53 A novel ensemble approach for enhanced heart disease prediction

Nomula Nagarjuna Reddy[a], *Lingadally Nipun*[b], *Samanvitha Maddula*[c], *Pogiri Deepika*[d] *and Thoutireddy Shilpa*[e]

Department of Computer Science and Engineering, B V Raju Institute of Technology, Narsapur, Telangana, India

Abstract

Heart disease (HD) poses a significant health challenge across the world, responsible for approximately 17.9 million deaths annually, necessitating advanced predictive techniques for early diagnosis and intervention. This paper introduces a novel ensemble learning (EL)approach that integrates random forest (RF), XGBoost (XGB), and multilayer perceptron (MLP) classifiers to enhance the accuracy of HD prediction. By combining these models, we leverage their complementary strengths to address the limitations of individual methods. Evaluated on a comprehensive dataset of 1,190 patients, our ensemble model attained a remarkable accuracy of 95.4% and an F1 score of 95.8%, exceeding the efficiency of individual models, where RF achieved 94.5% accuracy and 94% F1 score. This significant improvement emphasizes the efficacy of EL approaches in medical diagnostics. The proposed approach not only advances predictive accuracy but also contributes to the development of robust clinical decision support systems. Our findings highlight the potential of EL to enhance early detection and patient outcomes in cardiovascular (CV) healthcare, setting a new benchmark for the incorporation of AI and ML in medical diagnostics. Future work will focus on refining the model with additional techniques and hyperparameter optimization to further elevate predictive performance.

Keywords: Ensemble learning, heart disease, machine learning, multilayer perceptron, random forest, XGBoost

Introduction

Heart disease (HD), responsible for approximately 17.9 million deaths annually, remains a leading global health concern. The diversity of CV conditions—including CAD, heart failure, and arrhythmias—emphasizes the critical importance of reliable early diagnostic methods. Despite advances in medical research, predicting HD continues to challenge healthcare due to its complex nature.

The human heart, a central organ in diagnosing and understanding CVD, exemplifies biological complexity. Recent innovations in ML and DL have shown great potential in healthcare analytics, particularly through ensemble learning, which combines multiple algorithms to enhance predictive performance. Building on prior research, which demonstrated the effectiveness of the CART algorithm achieving 87% accuracy, this work proposes a novel application of EL techniques to further improve predictive accuracy.

Our research leverages a voting classifier (VC) that integrates

RF, MLP, and XGB algorithms. This approach aims to capitalize on the strengths of each algorithm—RF's robustness in handling large datasets, MLP's efficiency, and XGB's superior performance. The study utilizes a comprehensive dataset comprising 11,190 patient records from multiple countries, featuring 11 clinical attributes and a target variable indicating heart disease presence.

By rigorously optimizing and evaluating the ensemble model using evaluation metrics, we demonstrate significant improvements in predictive performance without the need for feature extraction. This approach enhances diagnostic

[a]nagarjunareddynomula2@gmail.com, [b]nipun.poh@gmail.com, [c]maddulasamanvitha04@gmail.com, [d]pogirideepika2111@gmail.com, [e]shilpathoutireddy@gmail.com

DOI: 10.1201/9781003616399-53

accuracy while also providing valuable insights into the practical implementation of EL in medical diagnostics.

This study significantly contributes to the field by demonstrating the effectiveness of advanced ML techniques in improving HD prediction, with the goal of enhancing patient outcomes by providing more accurate and reliable diagnostic tools.

Research Objective

The objective of this proposed research is to develop and validate an advanced predictive model for HD diagnosis by employing a novel EL approach. This method integrates three distinct ML classifiers—RF, XGB, and MLP—into a VC framework. The primary goal is to improve the accuracy and reliability of HD predictions, thereby enhancing the effectiveness of clinical decision support systems. Through this approach, the paper seeks to demonstrate how combining the strengths of these individual classifiers can lead to significant improvements in diagnostic accuracy, ultimately contributing to better patient outcomes in CV care.

Literature Review

HD poses significant health risks, underscoring the need for accurate predictive models for early detection and intervention. Recent research demonstrates the effectiveness of ML and DL techniques in predicting. For instance, the CART algorithm achieved 87% accuracy in HD prediction [1], while a voting ensemble with chi-square feature selection achieved 92.11% accuracy, enhancing both performance and efficiency [2].

RF, with 88.52% accuracy, outperformed other ML algorithms, emphasizing the importance of data quality and feature selection [3]. A MLPNN with an arithmetic optimization algorithm also surpassed traditional methods with 88.89% accuracy [4]. Additionally, MLP and RF were top performers in HD prediction, while DT lagged behind [5].

XGB was the most accurate for predicting myocardial infarction, with 94.80% accuracy and 90.0% AUC, highlighting ML's potential in CVD prediction [6]. Research indicates that feature selection techniques significantly enhance the accuracy of HD prediction [7]. Improved prediction accuracy has been achieved using six ML algorithms optimized with GridSearchCV and five-fold cross-validation, particularly with a soft voting ensemble classifier [8].

Comparative studies of gradient boosting, XGBoost, and AdaBoost for HD prediction found gradient boosting to be the most accurate at 92.20% [9]. A hybrid ML method combining KNN and SVM achieved an 81% prediction accuracy [10]. Additionally, an extra tree classifier with hyperparameter optimization predicted HD with 98.15% accuracy, emphasizing the need for broader datasets and integration of clinical and genetic data [11].

Keras-based DL models have effectively diagnosed HD using Convolutional Neural Networks (CNNs) on image data [12]. An optimized ANN model improved HD prediction accuracy to 93.44%, surpassing SVM by 7.5%, and reducing training time [13]. Advanced boosting techniques like AdaBoost achieved 95% accuracy [14]. Literature highlights the need for advanced ML models for HD detection, with XGBoost excelling in accuracy and precision, particularly in addressing imbalanced datasets [15].

In conclusion, this study validates the efficacy of ML and DL algorithms for predicting HD. By integrating findings from earlier research, this study offers a more comprehensive perspective on HD prediction, underscoring the potential of ML and DL methods to advance early detection and intervention approaches. The aim is to create precise predictive models for HD through EL, specifically employing a VC with RF, MLP, and XGBoost, to enhance diagnostic accuracy and improve patient outcomes in CV care.

Methodology

Dataset

We utilize a dataset consisting of 1,190 patient records collected from several countries, such as the US, UK, Switzer- land, and Hungary. The dataset includes 11 features along with a target variable indicating HD. These features reflect important patient attributes crucial for our analysis and are outlined in Table 53.1.

Table 53.1 Description of features in dataset

Feature name	Description
ST slope	Inclination of the ST segment at maximum exercise intensity on an ECG, indicating cardiac health.
Chest pain type	Typical angina, atypical angina, non-anginal pain.
Old peak	Represents the presence of myocardial ischemia.
Resting ECG	Reveals heart's rhythm and electrical conduction.
Age	The age of a person, a critical HD factor.
Fasting blood sugar (FBS)	High FBS level (mg/dL) indicates insulin and diabetes risk.
Max heart rate	Achieved during exercise (bpm), showing heart fitness.
Sex	Gender of a person, influences CVD.
Resting BP S	Resting systolic blood pressure (mmHg), important for assessing hypertension risk.
Exercise angina	It signals decreased blood flow to the heart during exertion.
Cholesterol	High serum cholesterol level (mg/dL), reflects CVD risk.

Source: M. Siddhartha, "Heart Disease Dataset (Comprehensive)," "IEEE Dataport, Nov. 5, 2020. [Online]. Available: https://dx.doi.org /10.21227/dz4t-cm36

System framework

In this study, we developed a robust framework (as shown in Figure 53.1) for HD prediction using a diverse dataset of 11,190 patient records from multiple countries. Each record includes 11 features and a target variable indicating HD presence. Dataset underwent preprocessing to ensure high data quality. We employed 3 ML algorithms—RF, XGB, and MLP. An ensemble approach using a VC was chosen to improve predictive performance by integrating the advantages

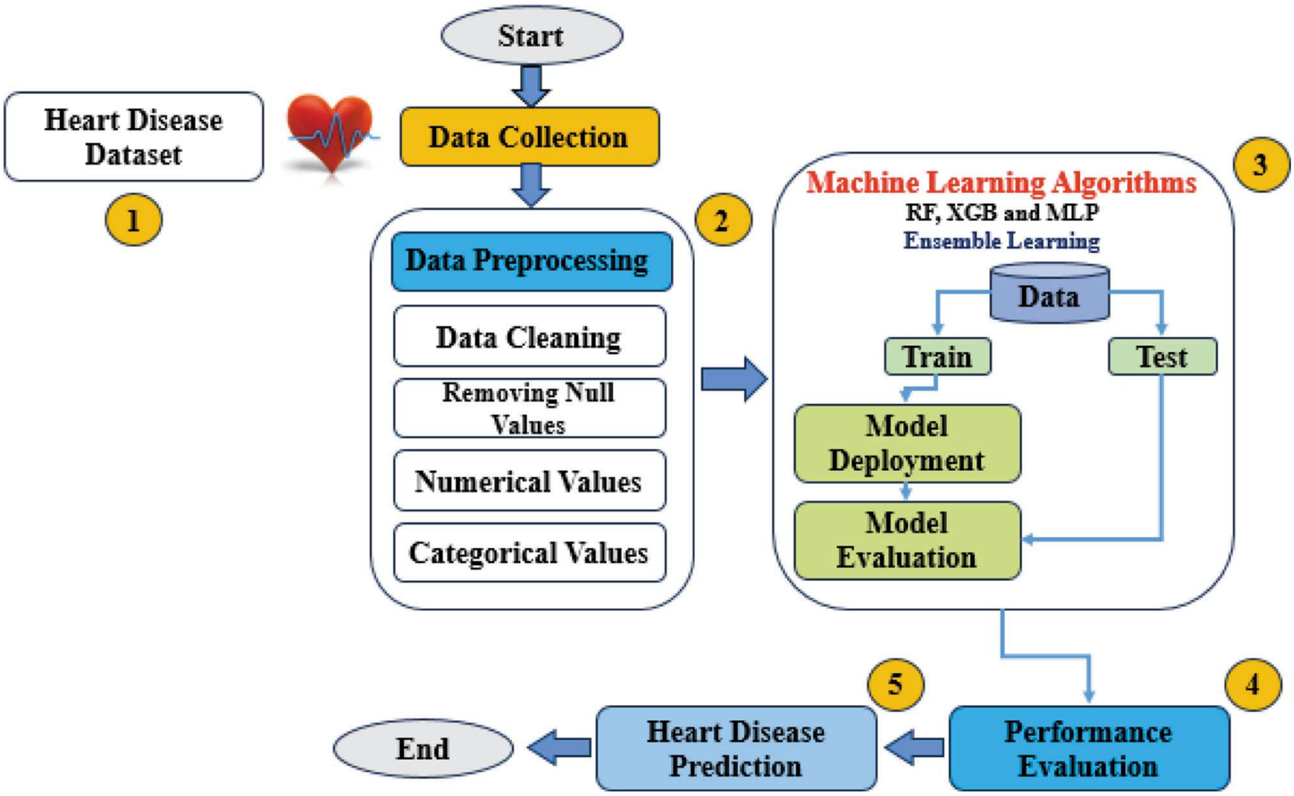

Figure 53.1 System framework
Source: Author

of the three algorithms. After training, the models were deployed for predicting new, unseen data to simulate real-world application. A comprehensive performance evaluation was conducted for our prediction framework. This approach aims to enhance healthcare decisions and facilitate prompt interventions.

Data preprocessing

Enhancing the accuracy of ML models is crucial. Practical datasets frequently have issues such as noise, missing data, and improper formats, which necessitate thorough cleaning and formatting. Key steps include handling missing values through removal or imputation, encoding categorical variables into numerical formats, and scaling features for uniform contribution during model training. This involved importing and cleansing the dataset.

Proposed ensemble methodology

Random forest: It is an EL technique for classification and regression, combining multiple DTs (Figure 53.2) to enhance predictive accuracy and reduce overfitting. Each tree is trained on distinct data subsets, and the final prediction is derived from aggregating these outputs. This method excels with large datasets, manages missing values, and handles high dimensional data.

Parameter optimization is key to mitigating overfitting, especially with noisy data. It is versatile, with applications ranging from customer churn prediction to disease diagnosis, and it provides feature importance scores for better interpretability. In our study, RF is chosen for its high accuracy and reliability in predicting HD, addressing model overfitting and data variability while offering valuable insights into significant predictors.

Multilayer perceptron: It is an ANN consisting of an input layer, hidden layers, and an output layer (Figure 53.3). It performs feedforward computations and uses Backpropagation for training, adjusting weights to minimize errors. MLPs excel at modeling non-linear relationships and handling complex datasets, making them suitable for tasks where conventional linear models fall short.

The choice of MLP is justified by its capability to approximate any continuous function with sufficient neurons and layers, capturing intricate patterns within data. Selecting an appropriate network structure, including multiple hidden layers and nodes, is crucial for balancing model complexity and preventing overfitting.

To implement an MLP, configure the network with suitable parameters, apply Backpropagation for training, and fine- tune the model to align with the specific characteristics of the dataset. This ensures robust predictive performance and adaptability to complex problems.

XGBoost: It is a powerful ML algorithm built on gradient-boosting decision trees, optimized

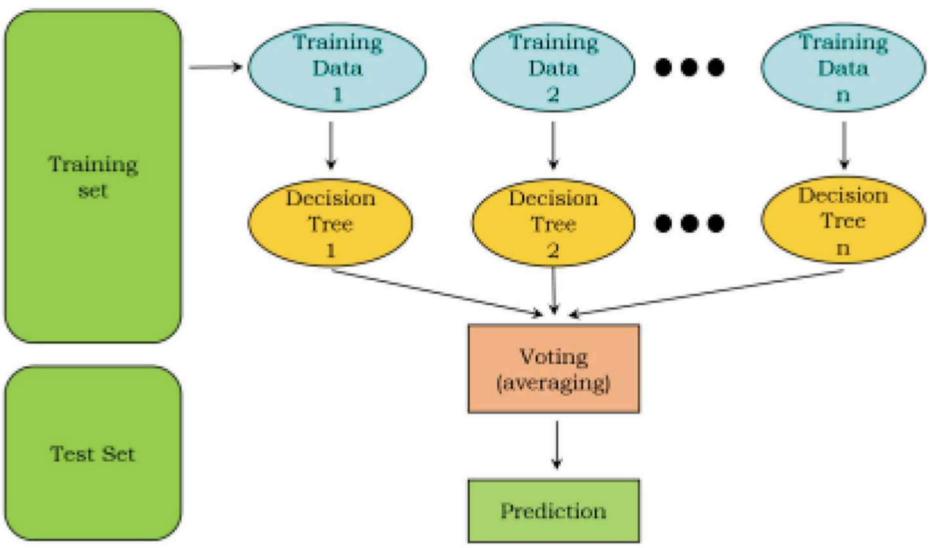

Figure 53.2 Random forest
Source: Author

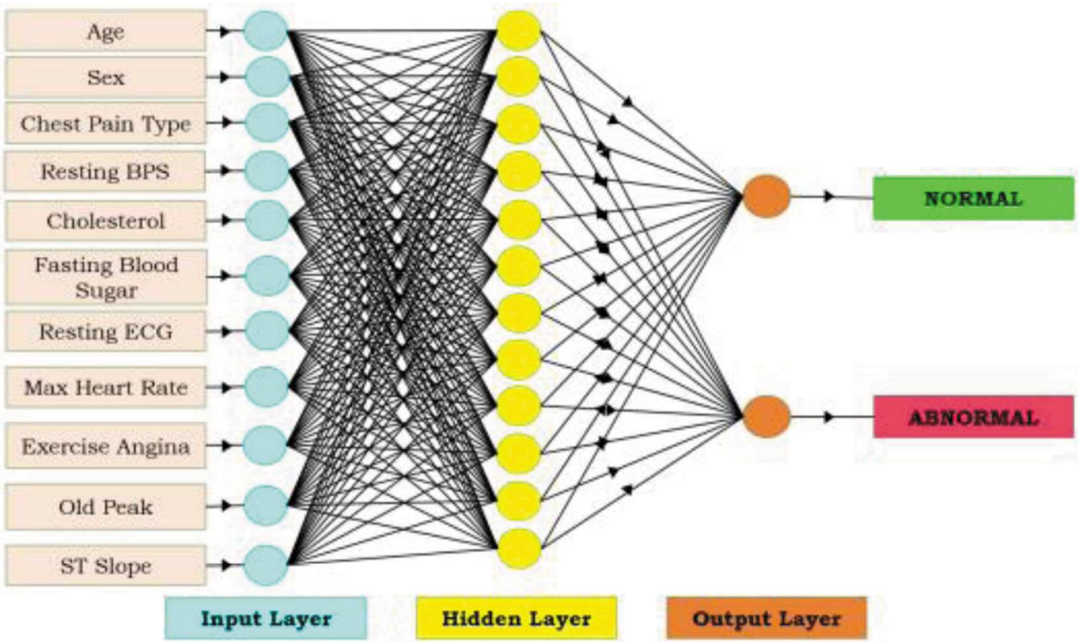

Figure 53.3 Multilayer perceptron
Source: Author

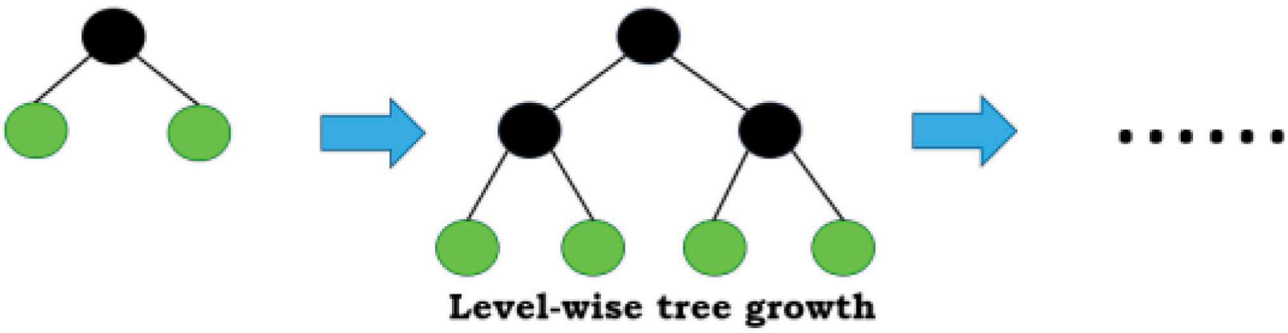

Figure 53.4 XGBoost
Source: Author

for speed, efficiency, and scalability. It iteratively constructs models that correct the errors of their predecessors, enhancing overall predictive accuracy. This ensemble approach combines predictions from multiple trees for a more accurate and robust model (Figure 53.4).

XGB's superior performance in predictive accuracy and computational speed, compared to alternatives like RF and GB Machines, makes it a preferred choice. It efficiently handles large datasets and complex features, improving model performance and resilience against overfitting. Additionally, XGB's interpretability and feature importance analysis aid in identifying key predictors in heart disease, aligning with our research objectives to enhance accuracy and reliability.

Ensemble learning

Ensemble learning (EL) is a robust technique in ML that combines several models to improve prediction accuracy. By aggregating outputs from various models, ensemble methods effectively address individual model weaknesses such as high variance and bias, resulting in more reliable and generalizable predictions. This method leverages diverse perspectives, whether from different algorithms, subsets of data, or introduced randomness during training, to form a cohesive prediction.

In this study, we employ a combination of RF, MLP, and XGB. RF is used to stabilize predictions and contribute valuable feature importance insights, while XGB is included for its superior accuracy and efficiency. MLP is selected for its capability to capture complex patterns through nonlinear relationships. By utilizing a VC with hard voting, we combine these models to leverage their individual strengths, thereby enhancing predictive performance for HD detection and addressing the limitations of each model.

This ensemble approach aims to increase accuracy and reliability of predictions, making it a powerful strategy for practical applications in medical diagnostics.

Voting classifier: It is an ensemble method that boosts prediction accuracy by aggregating outputs from multiple base models, such as RF, MLP, XGB, each trained independently. It uses majority voting to show the final class label from these models. Base models are trained with optimized hyperparameters, and their predictions are combined through hard voting, selecting the label with the most votes. This process leverages the strengths of each model while minimizing individual weaknesses, resulting in more robust and reliable predictions. By integrating diverse predictors, this technique effectively reduces model bias and variance, making it particularly valuable for HD prediction.

Model training and testing

The dataset is divided into 80% for training to develop the models and 20% for testing to assess their performance.

Performance analysis

We use the following metrics to evaluate model performance, each providing distinct insights into the model's reliability and diagnostic capability. It delineates true negatives (TrNe), false negative (FaNe), and false positive (FaPo), true positives (TrPo).

1) **Accuracy:** It illustrates the proportion of correct predictions relative to all cases. Although it provides a broad assessment of model performance, it can be misleading in datasets with imbalanced classes, where one class may dominate the dataset.

$$\mathrm{Acc} = \frac{\mathrm{TrPo+TrNe}}{\mathrm{TrPo+TrNe+FaPo+FaNe}} \quad (1)$$

2) **Precision:** It quantifies the ratio of TP predictions among all +ve predictions. A high precision indicates fewer FP, which is essential for accurately identifying individuals with heart disease.

$$\mathrm{Pre} = \frac{\mathrm{TrPo}}{\mathrm{TrPo+FaPo}} \quad (2)$$

3) **Recall:** It is the model's ability to detect all actual +ve cases. It is crucial for minimizing missed diagnoses and ensuring that as many positive instances as possible are captured.

$$\mathrm{Re} = \frac{\mathrm{TrPo}}{\mathrm{TrPo+FaNe}} \quad (3)$$

4) **F1-score:** It integrates the above 2 metrics into a single metric, offering a balanced measure of model performance. It provides a comprehensive evaluation by managing the balance between precision and recall, reflecting the model's overall effectiveness.

$$\mathrm{F1} = 2 \times \frac{\mathrm{Pre \times Re}}{\mathrm{Pre+Re}} \quad (4)$$

In summary, these offer key understanding into the performance of HD prediction. These metrics help in analysing model advantages and disadvantages, guiding the selection and optimization of algorithms to enhance diagnostic accuracy and patient care.

Results

Evaluation metrics of various ML models and their combinations are summarized in Figure 53.5. Among the individual models, SVM achieved an acc of 84.9%. NB slightly outperformed SVM with an acc of 85.7%. Logistic regression (LR) recorded an accuracy of 86.1%. A significant improvement in performance was observed with the MLP, which achieved an acc of 91.2%. XGB further improved the metrics, reaching an acc of 92.8%. The RF model surpassed all individual models with an acc of 94.5% and an F1 of 94%.

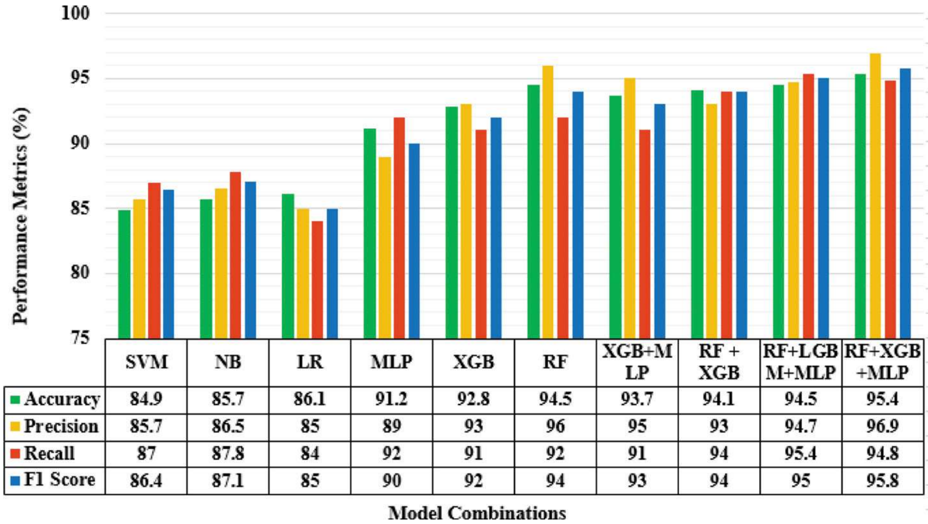

Figure 53.5 Performance analysis
Source: Author

The ensemble models show the best performance when compared to individual models. The combination of MLP and XGB achieved an acc of 93.7%. The ensemble of RF and XGB showed an acc of 94.1%. Notably, the ensemble of RF, LGBM, and MLP achieved an acc of 94.5%. The highest performance was recorded by ensemble of RF, XGB, and MLP, with an acc of 95.4% and an F1 of 95.8% (Figure 53.6 shows confusion matrix and Figure 53.7 shows model comparison).

Discussion

The results clearly denote that EL significantly improve the predictive performance for HD detection compared to individual models. Among the individual models, RF showed the highest accuracy and F1 score, confirming its effectiveness. in handling complex datasets with its feature importance, as shown in Figure 53.8. The improved performance of the MLP highlights its capability in capturing non-linear patterns within the data, evidenced by Figure 53.9, while XGB proved efficient in dealing with imbalanced datasets, as shown in Figure 53.10.

The ensemble models, particularly the combination of RF, XGB, and MLP, achieved the best results. This improvement can be attributed to the complementary strengths of the individual models. RF contributes robust overall performance, XGB efficiently manages imbalanced data and enhances precision, while MLP captures complex, non-linear relationships. By aggregating the outputs of these models through hard voting, the ensemble method minimizes individual model weaknesses and boosts overall prediction accuracy and reliability, as shown in Figure 53.11.

The ensemble of RF, XGB, and MLP acquired an acc of 95.4% and an F1 of 95.8%, underscoring the strength of EL in clinical decision support systems. These findings align with existing literature, which advocates for the use of ensemble techniques to enhance predictive performance. The integration of diverse models reduces model bias and variance, leading to more robust and reliable predictions. This study reinforces the value of ensemble methods in the early detection and intervention of HD, ultimately contributing to improved patient outcomes in cardiovascular healthcare.

In summary, the application of EL, specifically through the VC framework integrating RF, XGB, and MLP, offers a powerful approach to HD prediction. The results validate the capability of this method in achieving high accuracy and F1 scores, thus supporting its implementation in predictive analytics for CVD. Future research could explore the integration of additional models and further optimization of hyperparameters to continue enhancing predictive performance.

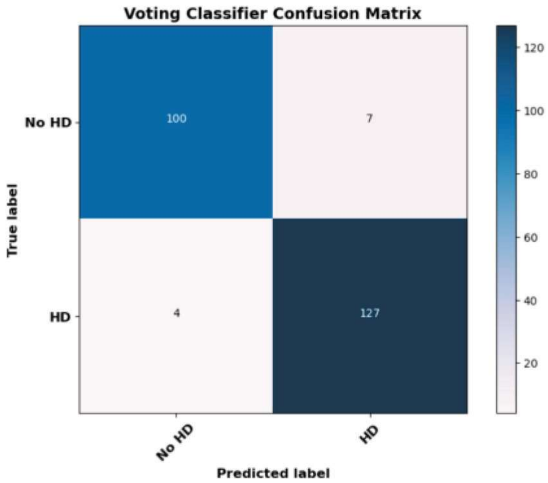

Figure 53.6 Voting classifier confusion matrix
Source: Author

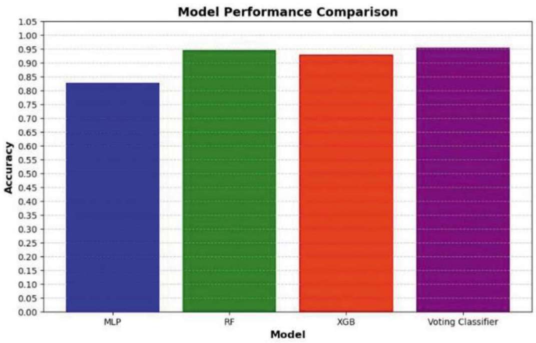

Figure 53.7 Model performance comparison
Source: Author

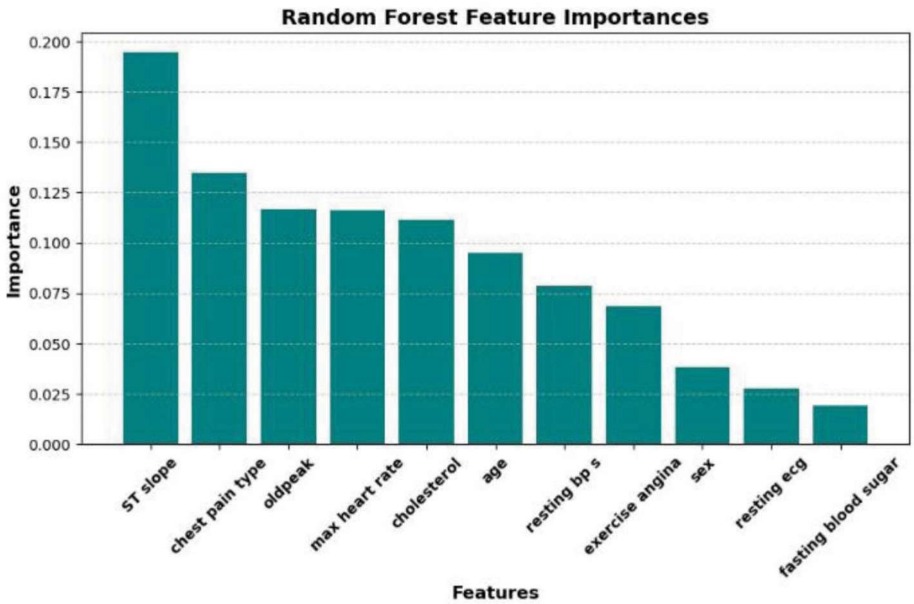

Figure 53.8 Random forest feature importance
Source: Author

Conclusion

This research offers an in-depth examination of ML and EL techniques for the prediction of HD, emphasizing the significance of accurate early detection methods in cardiovascular healthcare. The investigation demonstrated that ensemble methods significantly outperformed individual models performance.

Through the application of a VC, the ensemble model acquired a notable accuracy of 95.4% and an F1 score of 95.8%, highlighting its robustness and reliability in predicting HD. These results underscore the effectiveness of leveraging diverse model strengths to minimize individual weaknesses, thereby enhancing overall predictive performance. The research utilized a comprehensive dataset from multiple countries, encompassing 1,190 patient records with 11 features, ensuring the robustness and generalizability of the findings. A key novelty of this paper lies in the successful integration of RF, XGB, and MLP within a VC framework, demonstrating the enhanced performance of this ensemble method over individual classifiers. This approach showcases the innovative application of EL techniques in the medical domain, providing a robust and reliable tool for HD prediction.

However, there are some limitations to this study. The dataset, while comprehensive, may

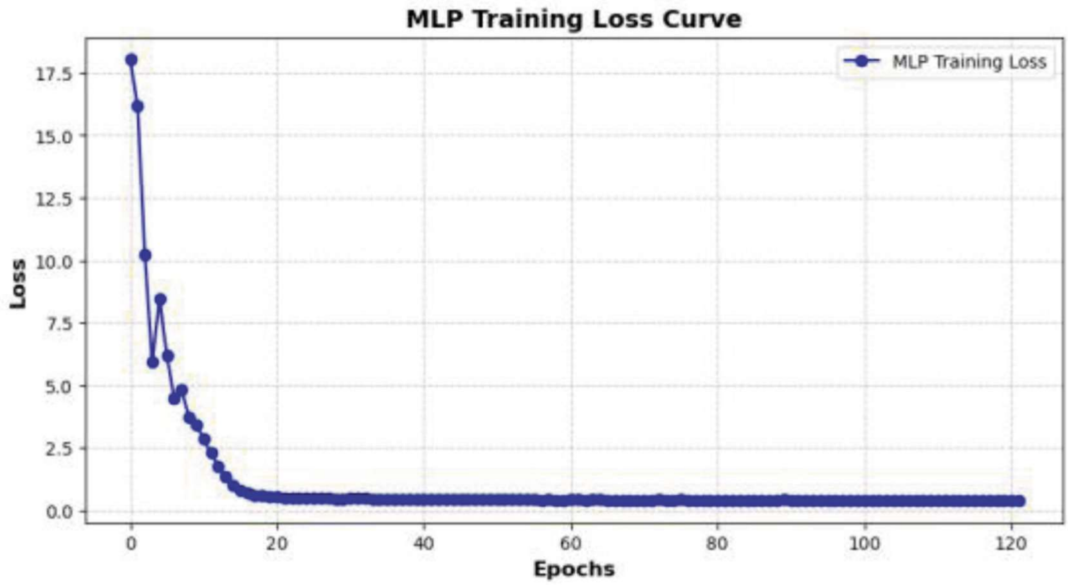

Figure 53.9 MLP training loss curve
Source: Author

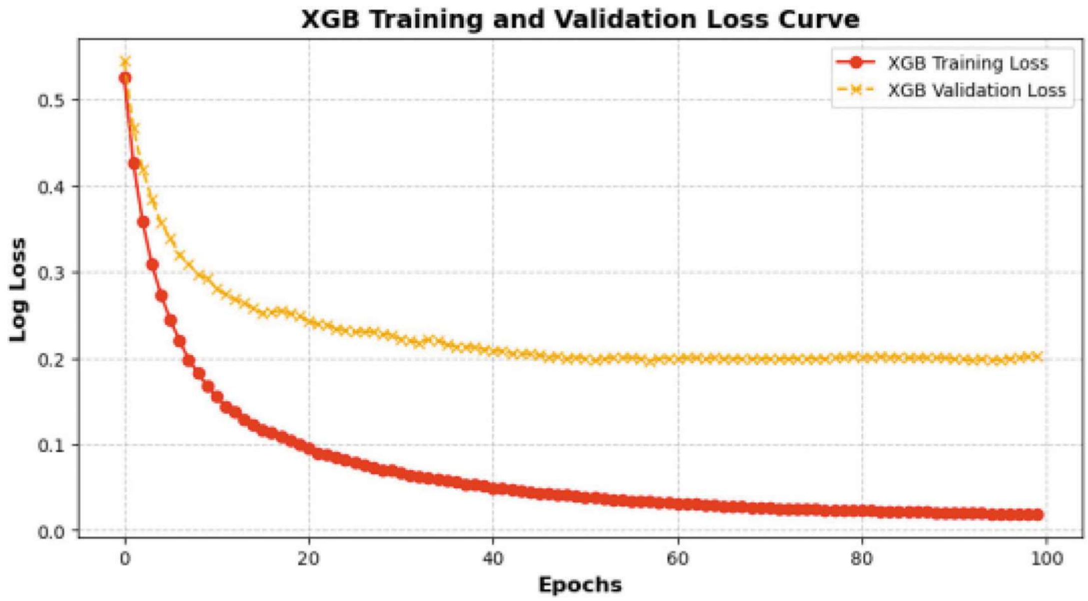

Figure 53.10 XGB training and validation curve
Source: Author

still have biases inherent to the specific populations from which it was collected. Additionally, the model's performance could vary with different or more diverse datasets. Another limitation is the absence of feature extraction techniques, which might further enhance the predictive capabilities of the models.

Future work could overcome these limitations by incorporating feature extraction methods and testing the models on more varied and extensive datasets. Additionally, exploring the integration of additional ML models and further optimizing ensemble strategies could lead to even better predictive accuracy.

In conclusion, this research confirms the capability of EL techniques for predicting HD, marking a substantial contribution to medical diagnostics. While the study has certain limitations, it offers

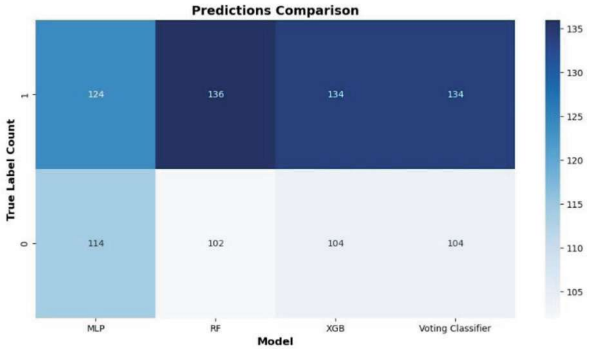

Figure 53.11 Prediction comparison matrix
Source: Author

important insights into the potential of EL methods in supporting clinical decision-making. This study paves the way for more sophisticated and reliable predictive tools, ultimately enhancing patient outcomes and contributing to the broader goal of effective CVD management.

References

[1] Ozcan, M., and Peker, S. (2023). A classification and regression tree algorithm for heart disease modeling and prediction. *Healthcare Analytics*, 3, 100130. doi: 10.1016/j.health.2022.100130.

[2] Korial, A. E.,Gorial, I. I., andHumaidi, A. J. (2024). An improved ensemble-based cardiovascular disease detection system with chi-square feature selection. *Computers*, 13(6), 126. doi: 10.3390/computers13060126.

[3] Stonier, A. A.,Gorantla, R. K., and Manoj, K. (2024). Cardiac disease risk prediction using machine learning algorithms. *Healthcare Technology Letters*, 11(4), 213–217. doi: 10.1049/htl2.12053.

[4] Alghamdi, F. A., Almanaseer, H., Jaradat, G., Jaradat, A., Alsmadi, M. K., Jawarneh, S., et al. (2024). Multilayer perceptron neural network with arithmetic optimization algorithm based feature selection for cardiovascular disease prediction. *Machine Learning and Knowledge Extraction*, 6(2), 987–1008. doi: 10.3390/make6020046.

[5] Mijwil, M., Faieq, A. K., and Aljanabi, M. (2024). Early detection of cardiovascular disease utilizing machine learning techniques: evaluating the predictive capabilities of seven algorithms. *Iraqi Journal For Computer Science and Mathematics*, 5(1), 263–276. doi: 10.52866/ijcsm.2024.05.01.018.

[6] Anjum, N., Siddiqua, C. U., Haider, M., Ferdus, Z., Raju, M. A. H., Imam, T., et al. (2024). Improving cardiovascular disease prediction through comparative analysis of machine learning models. *Journal of Computer Science and Technology Studies*, 6(2), 62–70. doi: 10.32996/jcsts.2024.6.2.7.

[7] Pathan, M. S., Nag, A.,Pathan, M. M., and Dev, S. (2022). Analyzing the impact of feature selection on the accuracy of heart disease pre- diction. *Healthcare Analytics*, 2, 100060. doi: 10.1016/j.health.2022.100060.

[8] Chandrasekhar, N., and Peddakrishna, S. (2023). Enhancing heart disease prediction accuracy through machine learning techniques and optimization. *Processes*, 11(4), 1210. doi: 10.3390/pr11041210.

[9] Ganie, S. M., Pramanik, P. K. D., Malik, M. B., Nayyar, A., and Kwak, K. S. (2023). An improved ensemble learning approach for heart disease prediction using boosting algorithms. *Computer Systems Science and Engineering*, 46(3), 3993–4006. doi: 10.32604/csse.2023.035244.

[10] Ahmed, R., Bibi, M., and Syed, S. (2023). Improving heart disease prediction accuracy using a hybrid machine learning approach: a comparative study of SVM and KNN algorithms. *International Journal of Computations, Information and Manufacturing (IJCIM)*, 3(1), 49–54. doi: 10.54489/ijcim.v3i1.223.

[11] Asif, D., Bibi, M.,Arif, M. S., and Mukheimer, A. (2023). Enhancing heart disease prediction through ensemble learning techniques with hyper- parameter optimization. *Algorithms*, 16(6), 308. doi: 10.3390/a16060308.

[12] Almazroi, A. A.,Aldhahri, E. A., Bashir, S., and Ashfaq, S. (2023). A clinical decision support system for heart disease prediction using deep learning. *IEEE Access*, 11, 61646–61659. doi: 10.1109/ACCESS.2023.3285247.

[13] Sarra, R. R., Dinar, A. M., and Mohammed, M. A. (2022). Enhanced accuracy for heart disease prediction using artificial neural network. *Indonesian Journal of Electrical Engineering and Computer Science*, 29(1), 375. doi: 10.11591/ijeecs.v29.i1.pp375-383.

[14] Nissa, N., Jamwal, S., and Neshat, M. (2024). A technical comparative heart disease prediction framework using boosting ensemble techniques. *Computation*, 12(1), 15. doi: 10.3390/computation12010015.

[15] Ogunpola, A., Saeed, F., Basurra, S.,Albarrak, A. M., and Qasem, S. N. (2024). Machine learning-based predictive models for detection of cardio vascular diseases. *Diagnostics*, 14(2), 144. doi: 10.3390/diagnostics14020144.