A Machine Learning Approach for Optimized Heart Disease Diagnosis with SMOTE and Voting Classifiers

A.S.M. Sabiqul Hassan
Department of Computer Science and Engineering
Northern University Bangladesh
Dhaka, Bangladesh

Tanzina Tazreen Meem
Department of Computer Science and Engineering
Bangladesh University of Business and Technology
Dhaka, Bangladesh

Md. Ruhul Amin
Department of Computer Science and Engineering
Southeast University
Dhaka, Bangladesh

Tasniah Mohiuddin
Department of Computer Science and Engineering
Bangladesh University of Business and Technology
Dhaka, Bangladesh

Muhammed Samsuddoha Alam
Department of Computer Science and Engineering
Northern University Bangladesh
Dhaka, Bangladesh

Mst. Najnin Sultana
Department of Computer Science and Engineering
Northern University Bangladesh
Dhaka, Bangladesh

ABSTRACT
Heart disease is globally considered a primary cause of a notable number of deaths. Each year, 17.9 million people die from heart disease, according to a report by the World Health Organization (WHO). In this study, a machine learning based optimal model has been developed that primarily includes SMOTE for handling class imbalance in the dataset, and ensemble learning strategies to improve the performance and reliability of heart disease diagnosis. This research work has been conducted on a publicly available dataset from the Kaggle online dataset repository, which includes relevant attributes for heart disease patients. Several base models: LR, KNN, DT, RF, and SVM have been trained for performance evaluation in terms of Accuracy, Precision, Recall, F1-score, and ROC-AUC values. SMOTE has been applied to address the class imbalance issue in the dataset and Soft Voting and Hard Voting classifiers have been used to optimize the model performance by combining all base classifiers. Finally, the Soft Voting classifier has achieved the optimal result: an Accuracy of 70.5%, Precision of 69.8%, Recall of 72.2%, F1-score of 71%, and ROC-AUC of 77%. This optimal model can be used as a decision making tool in the healthcare sector for the early diagnosis of heart diseases followed by necessary steps to prevent those diseases.

Keywords
Heart Disease Diagnosis, Classification, Class Imbalance, Ensemble Learning, Decision Making

1. INTRODUCTION
Heart disease, also known as cardiovascular disease (CVD), is a crucial global concern in the healthcare sector, as 17.9 million people die each year, representing around 32% of global deaths [1]. The mortality rate can be reduced, and the quality of life for patients can be improved through early diagnosis and timely intervention of heart disease. Although there are some effective traditional diagnostic methods such as angiography, echocardiography, and electrocardiograms, these are time-consuming, costly, and require expert interpretation [2]. Recently, prediction and diagnosis of heart diseases based on clinical data has achieved notable performance improvements using different machine learning techniques. These techniques are used to process large volumes of medical data and extract patterns that cannot be discovered manually. Therefore, experts can predict heart diseases more accurately [3].

However, class imbalance is a significant issue for medical diagnosis using machine learning approaches. In such scenarios, there can be a large difference between the disease prediction results (Yes or No) which can cause the model to become biased in the identification process of heart disease for many patients [4]. To overcome this issue, SMOTE (Synthetic Minority Oversampling Technique) has been widely applied by experts. It increases the minority class artificially by creating synthetic examples. It also helps to balance the dataset and to improve the generalization ability of the machine learning models used [5].

In addition, the use of ensemble learning techniques for example, a voting classifier can enhance the prediction accuracy by integrating the results of different base machine learning models [6]. Different machine learning algorithms for example, LR, RF, SVM, and KNN have been explored in numerous studies to build models for heart disease prediction. These models can often outperform the performance of individual models after the combination of SMOTE and ensemble techniques in terms of various performance metrics [7].

The objective of this research is to build an efficient heart disease detection model using machine learning classifiers where imbalanced data is handled by SMOTE and the performance of the model is optimized by ensemble learning techniques specially, Soft Voting and Hard Voting classifiers. The optimal model has been created using a publicly available dataset from the Kaggle online dataset repository [8] and it can be used later for decision making [9]–[11] purposes in heart disease prediction.

The paper is structured in the following sequence: Section 2 discusses previous studies conducted by experts on heart disease prediction; Section 3 describes the step by step development of the proposed optimal model; Section 4 analyzes the performance of the developed model; and Section 5 concludes the research and provides guidance for future improvements.

2. RELATED WORKS

Recently, a trend has been noticed in the healthcare sector for heart disease prediction by integrating machine learning techniques. The effectiveness of supervised machine learning algorithms has been demonstrated in several studies for the early diagnosis and prediction of heart diseases. In the study conducted by S. Mohan et al. [12], heart disease had been predicted using an ensemble classifier created with RF, SVM, and NB classifiers. Their model had achieved an accuracy of 88.7% on the UCI heart disease dataset by incorporating hybrid feature selection techniques. The main focus of the study was on the importance of integrating multiple classifiers to improve generalization and reduce bias in prediction.

According to the proposal by M. S. Alom et al. [13], they had developed a diagnostic system for heart disease prediction using machine learning classifiers: LR, SVM, and DT. Their model had been applied to real clinical datasets and had been stated that ensemble learning techniques could produce better performance in terms of F1-score compared to individual models. They had also observed that class imbalance negatively impacted recall for the minority classes and posed a significant challenge. D. Elreedy et al. [14] had introduced the SMOTE to address the class imbalance issue and later, it had been adopted as a standard practice in healthcare applications. The generation of synthetic examples for the minority classes using SMOTE had significantly improved sensitivity in heart disease prediction models.

The work conducted by W. Baddah et al. [15] had implemented SMOTE with a RF classifier on a heart failure related imbalanced dataset and had achieved improved recall and AUC scores. Another study conducted by B. Warner et al. [16] had analyzed the performance of several machine learning classifiers: KNN, SVM, and DT. These models had been integrated into a voting classifier which generated better performance with an overall accuracy of 89.7% and demonstrated the benefits of ensemble learning. In [15]–[16], SMOTE and voting classifier techniques had been implemented separately but these two techniques had not been implemented together. As the integration of these two techniques increases the performance of the model, both techniques have been implemented together in this paper to evaluate their combined effects.

Similarly, the study conducted by O. Hrizi et al. [17] had combined a voting based ensemble method with feature engineering techniques for heart disease detection. They had highlighted that model diversity was influential in ensemble performance and the use of heterogeneous classifiers provided better results. For real-time disease prediction, some recent developments had integrated machine learning based systems into web or mobile applications. The study by K. Lakshmanan and P. Gomathi [18] had proposed a real-time heart disease prediction system that used multi-layer perceptron neural networks. However, they had cited the lack of interpretability of deep models as a limitation for clinical adoption.

This research proposes an optimal model, developed based on previous studies by integrating SMOTE to overcome the class imbalance issue, several supervised machine learning classifiers as base models and ensemble learning strategies: Hard Voting and Soft Voting for enhanced performance. The main impact of the research is the usability of the developed optimal model as a decision making [9]–[11] tool for the early diagnosis of heart disease patients.

3. PROPOSED APPROACH

In this research, a machine learning based heart disease prediction system has been developed by integrating SMOTE to handle class imbalance issues and voting classifier techniques for enhanced performance. The dataset used in this project has been collected from the Kaggle online dataset repository [8]. The best performing model can be used later in decision making [9]–[11] for heart disease patients as a diagnostic tool.

The full coding part has been implemented on Google Colab platform using Python and its various libraries: NumPy, Pandas, Matplotlib, and Seaborn. The development process of the optimal machine learning model is described below in Figure 1.

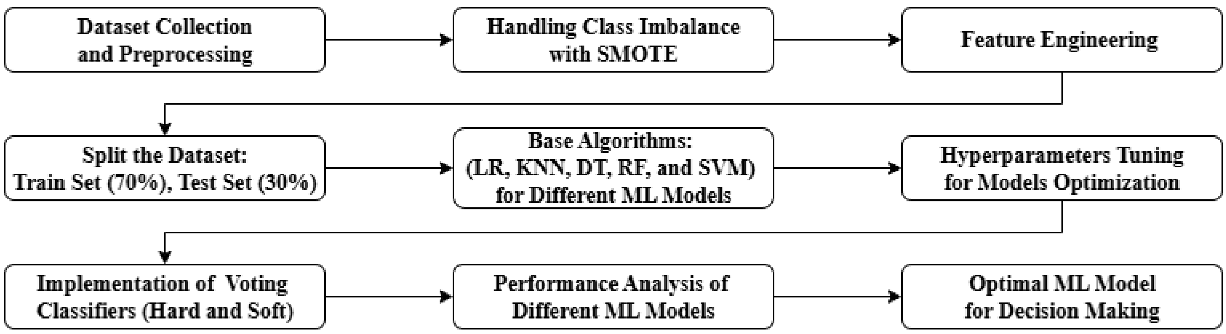


Figure 1: Workflow of the Optimal Machine Learning Model Development

3.1 Dataset Collection and Preprocessing

The dataset used in this research has been collected from the Kaggle online dataset repository [8] containing many health indicators and risk factors related to heart disease. This dataset includes the record of 10,000 heart disease instances and 21 related features. There are both categorical and numerical features, including a target variable that represents the status (Yes or No) of heart disease. Table 1 represents the description of the dataset.

Table 1. Description of the Dataset

Feature	Data Example	Data type
Age	56, 69, 32, etc.	Numerical
Gender	Male, Female	Categorical
Blood Pressure	120, 153, 146, etc.	Numerical
Cholesterol Level	155, 286, 216, etc.	Numerical
Exercise Habits	High, Medium, Low	Categorical
Smoking	Yes, No	Categorical
Family Heart Disease	Yes, No	Categorical
Diabetes	Yes, No	Categorical

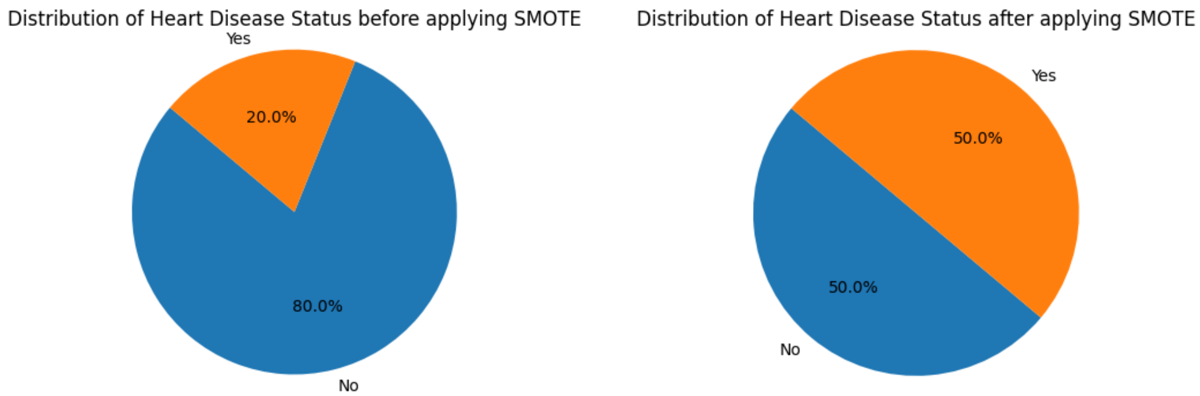
BMI	24.9, 25.2, 29.9, etc.	Numerical
High Blood Pressure	Yes, No	Categorical
Low HDL Cholesterol	Yes, No	Categorical
High LDL Cholesterol	Yes, No	Categorical
Alcohol Consumption	High, Medium, Low	Categorical
Stress Level	High, Medium, Low	Categorical
Sleep Hours	7.6, 8.7, 4.4, etc.	Numerical
Sugar Consumption	High, Medium, Low	Categorical
Triglyceride Level	342, 133, 393, etc.	Numerical
Fasting Blood Sugar	157, 92, 94, etc.	Numerical
CRP Level	12.9, 9.4, 12.7, etc.	Numerical
Homocysteine Level	12.4, 19.3, 11.2, etc.	Numerical
Heart Disease Status	Yes, No	Categorical

The preprocessing of the dataset includes several important steps: handling missing values, encoding categorical features and feature scaling for numerical features. There are two usual methods for handling missing values: removal and imputation [19]. In this project, the second method has been embraced for handling missing values. For example, all missing values for categorical features have been replaced with the mode of those feature values and numerical features have been replaced with the median of those feature values [20].

There are two types of categorical features conversion techniques: one-hot encoding and label encoding [21]. One-hot encoding replaces the current category values with a binary format, and label encoding assigns a particular numerical value to each available category. To achieve a better performance, numerical features have been scaled using StandardScaler to normalize numerical values [22]. The preprocessing steps followed in the project are consistent with standard practices in medical data modeling.

3.2 Handling Class Imbalance Issue with SMOTE

Disease diagnosis related medical datasets often face a class imbalance issue. In such cases, there is a notable difference in the number of positive and negative results. In this paper, SMOTE technique has been followed to overcome that issue [4]–[5]. In this approach, synthetic samples of the minority class are generated by interpolating between existing minority instances. Thus, it helps classifiers learn more accurate decision boundaries without over-fitting. The pie charts given below show the effect of using SMOTE on the collected imbalanced dataset, and the imbalance issue has been significantly improved after the application of SMOTE.



(a) **(b)**

Figure 2: Distribution of Heart Disease Status (a) before the application of SMOTE and (b) after the application of SMOTE

3.3 Feature Selection and Split the Dataset

Feature selection has been implemented to enhance the model performance and reduce the computational complexity using an integration of two techniques: (1) Correlation Analysis to identify multicollinearity, and (2) Recursive Feature Elimination (RFE) with cross-validation to select the most influential features for model training purposes [23].

After the feature engineering steps, the following features have been selected for further processing: Exercise Habits, BMI, Alcohol Consumption, Stress Level, Sugar Consumption, Gender (Male), Smoking (Yes), Family Heart Disease (Yes), Diabetes (Yes), High Blood Pressure (Yes), Low HDL Cholesterol (Yes), High LDL Cholesterol (Yes). Then, the dataset has been divided into two partitions: a training set and a testing set, where the training set contains 70% of the total heart disease diagnosis instances, and the training set contains the remaining 30% of the dataset.

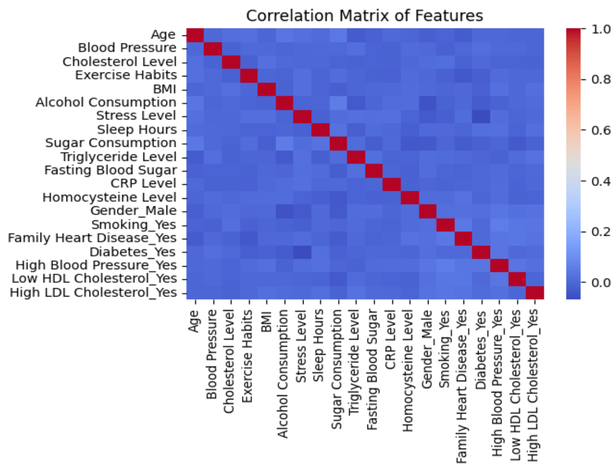


Figure 3: Correlation Matrix of Features

3.4 Applied Machine Learning Algorithms

Several base prediction models for heart disease diagnosis have been trained using the following five mostly used machine learning algorithms. During the training of these models, stratified 5-fold cross-validation has been applied to maintain class distribution across folds [24]. As hyperparameter tuning was conducted using Grid Search optimization for each base classifier, the best-performing configurations had been identified [25].

3.4.1 Logistic Regression (LR)
It is a statistical method used for binary classification tasks [26]. A logistic function is used to model the probability of a categorical dependent variable. Several classification tasks are employed using this algorithm for example, disease diagnosis and spam detection.

3.4.2 K-Nearest Neighbors (KNN)
It is a non-parametric, instance based learning algorithm [27]. It classifies the new data points based on the majority class result of its surrounding 'k' nearest neighbors in the selected feature space. For its simplicity and effectiveness, it is used for pattern recognition and classification tasks.

3.4.3 Decision Tree (DT)
It creates a flowchart like tree structure for both classification and regression tasks [28]. The tree contains some components: internal nodes which represent decision rules, branches represent outcomes and leaf nodes represent class labels. It performs well and interprets both categorical and numerical data.

3.4.4 Random Forest (RF)
It is an ensemble learning technique that contains multiple decision trees [29]. The results of these trees are merged for better prediction. This algorithm can overcome over-fitting issue to improve generalization of a base model.

3.4.5 Support Vector Machine (SVM)
It is a supervised learning algorithm that can identify the optimal hyperplane and separate classes in high-dimensional space for best possible result [30]. The use of kernel tricks made this algorithm effective in high-dimensional and non-linear classification.

3.5 Ensemble Learning Techniques
Ensemble learning techniques: Hard Voting and Soft Voting classifiers have been implemented for further improvement of predictive performance and generalization [31]. The Hard Voting classifier predicts based on the majority class results predicted by the base models. Moreover, the Soft Voting classifier finds the average probability outputs of each classifier to provide the final prediction. These two techniques enhance the performance by reducing individual model bias and provide better prediction for disease diagnosis.

4. RESULTS AND DISCUSSION
After the analysis and preprocessing of the collected dataset, five base classifiers: LR, KNN, DT, RF, and SVM have been trained. SMOTE and voting classifier techniques have been applied on the dataset to address class imbalance issues with an expectation for achieving more accurate results.

4.1 Performance Evaluation Metrics
The performance of these base classifiers including both voting classifiers are measured using the confusion matrix. A confusion matrix includes several parameters: TP (True Positive), FP (False Positive), TN (True Negative), and FN (False Negative), which summarize the performance of each trained model.

Table 2. Statistics of Confusion Matrix for Different Classifiers

Classifiers	TP	FP	TN	FN
LR	1545	903	1497	855
KNN	1761	973	1427	639
DT	1539	847	1553	861
RF	1730	673	1727	670
SVM	1712	657	1743	688
Hard Voting	1733	735	1665	667
Soft Voting	1733	748	1652	667

Based on the parameters of the confusion matrix for all classifiers, several evaluation metrics: Accuracy, Precision, Recall, F1-score, and ROC-AUC are calculated as follows [32]–[34].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1-score = 2 * \frac{Precision * Recall}{Precision+Recall} \quad (4)$$

Accuracy mentioned in the Eq. (1) measures the percentage of correctly predicted instances for both true positives and true negatives among all predicted instances. It performs well when the classes are evenly distributed in the dataset. Precision mentioned in the Eq. (2) measures the percentage of true positives among all predicted positives. It reflects the model's capability in avoiding false positives. Recall mentioned in the Eq. (3) measures the percentage of actual positives among the correctly predicted by a model. It is important particularly in situations like disease detection where the case of a missing value can be costly.

F1-score mentioned in the Eq. (4) is calculated as the harmonic mean of precision and recall values. It can balance the trade-off between these two metrics and performs better when the class distribution in the dataset is imbalanced. ROC-AUC measures the capability of a classification model to distinguish between classes. The ROC curve plots the true positive rate against the false positive rate. The values of AUC range from 0.5 (random) to 1 (perfect classifier).

Table 3. Statistics of Evaluation Metrics across Different Classifiers

Classifiers	Accuracy	Precision	Recall	F1-score	ROC-AUC
LR	0.633750	0.631127	0.643750	0.637376	0.681491
KNN	0.664167	0.644111	0.733750	0.686015	0.726476
DT	0.644167	0.645013	0.641250	0.643126	0.678770
RF	0.720208	0.719933	0.720833	0.720383	0.782857

SVM	0.719792	0.722668	0.713333	0.717970	0.779580
Hard Voting	0.707917	0.702188	0.722083	0.711997	N/A
Soft Voting	0.705208	0.698509	0.722083	0.710100	0.770795

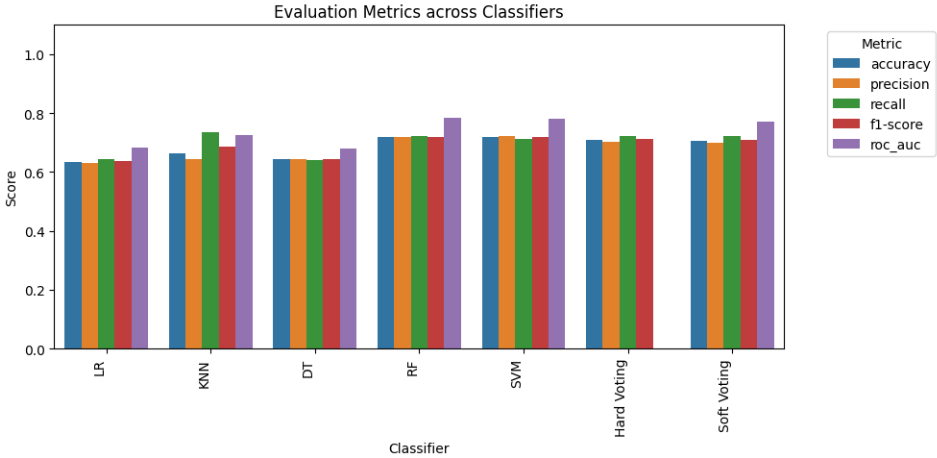


Figure 4: Statistics of Evaluation Metrics across Different Classifiers

4.2 Performance Comparison

The metrics mentioned in the section 4.1 are used to analyze the performance of all developed classifiers. Table 3 and Figure 4 represent the statistics of evaluation metrics across different classifiers, and Figure 5 represents the statistics of the ROC-AUC curve across different classifiers.

Firstly, the accuracy values for LR, KNN, DT, RF, SVM, Hard Voting, and Soft Voting classifiers are 0.633750, 0.664167, 0.644167, 0.720208, 0.719792, 0.707917, and 0.705208, respectively. RF and SVM classifiers achieved the highest accuracy around 0.72. Hard Voting and Soft Voting classifiers achieved the second highest accuracy around 0.70, which is almost close to the highest accuracy value. LR, KNN, and DT classifiers achieved lower accuracy values and almost all values are close to each other around 0.64. Secondly, the precision values for LR, KNN, DT, RF, SVM, Hard Voting, and Soft Voting classifiers are 0.631127, 0.644111, 0.645013, 0.719933, 0.722668, 0.702188, and 0.698509, respectively. The precision values of all classifiers are almost found in similar pattern to their accuracy values.

Thirdly, the recall values for LR, KNN, DT, RF, SVM, Hard Voting, and Soft Voting classifiers are 0.643750, 0.733750, 0.641250, 0.720833, 0.713333, 0.722083, and 0.722083, respectively. The KNN classifier achieved the highest recall around 0.73. The second highest recalls are found for RF, SVM, Hard Voting, and Soft Voting classifiers but they are very close to the highest recall value around 0.72. LR and DT classifiers achieved lower recall values around 0.64. Additionally, the F1-score values for LR, KNN, DT, RF, SVM, Hard Voting, and Soft Voting classifiers are 0.637376, 0.686015, 0.643126, 0.720383, 0.717970, 0.711997, and 0.710100, respectively. The RF classifier achieved the highest F1-score around 0.72. The second highest values are observed for the SVM, Hard Voting, and Soft Voting classifiers around 0.71 and these values are close to the highest one. LR, KNN and DT classifiers achieved lower F1-score values around 0.64.

Finally, the ROC-AUC values for LR, KNN, DT, RF, SVM, Hard Voting, and Soft Voting are 0.681491, 0.726476, 0.678770, 0.782857, 0.779580, N/A, and 0.770795, respectively. The RF, SVM, and Soft Voting classifiers produced almost similar values and achieved the highest ROC-AUC scores around 0.77. The KNN classifier achieved the second highest value around 0.72. The ROC-AUC value for Hard Voting is not available. Because, the ROC-AUC value requires probability estimates and Hard Voting classifier only considers the majority class votes across the base models which provides discrete outputs not probabilities [35]. LR and DT classifiers achieved lower ROC-AUC values around 0.68.

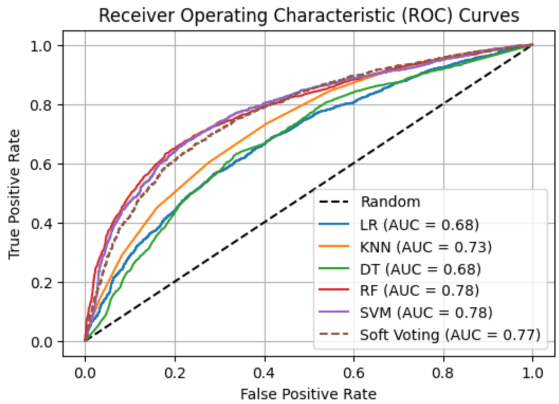


Figure 5: Statistics of ROC Curves across Different Classifiers

4.3 Best Performing Model

The statistics in Table 3 are considered to identify the best performing model across different developed classifiers. The Soft Voting classifier demonstrated better performance in almost all metrics, especially for the ROC-AUC value, which is considered a robust metric for classifiers developed on the imbalanced datasets and can accurately distinguish between positive and negative cases [36]. The RF and SVM classifiers performed well among the base classifiers, with constantly high scores across all metrics. The Hard Voting classifier showed performance close to the best one but it could not provide a ROC-AUC value due to its limitation. In contrast, the LR, KNN, and DT classifiers have lower performance scores compared to the best performing base classifiers. Based on the analysis and discussions, the Soft Voting classifier is considered to be the best performing model as it outperformed other models with respect to all evaluation metrics: Accuracy, Precision, Recall, F1-score, and particularly in the ROC-AUC value.

5. CONCLUSION AND FUTURE WORK

In this research, an optimal model for heart disease diagnosis has been developed using five mostly used machine learning algorithms. The performance of the model has been improved by applying SMOTE to handle the imbalance in the dataset and using voting classifier techniques for enhanced performance. The main findings of the result analysis show that ensemble learning techniques: Hard Voting and Soft Voting classifiers performed better than individual base models, and the Soft Voting classifier technique outperformed all models in terms of all evaluation metrics. This optimal model can be used as a decision-making tool for heart disease diagnosis. Patients can check the likelihood of heart disease in advance using that model and take necessary steps to prevent it.

However, there are some limitations in this research, and these can be addressed for further improvement of the model. Firstly, the dataset is small, and it is moderately noise free. A larger dataset may contain noise which can enhance the generalizability of the model. Secondly, the integration of deep learning models could provide better performance than traditional machine learning models. Thirdly, to ensure data security and enable model training without centralized data sharing across multiple hospitals or healthcare institutions, a federated learning approach can be considered. Finally, deploying the optimal model as a flask based web application can demonstrate its applicability in real-life heart disease diagnosis.

6. ACKNOWLEDGMENTS

The authors are thankful to all individuals for their assistance during the research period and acknowledge the importance of all materials used in this research.

7. REFERENCES

[1] World Health Organization, “Cardiovascular diseases (CVDs),”[Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). [Accessed: Jul. 31, 2025].

[2] N. E. Almansouri et al., “Early Diagnosis of Cardiovascular Diseases in the Era of Artificial Intelligence: An In-Depth Review,” *Cureus*, vol. 16, no. 3, Mar. 2024, Art. no. e55869, doi: 10.7759/cureus.55869.

[3] Y. Mao et al., “Machine learning algorithms for heart disease diagnosis: A systematic review,” *Current Problems in Cardiology*, vol. 50, no. 8, Aug. 2025, Art. no. 103082, doi: 10.1016/j.cpcardiol.2025.103082.

[4] K. Fujiwara, “Knowledge distillation with resampling for imbalanced data classification: Enhancing predictive performance and explainability stability,” *Results in Engineering*, vol. 24, Dec. 2024, Art. no. 103406, doi: 10.1016/j.rineng.2024.103406.

[5] S. Matharaarachchi, “Enhancing SMOTE for imbalanced data with abnormal minority instances,” *Machine Learning with Applications*, vol. 18, Dec. 2024, Art. no. 100597, doi: 10.1016/j.mlwa.2024.100597.

[6] Q. An, “Ensemble learning method for classification: Integrating data envelopment analysis with machine learning,” *Computers & Operations Research*, vol. 169, Sep. 2024, Art. no. 106739, doi: 10.1016/j.cor.2024.106739.

[7] D. B. Olawade et al., “Comparative analysis of machine learning models for coronary artery disease prediction with optimized feature selection,” *International Journal of Cardiology*, vol. 436, May. 2025, Art. no. 133443, doi: 10.1016/j.ijcard.2025.133443.

[8] Oktay Rdeki, “Heart Disease Dataset,” *Kaggle*, [Online]. Available:https://www.kaggle.com/datasets/oktayrdeki/heart-disease. [Accessed: Jul. 19, 2025].

[9] A. Kusiak, “Data mining and decision making,” *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 4730, pp. 155-165, May. 2002. [Online]. Available: https://www.researchgate.net/publication/228697239_Data_mining_and_decision_making. [Accessed: Jul. 19, 2025].

[10] M. H. Kabir. “Study on the Performance of Classification Algorithms for Data Mining,” *IOSR Journal of Computer Engineering (IOSR-JCE)*, vol. 21, no. 3, pp. 23-30, Jul. 2019, doi: 10.9790/0661-2103062330.

[11] A. S. M. S. Hassan et al., “Study on the Performance of Supervised Machine Learning Algorithms in Mobile Price Range Classification,” *International Journal of Computer Applications*, vol. 187, no. 1, pp. 39-45, May. 2025, doi: 10.5120/ijca2025924768.

[12] S. Mohan, et al., “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques,” *IEEE Access*, vol. 7, pp. 81542-81554, Jun. 2019, doi: 10.1109/ACCESS.2019.2923707.

[13] M. S. Alom et al., “Enhanced Heart Disease Prediction Using Advanced Classifiers and Ensemble Learning Techniques,” in *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Chattogram, Bangladesh, Feb. 13-15, 2025, doi: 10.1109/ECCE64574.2025.11013799.

[14] D. Elreedy et al., “A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning,” *Machine Learning*, vol. 113, pp. 4903-4923, Jan. 2023, doi:10.1007/s10994-022-06296-4.

[15] W. Baddah et al., “Optimizing Heart Disease Prediction Models through SMOTE: Addressing Data Imbalance,” in *2024 4th International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, Sana'a, Yemen, Aug. 6-7, 2024, doi: 10.1109/eSmarTA62850.2024.10638899.

[16] B. Warner et al., “Ensemble Learning with Highly Variable Class-Based Performance,” *Machine Learning and Knowledge Extraction*, vol. 6, no. 4, pp. 2149-2160, Sep. 2024, doi: 10.3390/make6040106.

[17] O. Hrizi et al., “Federated and ensemble learning framework with optimized feature selection for heart disease detection [J],” *AIMS Mathematics*, vol. 10, no. 3, pp. 7290-7318, Mar. 2025, doi: 10.3934/math.2025334.

[18] K. Lakshmanan and P. Gomathi, “Multi-cascaded heart disease prediction using hybrid deep learning and optimization techniques,” *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 1, pp. 1–36, Jul. 2025, doi: 10.1080/10255842.2025.2525981.

[19] L. Ren et al., “A review on missing values for main challenges and methods,” *Information Systems*, vol. 119,

Oct. 2023, Art. no. 102268, doi: 10.1016/j.is.2023.102268.

[20] M. Alwateer et al., "Missing Data Imputation: A Comprehensive Review," *Journal of Computer and Communications*, vol. 12, no. 11, pp. 53-75, Nov 2024, doi: 10.4236/jcc.2024.1211004.

[21] F. Bolikulov et al., "Effective Methods of Categorical Data Encoding for Artificial Intelligence Algorithms," *Mathematics*, vol. 12, no. 16, Aug. 2024, Art. no. 2553, doi:10.3390/math12162553.

[22] J. M. H. Pinheiro et al., "The impact of feature scaling in machine learning: Effects on regression and classification tasks," *arXiv preprint arXiv:2506.08274v3*, Jul. 2025. [Online]. Available: https://arxiv.org/abs/2506.08274.

[23] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 4, pp. 1060-1073, Apr. 2022, doi: 10.1016/j.jksuci.2019.06.012.

[24] Mahesh T R et al., "The stratified K-folds cross-validation and class-balancing methods with high-performance ensemble classifiers for breast cancer classification," *Healthcare Analytics*, vol. 4, Dec. 2023, Art. no. 100247, doi: 10.1016/j.health.2023.100247.

[25] J. Wu et al., "Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization," *Journal of Electronic Science and Technology*, vol. 17, pp. 26-40, Mar. 2019, doi: 10.11989/JEST.1674-862X.80904120.

[26] T. G. Nick, and K. M. Campbell, "Logistic Regression," *Methods in molecular biology (Clifton, N.J.)*, vol. 404, pp. 273-301, Feb. 2007, doi: 10.1007/978-1-59745-530-5_14.

[27] N. S. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, Feb. 2012, doi: 10.1080/00031305.1992.10475879.

[28] J.R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81–106, Mar 1986, doi: 10.1007/BF00116251.

[29] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 5–32, Oct 2001, doi: 10.1023/A:1010933404324.

[30] C. Cortes, and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[31] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825−2830, Nov 2011. [Online]. Available: http://jmlr.org/papers/v12/pedregosa11a.html.

[32] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *arXiv preprint arXiv: arXiv: 2010.16061v1*, Oct. 2020. [Online]. Available: https://arxiv.org/abs/2010.16061.

[33] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427-437, Jul. 2009, doi: 10.1016/j.ipm.2009.03.002.

[34] P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145-1159, Jul. 1997, doi: 10.1016/S0031-3203(96)00142-2.

[35] L. Rokach, "Ensemble-based classifiers," *Artificial Intelligence Review*, vol. 33, pp. 1–39, Nov. 2010, doi: 10.1007/s10462-009-9124-7.

[36] G. Seni and J. F. Elder, "Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions," *Morgan & Claypool*, vol. 2, no. 1, Jan. 2010, doi: 10.2200/S00240ED1V01Y201001DMK002.