classified, and β is the vector parameter. Now, the classification problem is equivalent to finding vector parameters.

(2) KNN is one of the simplest algorithms that stores all the classes and classifies the new ones based on the nearest neighbors calculating distance function. It assumes that similar entities reside close to one another in identical classes. The letter "*k*" represents the closest neighbors used to categorize an instance. In this algorithm, two quantities are necessary, i.e., the distance between two entities and the neighbors' quantity (k). Typically, the Euclidean distance given in the following equation is used.

$$D(X, Y) = \sqrt{\sum_{X=0}^{N} \left(X_K - Y_K\right)^2} \tag{2}$$

As shown in Fig. 6, KNN calculates an entity's distance from neighboring points and classifies it based on the nearest neighbors. Clean, normalized data is essential for KNN to prevent bias from outliers and high-value entities. During training, KNN stores the data, and in testing, it compares the test instance to the stored data, identifying the nearest neighbors to predict the majority label. The choice of 'K' and the distance metric significantly impacts the performance of the KNN method.

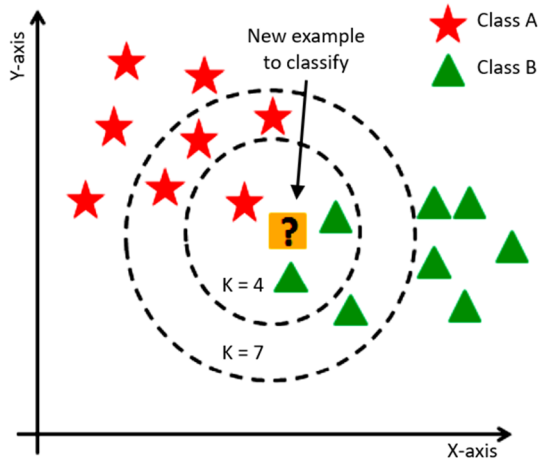

Fig. 6 Classifier example with K = 4 and 7 neighbors

(3) DT is a tree-structured supervised learning model for classification and regression tasks. Internal nodes represent the dataset's features, branches represent decision paths, and leaf nodes provide the outcomes. Decision nodes have multiple branches (as shown in Fig. 7), while leaf nodes indicate the final decisions without further branching. Each decision or test is based on the characteristics of the dataset.

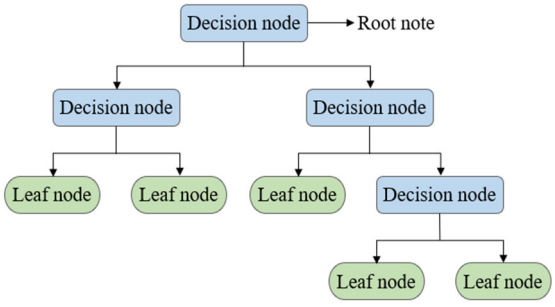

Fig. 7 Decision tree structure

(4) RF builds multiple numbers of the individual DT in the training stage. Each tree predicts the results, and the class with the most predictable results is considered the model output. Fig. 8 shows the structure of RF using trees.

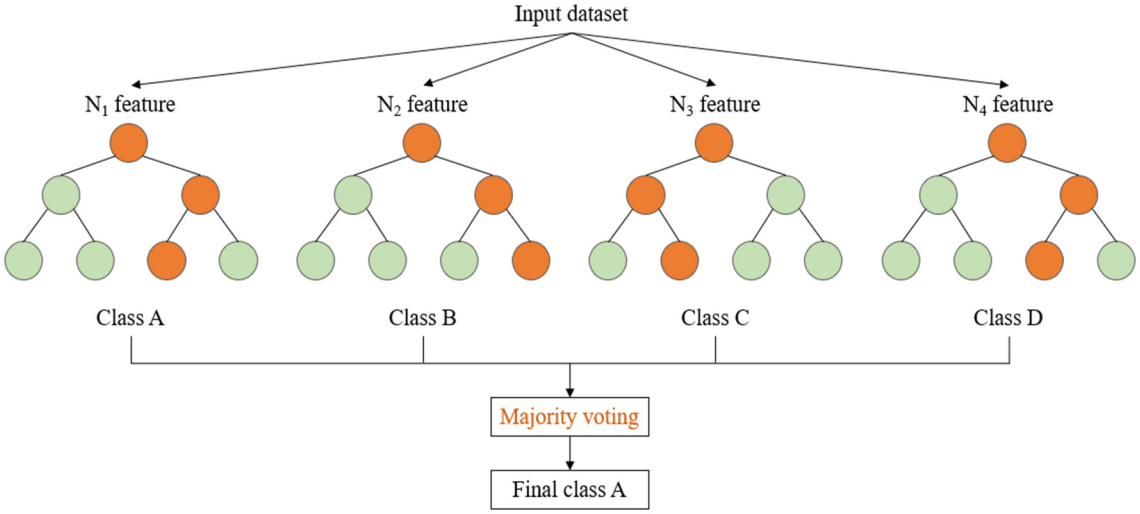

Fig. 8 Random forest using multiple trees

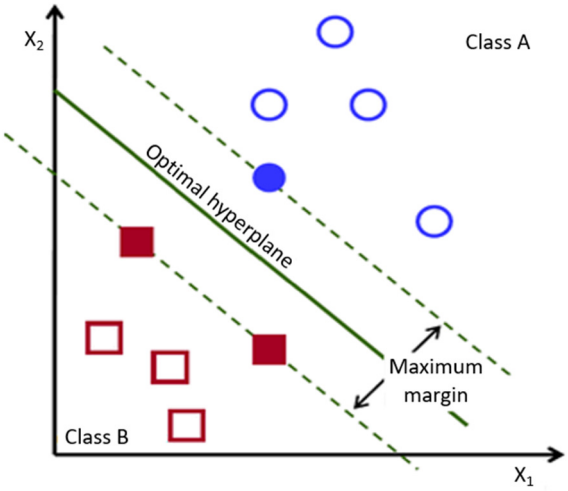

Fig. 9 Hyperplane-based classification in SVM

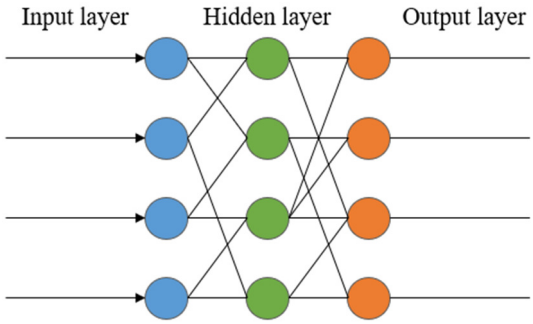

Fig. 10 MLP layer representation

(5) The SVM classifier employs a hyperplane for data classification. It looks for the best hyperplane in the area with the most significant distance from the data points. Fig. 9 illustrates the model's attempt to fit the hyperplane with the most significant possible margin. The classification accuracy may be impacted by data points located closer to the plane.

(6) NB is a probabilistic approach based on Bayes' theorem, assuming all features are independent. It combines prior knowledge about classes with new evidence using training data. First, it calculates a probability table for each data point and then determines the posterior probability for each class. The predicted class is the one with the highest posterior probability. In this study, an ensemble approach incorporates both the traditional Gaussian NB method and its optimized variant.

(7) XGBoost uses an ensemble of K classification and regression trees, enhancing learning by combining the judgments of weak classifiers. It reduces the computational cost and time of gradient boosting, making it a powerful tool for achieving state-of-the-art results in various fields.

(8) Unlike XGBoost, MLP is a feed-forward NN with input, hidden, and output layers. Neurons are trained using BP, allowing MLPs to solve non-linearly separable problems by modeling continuous functions. This study randomized input vectors during training to achieve global learning. Fig. 10 illustrates the MLP layer structure.

4.3. Optimization of Hyperparameter using a grid search

Cross-validation and grid search were used to optimize classification parameters during hyperparameter tuning. Grid search systematically divides the hyperparameter domain into a grid, generating models for each parameter combination and using cross-validation to evaluate performance. This method thoroughly explores a selected portion of the algorithm's hyperparameter space, providing consistent parameter values for dataset analysis.

Grid search systematically generates each candidate's parameter setting based on the parameters to be optimized. For instance, setting a sigma range in SVM to 5 will only allow five possible values. Therefore, the grid search approach provides $5 \times 5 = 25$ permutations of parameter settings for a random classifier with two parameters and a sigma with five possible values. Then, it evaluates the parameter setting of each candidate. Find the optimal parameters among all. A grid search with 10-fold cross-validation was conducted to find optimal hyperparameters for the heart disease dataset. The best hyperparameters obtained are shown in Table 4. Models were generated using feature selection and hyperparameter tuning before evaluating performance on the test set.

Table 4 Hyperparameters tunning and its optimum value

Classifier	Initial hyperparameters	Optimum value of hyperparameters
LR	'C': [1, 5, 10, 15, 20, 25]	'C' = 1
DT	'criterion': ['gini', 'entropy']	'criterion' = 'gini'
SVM	'C': [0.01, 0.1, 0.25, 0.5, 0.75, 1, 10, 100] 'gamma': [1, 0.75, 0.5, 0.25, 0.1, 0.01, 0.001] 'kernel': ['rbf', 'poly', 'linear']	'C' = 0.1 'gamma' = 1 'kernel' = 'linear'
KNN	'n_neighbors': list(range(1, 56)) 'leaf_size': list(range(1, 50))	'n_neighbors' = 12 'leaf_size' = 1
NB	Smoothing: default 1e-9	Smoothing = 1e-9
RF	'n_estimators': [1, 5, 10, 15, 20, 25, 30], 'max_depth': [3, 4, 5], 'max_leaf_nodes': [10, 15, 20], 'min_samples_leaf': [10, 15, 20, 25]	'n_estimators': [25] 'max_depth': [4] 'max_leaf_nodes': [20] 'min_samples_leaf': [20]
XGBoost	'learning_rate': [0.05, 0.10, 0.15, 0.20] 'max_depth': [3, 4, 5] 'gamma': [0.0, 0.1, 0.2, 0.3]	'learning_rate': [0.10] 'max_depth': [5] 'gamma': [0.3]
MLP	'C': [0.01, 0.1, 0.25, 0.5, 0.75, 1, 10, 100] 'gamma': [1, 0.75, 0.5, 0.25, 0.1, 0.01, 0.001] 'kernel': ['rbf', 'poly', 'linear']	'C' = [0.1] 'gamma': [0.1] 'kernel': ['linear']

4.4. Experimental evaluation and performance analysis

The performance of the suggested system was assessed using 30% of the data for testing and 70% for training. The recall, precision, accuracy, and F1-score of the performance are evaluated using the confusion matrix. Accuracy defines the ratio of correctly identified labels to the total number of records. Precision is computed by taking the ratio of correctly identified heart disease labels to the total predicted heart disease labels. The recall is computed by the ratio of truly identified heart disease labels to all labels in the dataset with heart disease. The F1-score defines the weighted average of precision and recall. Accuracy, precision, recall, and F-measure are defined mathematically as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

True Positive (TP) represents the number of instances correctly predicted as positive (heart disease). False Positive (FP) refers to the number of instances incorrectly predicted as positive (heart disease) when they were negative. True Negative (TN) is the number of instances correctly predicted as negative (no heart disease), and False Negative (FN) refers to the number of instances incorrectly predicted as negative (no heart disease) when they were positive.

Table 5 Precision score ML algorithms

Precision	HT without feature selection	HT with feature selection	Normal without feature selection	Normal with feature selection
LR	0.792731	0.821367	0.792731	0.821367
DT	0.803762	0.793696	0.803762	0.793696
SVM	0.871708	0.902577	0.851565	0.891429
KNN	0.883295	0.903718	0.862873	0.89472
NB	0.865055	0.885839	0.865055	0.885839
RF	0.804549	0.872304	0.83811	0.861491
XGBoost	0.821914	0.813455	0.802085	0.842376
MLP	0.836731	0.862032	0.872178	0.921562
Hybrid ensemble	0.891429	0.93414	0.862973	0.911993

Table 6 Recall score ML algorithms

Recall	HT without feature selection	HT with feature selection	Normal without feature selection	Normal with feature selection
LR	0.792079	0.821782	0.792079	0.821782
DT	0.78297	0.792871	0.782987	0.792871
SVM	0.872178	0.901881	0.852376	0.89198
KNN	0.88198	0.901782	0.862178	0.891881
NB	0.862178	0.88198	0.862178	0.88198
RF	0.802772	0.872079	0.832475	0.862178
XGBoost	0.812673	0.812673	0.792871	0.842376
MLP	0.832574	0.862277	0.872178	0.921683
Hybrid ensemble	0.89198	0.931584	0.862277	0.911782

Table 7 F1-score ML algorithms

F1-score	HT without feature selection	HT with feature selection	Normal without feature selection	Normal with feature selection
LR	0.79237	0.821246	0.792327	0.821246
DT	0.78332	0.793214	0.78332	0.793214
SVM	0.870684	0.900167	0.850746	0.891385
KNN	0.879686	0.89717	0.859655	0.889194
NB	0.858429	0.878572	0.858429	0.878572
RF	0.803377	0.870185	0.833452	0.861523
XGBoost	0.813778	0.812992	0.794058	0.842376
MLP	0.83349	0.860202	0.872178	0.920871
Hybrid ensemble	0.891385	0.929749	0.862549	0.91056

Table 8 Accuracy analysis of various ML algorithms

Accuracy	HT without feature selection	HT with feature selection	Normal without feature selection	Normal with feature selection
LR	0.792079	0.821782	0.792079	0.821782
DT	0.78297	0.792871	0.78297	0.792871
SVM	0.872178	0.901881	0.852376	0.89198
KNN	0.88198	0.901782	0.862178	0.891881
NB	0.862178	0.88198	0.862178	0.88198
RF	0.802772	0.872079	0.832475	0.862178
XGBoost	0.812673	0.812673	0.792871	0.842376
MLP	0.822574	0.852277	0.862178	0.911683
Hybrid ensemble	0.89198	0.931584	0.862277	0.911782

The literature review [8-9, 13, 15] indicates that various ML models, including KNN, SVM, DT, NB, AdaBoost, and gradient boost, have been tested. These algorithms use different feature elimination and selection methods. In Salhi et al. [13], the authors ranked features using the Pearson correlation method. However, out of 14, they eliminated only one feature in the experiment. The major weakness of the Pearson correlation is that it considers all features independent and, therefore, fails to eliminate redundant features. Muhammad et al. [15] used a fast correlation filter (FCF) to choose the top features from the list of 14. FCF calculates symmetrical uncertainty to find the high correlation, and six features were selected based on the ranking results. Among all ML models, RF succeeded with 88.48% accuracy. However, their recall rate is limited to 85.57%.

Alfadli and Almagrabi [28] used chi-squared distance for feature ranking and selected seven highly ranked features. They trained multiple ML models using different hyperparameter configurations. Their accuracy is limited to 83.15%. Sarra et al. [29] tested different ML models. They also used chi-squared distance to find the important features from the 14 features. Their study showed that SVM was the most accurate among KNN and ANN, with 89.47% accuracy using six feature sets. Thus, chi-square (X2-distance) performed well in ranking the features, but its weakness is that it works well for small-size feature sets. Reduced processing load and improved system performance were achieved by removing superfluous and redundant characteristics and selecting the essential ones [30].

Further improving system performance, a unique feature weight for each class was computed using the conditional probability technique. The prediction of heart disease was also trained into a deep-learning ensemble model. The authors reduced the feature sets to 14 out of 23 and received 83.5% accuracy. Therefore, the RFE method is used in the proposed model. Six relevant features were obtained and used to train various ML models. To improve the performance, the hyper-parameters of all models are optimized using the greedy search algorithm. Another can handle the weakness of one model, and therefore, their ensemble approach succeeded in performing better than other models. Table 9 shows the comparative studies of different models and proposed methods.

Table 9 Comparison of models utilizing feature reduction method and ML models

Ref.	year	Model	Feature selection method	Reduced features number	Accuracy	Precision	Recall
Tarawneh and Embarak [9]	2019	SVM, KNN, ANN, NB, DT	Symmetrical uncertainty and one R selection	12	83.75	81.15	82.35
Burse et al. [14]	2019	MLPSN	-	14	90.44	-	-
		SVM	PCA	4	88.32	-	-
Shah et al. [8]	2020	KNN, NB, DT	-	14	90.78	-	-
Muhammad et al. [15]	2020	DT, KNN, ETC, ANN, LR, RF, NB, SVM, GB, AB	Correlation-based filter	6	88.48	90.87	85.57
Ali et al. [30]	2020	Ensemble deep learning	Information gain	14	83.5	84.5	82.5
Salhi et al. [13]	2021	KNN, SVM, ANN	Pearson correlation method	13	93	92	94
Sarra et al. [29]	2022	SVM	X2 feature selection	6	89.47	89.40	89.40
Alfadli and Almagrabi [28]	2023	Ensemble approach	X2 feature selection	7	83.15	83.97	86.00
Proposed	2024	Ensemble approach	Recursive elimination	6	93.15	93.15	92.97

5. Conclusions and Future Work

This study significantly advances heart disease prediction by implementing a robust hybrid ensemble method. The implementation of hyperparameter tuning and feature selection significantly enhanced prediction accuracy. Several ML techniques were evaluated, including NB, XGBoost, KNN, RF, DT, LR, SVM, and MLP. Despite reducing features to six using RFE, the hybrid model achieved strong results: 93.15% accuracy, 93.15% precision, and 92.97% recall. DT using RFE

had a minimum accuracy of 79.15%, with the ensemble approach offering a 17.52% improvement. This success stems from hyperparameter tuning and feature selection optimizing model performance, demonstrating that focusing on key features enhances early heart disease detection and management.

Future efforts will explore alternative optimization methods for feature selection and hyperparameter tuning. Integrating CNNs, RNNs, wearable device data, IoT for real-time predictions, GA, and more diverse populations will improve accuracy. A user-friendly app for heart health monitoring and real-time risk assessment is also planned. These extensions enhance the model's practicality and impact on cardiovascular health management.

Conflicts of Interest

The authors declare no conflict of interest.

Statement of Ethical Approval

(a) Statement of human rights

For this type of study, statement of human rights is not required.

(b) Statement on the welfare of animals

For this type of study, statement on the welfare of animals is not required.

Statement of Informed Consent

For this type of study, informed consent is not required.

References

[1] WHO, Cardiovascular Diseases, https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1, accessed on 2021.
[2] American Heart Association, Classes of Heart Failure, https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure/classes-of-heart-failure, accessed on 2021.
[3] American Heart Association, Heart Failure, https://www.heart.org/en/health-topics/heart-failure, accessed on 2021.
[4] S. Shalev-Shwartz and S Ben-David, Understanding Machine Learning: From Theory to Algorithms, New York: Cambridge University Press, 2014.
[5] T. Hastie, R. Tibshirani, and J. H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., New York: Springer, 2009.
[6] S. Marsland, Machine Learning: An Algorithmic Perspective, 2nd ed., Boca Raton, FL: CRC Press, 2015.
[7] UCI, Heart Disease, https://archive.ics.uci.edu/dataset/45/heart+disease, accessed on 2021
[8] D. Shah, S. Patel, and S. K. Bharti, "Heart Disease Prediction Using Machine Learning Techniques," SN Computer Science, vol. 1, no. 6, article no. 345, 2020.
[9] M. Tarawneh and O. Embarak, "Hybrid Approach for Heart Disease Prediction Using Data Mining Techniques," Advances in Internet, Data and Web Technologies: The 7th International Conference on Emerging Internet, Data and Web Technologies, pp. 447-454, 2019.
[10] A. K. Ramotra and V. Mansotra, "A Hybrid Cluster and PCA-Based Framework for Heart Disease Prediction Using Logistic Regression," Rising Threats in Expert Applications and Solutions: Proceedings of FICR-TEAS 2020, pp. 111-117, 2021.
[11] S. Singh and V. Shrivastava, "The Analysis of Machine Learning Techniques for Heart Disease Prediction," Proceedings of International Conference on Communication and Computational Technologies: ICCCT-2019, pp. 137-143, 2020.
[12] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. Quinn, and M. A. Moni, "Heart Disease Prediction Using Supervised Machine Learning Algorithms: Performance Analysis and Comparison," Computers in Biology and Medicine, vol. 136, article no. 104672, 2021.

[13] D. E. Salhi, A. Tari, and M. T. Kechadi, "Using Machine Learning for Heart Disease Prediction," Advances in Computing Systems and Applications: Proceedings of the 4th Conference on Computing Systems and Applications, pp. 70-81, 2021.
[14] K. Burse, V. P. S. Kirar, A. Burse, and R. Burse, "Various Preprocessing Methods for Neural Network Based Heart Disease Prediction," Smart innovations in communication and Computational Sciences: Proceedings of ICSICCS-2018, pp. 55-65, 2019.
[15] Y. Muhammad, M. Tahir, M. Hayat, and K. T. Chong, "Early and Accurate Detection and Diagnosis of Heart Disease Using Intelligent Computational Model," Scientific Reports, vol. 10, article no. 19747, 2020.
[16] A. Singh and R. Kumar, "Heart Disease Prediction Using Machine Learning Algorithms," International Conference on Electrical and Electronics Engineering, pp. 452-457, 2020.
[17] C. Krittanawong, H. U. H. Virk, S. Bangalore, Z. Wang, K.W. Johnson, R. Pinotti, et al., "Machine Learning Prediction in Cardiovascular Diseases: A Meta-Analysis," Scientific Reports, vol. 10, article no. 16057, 2020.
[18] S. S. Yadav, S. M. Jadhav, S. Nagrale, and N. Patil, "Application of Machine Learning for the Detection of Heart Disease," 2nd International Conference on Innovative Mechanisms for Industry Applications, pp. 165-172, 2020.
[19] A. Akella and S. Akella, "Machine Learning Algorithms for Predicting Coronary Artery Disease: Efforts Toward an Open Source Solution," Future Science OA, vol. 7, no. 6, article no. FSO698, 2021.
[20] M. T. García-Ordás, M. Bayón-Gutiérrez, C. Benavides, J. Aveleira-Mata, and J. A. Benítez-Andrades, "Heart Disease Risk Prediction Using Deep Learning Techniques With Feature Augmentation," Multimedia Tools and Applications, vol. 82, no. 20, pp. 31759-31773, 2023.
[21] F. Yaghouby, A. Ayatollahi, and R. Soleimani, "Classification of Cardiac Abnormalities Using Reduced Features of Heart Rate Variability Signal," World Applied Sciences Journal, vol. 6, no. 11, pp. 1547-1554, 2009.
[22] B. M. Asl, S. K. Setarehdan, and M. Mohebbi, "Support Vector Machine-Based Arrhythmia Classification Using Reduced Features of Heart Rate Variability Signal," Artificial Intelligence in Medicine, vol. 44, no. 1, pp. 51-64, 2008.
[23] U. Nagavelli, D. Samanta, and P. Chakraborty, "Machine Learning Technology-Based Heart Disease Detection Models," Journal of Healthcare Engineering, vol. 2022, article no. 7351061, 2022.
[24] N. Biswas, M. M. Ali, M.A. Rahaman, M. Islam, M. R. Mia, S. Azam, et al., "Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques," BioMed Research International, vol. 2023, article no. 6864343, 2023.
[25] A. A. Ahmad and H. Polat, "Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm," Diagnostics, vol. 13, no. 14, article no. 2392, 2023.
[26] K. Saikumar, V. Rajesh, G. Srivastava, and J. C. W. Lin, "Heart Disease Detection Based on Internet of Things Data Using Linear Quadratic Discriminant Analysis and a Deep Graph Convolutional Neural Network," Frontiers in Computational Neuroscience, vol. 16, article no. 964686, 2022.
[27] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Graph Based Approach for Heart Disease Prediction," Proceedings of the Third International Conference on Trends in Information, Telecommunication and Computing, pp. 465-474, 2013.
[28] K. M. Alfadli and A. O. Almagrabi, "Feature-Limited Prediction on the UCI Heart Disease Dataset," Computers, Materials & Continua, vol. 74, no. 3, pp. 5871-5883, 2023.
[29] R. R. Sarra, A. M. Dinar, M. A. Mohammed, and K. H. Abdulkareem, "Enhanced Heart Disease Prediction Based on Machine Learning and X2 Statistical Optimal Feature Selection Model," Designs, vol. 6, no. 5, article no. 87, 2022.
[30] F. Ali, S. El-Sappagh, S. R. Islam, D. Kwak, A. Ali, M. Imran, et al., "A Smart Healthcare Monitoring System for Heart Disease Prediction Based on Ensemble Deep Learning and Feature Fusion," Information Fusion, vol. 63, pp. 208-222, 2020.