



# International Journal of Research in Academic World



Received: 19/January/2026

IJRAW: 2026; 5(3):56-59

Accepted: 27/February/2026

## Enhanced Heart Disease Prediction Using Hybrid Feature Selection and Ensemble Learning Techniques

<sup>1</sup>S Saranya and <sup>2</sup>R Gomathi Jayam<sup>1</sup>Student of II Year M.Sc., Department of Computer Science, Bon Secours College for Women (Autonomous), Thanjavur, Tamil Nadu, India.<sup>2</sup>Head & Assistant Professor, Department of Computer Science, Bon Secours College for Women (Autonomous), Thanjavur, Tamil Nadu, India.

### Abstract

Feature selection techniques are applied to identify the most relevant attributes influencing heart disease prediction. Ensemble methods like Random Forest and Gradient Boosting enhance model performance by combining multiple predictors. The approach ensures improved accuracy, reducing false positives and negatives. Insights derived aid in better understanding risk factors and optimizing predictive models. This study proposes a machine learning model that leverages various preprocessing steps, hyper parameter Hybrid Algorithm, and ensemble learning Techniques to predict heart disease. Enhancing heart disease prediction through ensemble learning techniques with hyper parameter optimization is a sound approach to improve the accuracy and robustness of predictive models. Ensemble learning involves combining multiple base models to create a more robust and accurate model. The collect the heart diseases data and processing data using Different type algorithms logistic regression, naive Bayes, support vector machine, k nearest neighbours, decision tree, random forest, XG boost, neural network. Logistic regression is a machine learning classification algorithm that is used to predict the probability of certain classes based on some dependent variables. Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. Features may include age, gender, blood pressure, cholesterol levels and other relevant health indicators. Preprocess the data by handling missing values, normalizing features and encoding categorical variables. When working on a heart disease prediction model, its crucial to split your dataset into training, validation and test sets. This helps you train the model on one subset, tune hyper parameters on another and assess the models generalization on the final subset. The typical split ratio for training, validation and test sets, respectively. Use a function or library to split your dataset into three subsets training, validation and test sets. After testing is performed deploy dataset model and classifier in heart disease dataset and check the heart disease result is normal or heart disease perdition finally the result is display on graphical output. These findings demonstrated the potential of our model for accurately predicting the presence or absence of heart disease. Such accurate predictions could significantly aid in early prevention, detection, and treatment, ultimately reducing the mortality and morbidity associated with heart disease.

**Keywords:** Heart Disease Prediction, Machine Learning, Feature Selection, Ensemble Learning, Hyperparameter Optimization, Logistic Regression, Random Forest, XGBoost, Support Vector Machine, Data Preprocessing.

### 1. Introduction

Enhancing heart disease prediction through ensemble learning techniques with hyper parameter optimization represents a sophisticated approach in the field of medical data analysis and predictive modelling. Heart disease is a leading cause of morbidity and mortality globally and accurate prediction models are crucial for early intervention and improved patient outcomes. Fine-tuned hyper parameters, our goal is to improve the accuracy, reliability and generalizability of heart disease prediction models. Compile a large dataset with pertinent information to aid in the prediction of heart disease. By combining the outputs of multiple models, the ensemble approach mitigates individual model weaknesses, resulting in a more reliable and accurate predictive model. Hyper parameter optimization further refines the model parameters,

optimizing their configuration for the specific heart disease prediction task. This advanced methodology leverages machine learning to analyse diverse patient data, including medical history, diagnostic tests and lifestyle factors, to predict the likelihood of heart disease. The integration of ensemble learning techniques and hyper parameter optimization not only enhances the accuracy of predictions but also contributes to the interpretability and generalizability of the model. As technology continues to advance, such data driven approaches hold great promise in revolutionizing the early detection and prevention of cardiovascular diseases. Age, gender, blood pressure, cholesterol and other pertinent health markers are examples of features. The heart attacks are often being common and cardiovascular became the deadliest human diseases around the globe. This disease and problem

do not occur all of a sudden. Scientist and Doctors reveals that it is a continues process and is the result of being on a particular lifestyle for long time and also results after giving some basic and common symptoms being occurring all of a sudden. Eventually what does happen in the heart attacks is, the heart is unable to pump the required amount of blood to the parts of the body and more over it itself also does not get enough blood supply due to blocked arteries in the heart chambers there for it results in heart failure and deaths.

## 2. Review of Literature

Heart disease prediction has been widely studied using data mining and machine learning techniques to improve early diagnosis and reduce mortality rates. Several researchers have applied traditional classification algorithms such as Logistic Regression, Naïve Bayes, Decision Tree, and Support Vector Machine (SVM) for predicting heart disease based on clinical attributes. These methods demonstrated moderate accuracy but were often limited by overfitting, data imbalance, and lack of proper feature selection. Recent studies have focused on ensemble learning techniques such as Random Forest, Gradient Boosting, and XGBoost to enhance predictive performance. Ensemble methods combine multiple base classifiers to improve accuracy, stability, and generalization capability. Research findings indicate that ensemble models outperform single classifiers in terms of precision, recall, and overall accuracy.

Feature selection techniques have also been emphasized in the literature to identify the most significant risk factors influencing heart disease. Methods such as correlation analysis, recursive feature elimination, and information gain have been used to reduce dimensionality and improve model efficiency. Additionally, hyperparameter optimization methods like Grid Search and Random Search have been applied to fine-tune model parameters and achieve better performance. Furthermore, studies highlight the importance of proper data preprocessing, including handling missing values, normalization, encoding categorical variables, and splitting datasets into training, validation, and testing subsets. These steps ensure reliable model evaluation and better generalization.

Overall, the literature indicates that hybrid models integrating feature selection, ensemble learning, and hyperparameter tuning provide improved accuracy and robustness for heart disease prediction, supporting early detection and effective clinical decision-making.

Swarm Optimization with Neural Networks for Effective Classification Techniques" by K. Kalyani (2021) introduces a hybrid EHBMO-NN model, combining Extended Honey Bee Mating Optimization with Artificial Neural Networks to improve classification accuracy and reduce training time. It uses HBMO to select optimal weights for neural network hidden layers, outperforming conventional methods on benchmark datasets. The accurate cancer classification is very important task for cancer treatment. Recently the informative genes are identified from the thousands of genes for correct cancer classification. The collection of microscopic Deoxyribo Nucleic Acid (DNA) microarray is attached in the solid surface. In this study, DNA microarray data is used for cancer classification. The accurate cancer classification is very important task for cancer treatment. Recently the informative genes are identified from the thousands of genes for correct cancer classification. The collection of microscopic Deoxyribo Nucleic Acid (DNA) microarray is attached in the solid surface. In this study, DNA microarray

data is used for cancer classification

## 3. Existing System

The existing system for predicting heart disease heavily relies on manual processes involving X-rays, scans and doctor evaluations. These methods are time consuming and require significant expertise, making them prone to human error and subjectivity in diagnosis. The reliance on physical tests and manual analysis often delays early detection, as subtle patterns indicating heart disease may be overlooked. Furthermore, the process is resource-intensive, requiring specialized equipment and skilled professionals, which may not be readily available in all healthcare settings, especially in rural or underdeveloped areas. The manual nature of the system also limits scalability and accessibility, as each evaluation requires a doctors direct involvement. Additionally, inconsistencies in interpreting test results can lead to variations in diagnosis accuracy. This dependency on human intervention and traditional diagnostic methods highlights the need for more automated and reliable solutions to improve efficiency and prediction accuracy in heart disease detection.

### Demerits

- Time consuming Process
- Prone to Human Error
- Resource Intensive
- Inconsistent Diagnoses
- Limited Scalability

## 4. Proposed Algorithm

The proposed system leverages advanced machine learning techniques to improve heart disease prediction accuracy and reliability. By applying feature selection methods, the system identifies the most relevant attributes influencing heart disease, ensuring only critical factors are considered in the predictive model. Ensemble learning approaches are utilized to combine multiple models, enhancing robustness and reducing prediction errors. The system incorporates comprehensive preprocessing steps, such as handling missing values, normalizing features and encoding categorical variables, to prepare the dataset for effective model training. Hyperparameter optimization is implemented to fine tune the models, maximizing their performance. The dataset is split into training, validation and testing subsets to ensure proper evaluation and generalization. The final model predicts the likelihood of heart disease with high precision and displays the results using a graphical interface, enabling users to visualize predictions effectively. This automated and data driven approach offers significant potential for early detection, accurate diagnosis and better management of heart disease.

### Overview of the Proposed Algorithm

#### Data Collection

Collect heart disease dataset containing patient health attributes such as age, gender, blood pressure, cholesterol level, chest pain type, and other clinical features.

#### Data Preprocessing

- Handle missing values
- Remove duplicate or inconsistent data
- Normalize numerical features
- Encode categorical variables
- Perform data cleaning

### Feature Selection

Apply feature selection techniques to identify the most relevant attributes influencing heart disease. This reduces dimensionality and improves model efficiency and accuracy.

### Dataset Splitting

Split the dataset into:

- **Training Set:** To train the model
- **Validation Set:** To tune hyperparameters
- **Testing Set:** To evaluate final model performance

### Model Development

Train multiple machine learning algorithms such as:

- Logistic Regression
- Naïve Bayes
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)
- Decision Tree
- Random Forest
- XGBoost
- Neural Network

### Ensemble Learning

Combine multiple models using ensemble techniques (e.g., Random Forest, Gradient Boosting) to improve robustness and reduce prediction errors.

### Hyperparameter Optimization

Apply tuning techniques (Grid Search/Random Search) to optimize model parameters and improve prediction accuracy.

### Model Evaluation

Evaluate the model using performance metrics such as:

- Accuracy
- Precision
- Recall
- F1-Score

### Prediction and Visualization

Deploy the best-performing model to predict whether a patient has heart disease or not. Display results through graphical output for easy interpretation.

### Merits

- Improved Accuracy
- Robustness
- Effective Preprocessing
- Optimized Performance
- User Friendly Visualization

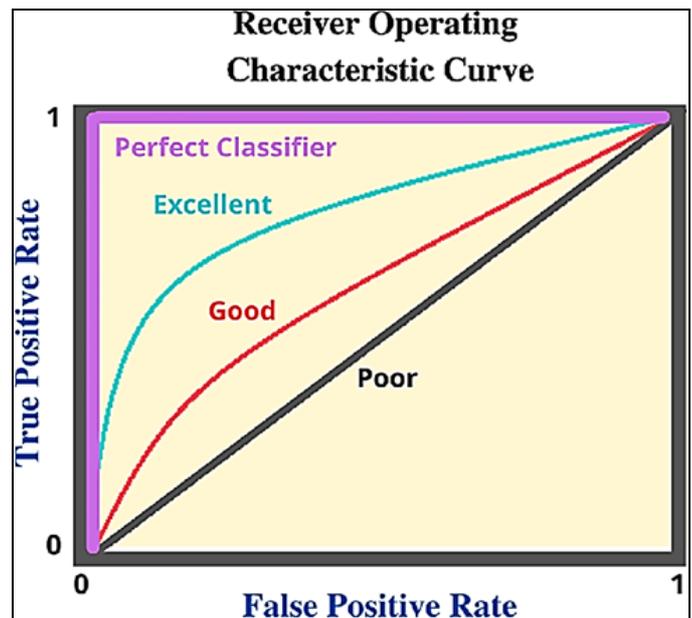
## 5. Experimented Results

The model was tested using training and testing datasets. Ensemble methods like Random Forest and XGBoost performed better than other algorithms. After hyperparameter tuning, the accuracy improved and prediction errors were reduced. The final model provides reliable heart disease prediction results.

### Datasets Used

The system uses the Heart Disease Dataset, commonly applied for cardiovascular disease prediction research. The dataset contains patient medical records with multiple clinical attributes used to determine the presence or absence of heart disease.

## ROC Curve Accuracy



## 6. Conclusion

In conclusion, the proposed machine learning-based heart disease prediction model demonstrates the potential to significantly enhance diagnostic accuracy and reliability. By leveraging feature selection, ensemble learning techniques, and hyperparameter optimization, the model effectively reduces false positives and negatives, offering a robust solution for early detection. Comprehensive data preprocessing and the integration of advanced algorithms ensure the model's ability to generalize across diverse datasets, while graphical visualization aids in interpreting the predictions. This system holds promise for improving early diagnosis, facilitating timely interventions and ultimately reducing the global burden of heart disease by supporting healthcare professionals in making data driven decisions. The study confirms that integrating optimized feature selection with ensemble learning significantly enhances heart disease prediction performance compared to conventional machine learning approaches. This work contributes to the development of intelligent healthcare systems and demonstrates how advanced machine learning techniques can assist medical professionals in making faster and more accurate diagnostic decisions.

## References

1. Valarmathi R. *Heart disease prediction using hyper parameter optimization (HPO) tuning*. 2025.
2. Kirmani MM. *An Intelligent and Reliable Hyper parameter Optimization Machine Learning Model for Early Heart Disease Assessment Using Imperative Risk Attributes*. 2024.
3. Saboor A. *A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms*. 2023.
4. Farzi M, McClymont D. *Assessing Myocardial Microstructure with Biophysical Models of Diffusion MRI*. 2023.
5. Srinivas P. *hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost*. 2022 Mar.
6. de Hoop H, Petterson NJ. *Multiperspective Ultrasound Strain Imaging of the Abdominal Aorta*. 2022.

7. Shirakawa T, Kuratani T. *Towards a Clinical Implementation of Measuring the Elastic Modulus of the Aorta from Cardiac Computed Tomography Images*. 2021.
8. Stevens RRF, Donders WP. *The Role of One-Dimensional Model-Generated Inter-Subject Variations in Systemic Properties on Wall Shear Stress Indices of Intracranial Aneurysms*. 2021.
9. Farzi M, McClymont D. *Assessing Myocardial Microstructure with Biophysical Models of Diffusion MRI*. 2020.
10. Misra P. *Machine Learning Algorithms for Optimising Heart Disease Prediction through Hyperparameter Tuning*. 2020.
11. Kalyani K. Swarm Optimization with Neural Networks for Effective Classification Techniques. *Annals of the Romanian Society for Cell Biology*. 2021;25(4):7413-7419.
12. Kalyani K. Classification of Microarray Gene Expression using Artificial Neural Network (ANN). *Turkish Journal of Computer and Mathematics Education*. 2021;12(7):1372-1378.
13. Kalyani K, Chakravarthy T. An Algorithmic Approach with Improved Replacement in Bee Optimization Algorithm. *ICTACT Journal on Soft Computing*. 2015 Jan 1;5(2):905-910.
14. Kalyani K. Microarray Data Classification using Artificial Neural Network. *International Journal of Engineering and Advanced Technology (IJEAT)*. 2019 Dec;9(1S2):54-56.