High-Performance Heart Disease Prediction Through Soft Voting Ensemble: An Interpretable Machine Learning Approach

Hoda El-Batrawy
Department of Information Technology, Faculty of Computers and Artificial Intelligence, Damanhour University, Damanhour, Egypt
hodaelbatrawy@cis.dmu.edu.eg

Arshad Ali
Faculty of Computer and Information Systems, Islamic University of Madinah, Al Madinah Al Munawarah, Saudi Arabia
dr.aali7pk@gmail.com (corresponding author)

Ghulam Mustafa
London Campus, University of the West of Scotland (UWS), UK
gmg7pk@gmail.com

Gahangir Hossain
Anuradha and Vikas Sinha Department of Data Science, University of North Texas, Denton, USA
gahangir.hossain@unt.edu

Wesam Ahmed
Department of Information Technology, Faculty of Computers and Artificial Intelligence, Hurghada University, Hurghada, Egypt
wesamahmed929@yahoo.com

Received: 29 January 2026 | Revised: 23 March 2026 and 4 April 2026 | Accepted: 5 April 2026

 DOI: https://doi.org/10.48084/etasr.17838

ABSTRACT

Heart disease continues to be one of the world's leading causes of death, making the creation of precise, comprehensible prediction models necessary to support risk assessment and early diagnosis. To address this need, ensemble learning combines several basic classifiers and has shown significant promise in enhancing prediction robustness and performance. In this context, this work assesses the effectiveness and interpretability of a voting-based ensemble classifier for heart disease prediction. The dataset utilized includes comprehensive health indicators such as age, providing a robust foundation for analysis. Three heterogeneous base classifiers were employed: Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Extra Trees. These classifiers were integrated into a soft voting ensemble by aggregating posterior class probabilities to generate final predictions. As a result, the soft voting ensemble achieved superior performance, with an accuracy of 98.03%, an F1-score of 98.01%, and a Receiver Operating Characteristic–Area Under the Curve (ROC-AUC) of 98.69%, outperforming individual base classifiers. This highlights the benefits of incorporating probabilistic predictions from base classifiers. Consequently, the model offers high predictive accuracy and interpretability, enabling its potential application in clinical decision-making. These findings suggest that ensemble learning approaches, and particularly soft voting mechanisms, can support healthcare providers in early detection and personalized treatment planning for heart disease.

Keywords-heart disease; classification; ensemble learning; soft voting; Extra Trees; Support Vector Machine (SVM); Multilayer Perceptron (MLP)

I. INTRODUCTION

Heart disease remains the primary cause of death throughout the world, as it leads to approximately 9.1 million fatalities each year, which makes it the largest reason for deaths linked to cardiovascular conditions [1, 2]. The economic impact of heart disease includes direct healthcare expenses and costs for long-term treatment, together with the loss of productivity that results from the condition, which causes worldwide economic damage. Early detection of individuals who have a high risk for heart disease provides critical advantages because it enables doctors to start preventive treatments and select effective treatment methods, while also reducing the medical costs that affect both individual patients and the entire community. The main method used in traditional heart disease risk assessment determines risk through statistical models. The population-level risk stratification tools demonstrate their usefulness, but they contain multiple built-in constraints which limit their ability to work effectively in clinical settings. These existing limitations establish a requirement for advanced analytical methods which can understand complex behaviors that occur within high-dimensional clinical data.

The prediction of heart disease risk through Machine Learning (ML) methods shows better results because these methods can detect intricate connections between multiple risk factors without needing researchers to define their mathematical relationships [3-5]. The ensemble methods outperform other ML techniques through their ability to merge various base learners for producing better and more dependable output than single models [2, 3, 6]. Voting ensemble classifiers use their diverse classifier base to create more accurate results because their base learners make different mistakes on various datasets which result from their unique inductive biases and different learning methods. Soft voting extends this approach by weighting predictions according to predicted class probabilities, thereby incorporating classifier certainty into the aggregation process and theoretically yielding superior performance when base learners produce well-calibrated probability estimates [3, 6, 7]. The primary objective of this study is to develop and assess a high-performance soft voting ensemble classifier that predicts the presence of heart disease through binary prediction. Recent literature confirms the promise of ML for heart disease prediction. Authors in [8] created prediction models which can precisely identify individuals who face heart disease risk through their application of various ML methods to a structured heart disease dataset. They created multiple ML models through the development of Naïve Bayes and other algorithms. The Naïve Bayes classifier emerged as the top model because its accuracy reached 90% after testing, and performance evaluation showed excellent results.

Authors in [9] employed data from the UC Irvine Machine Learning Repository. They used multiple ML models for performance assessment, which included Random Forest to measure accuracy together with other metrics. The Random Forest model achieved 94% accuracy at k = 10, whereas it reached 92% accuracy at k = 5. The best Receiver Operating Characteristic–Area Under the Curve (ROC-AUC) results of 95% and 94%, respectively, were obtained by Random Forest, Bagged Trees, and XGBoost. Authors in [10] assessed cardiovascular risk by analyzing demographic and clinical information from a sample of 709 Ecuadorian participants. They tested eight ML algorithms, which included Decision Trees, while they used SMOTE and a hybrid ROS–SMOTE method to address class imbalance. The hybrid approach of Gradient Boosting produced the highest performance results, which included 87% accuracy, 81% precision, 74% recall, and 75% F1-score. Authors in [11] used the Classification and Regression Tree (CART) algorithm for supervised ML to predict heart disease and provide input–output decision rules. This study ranks heart disease-causing factors according to their importance. The model achieves 87% prediction accuracy, which establishes its reliability across all performance assessment metrics. The proposed method enables healthcare providers and patients who have budget and schedule limitations to diagnose and treat heart disease. Authors in [12] developed an early heart disease detection ML system which uses various feature selection techniques. They used Chi-square and mutual information to choose feature subsets, and six ML models to find the best model and feature combination. The Random Forest model performed best with SF3 feature subsets, obtaining 94.51% accuracy, 94.87% sensitivity, 94.23% specificity, and 0.31 log loss.

The previous studies achieved positive results. Despite this, they still contain several critical research gaps. First, most studies examine only single ML models, such as Naïve Bayes, Random Forest, Gradient Boosting, CART, and Logistic Regression. They do not use ensemble methods to combine classifier strengths through soft voting, which can produce more reliable and widely applicable clinical predictions. Our study develops a soft voting ensemble framework together with an upsampling method. This approach aims to improve heart disease prediction accuracy, enhance system resilience, and increase clinical usefulness. To address the identified research gaps, we defined three distinct research objectives:

1. This study offers a soft voting ensemble classifier to predict heart disease. It combines three ML algorithms. Each algorithm brings a unique strength and produces different results.
2. We assessed the performance of the soft voting ensemble and its base classifiers. We used several metrics: accuracy, F1-score, and ROC-AUC. These helped us show the statistical significance of our findings.
3. The study seeks to determine the clinical utility of heart disease risk prediction using our proposed approach for early risk assessment and clinical decision support.

II. METHODOLOGY

Figure 1 illustrates the proposed heart disease prediction system workflow, showing its operational design. The dataset preprocessing stage successfully eliminated all inconsistencies in the data. The dataset was divided into two parts, the training set and the test set. The proposed classification technique was then evaluated.

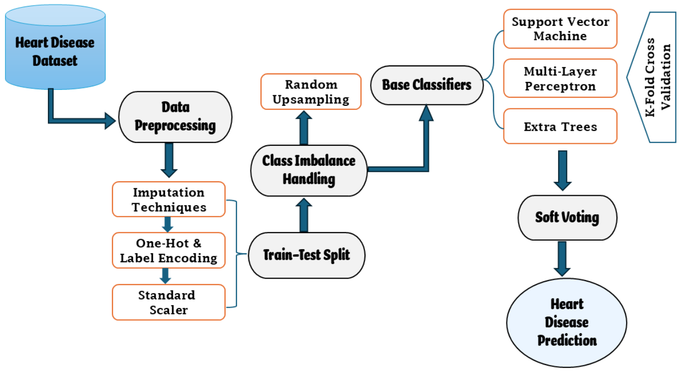

Fig. 1. Proposed heart disease prediction framework.

A. Dataset Description

Heart disease indicators and risk variables are included in the Heart Disease Data for Health Research dataset used in this study [13]. This dataset contains 10,000 entries and 21 columns: 9 are numerical variables (such as blood pressure), and 12 are categorical variables (such as smoking behavior and other classification-based features) that require encoding into numerical format. Initially, missing values in the dataset were addressed to ensure data completeness. The distribution of the target variable, shown in Figure 2, indicates a class imbalance, with the majority of samples labeled as "no disease." This imbalance may favor the majority class and reduce heart disease prediction accuracy. To mitigate this issue, random oversampling was applied, a commonly used ML data augmentation method [14]. Figure 3 presents a correlation heatmap that illustrates the relationships among input variables and the target variable for heart disease prediction.

B. Machine Learning Models

We employed a variety of ML techniques, including the Support Vector Machine (SVM), a Multilayer Perceptron (MLP), and the Extra Trees classifier, to predict heart disease [15].

1) Support Vector Machine

It is a supervised ML algorithm [16]. The algorithm searches for an optimal boundary, referred to as a hyperplane, to separate different classes in the data. It is particularly effective for binary classification tasks [17].

2) Multilayer Perceptron

It uses fully connected dense layers to transform input data across multiple representations [18]. The architecture consists of an input layer, one or more hidden layers, and an output layer. The MLP learns complex nonlinear relationships between input and output data [19].

3) Extra Trees

It is an ensemble learning model that integrates multiple decision trees using a high degree of randomization. Extra Trees generates tree-based splits by selecting random thresholds rather than choosing optimal split points. It builds each tree using the full dataset while introducing randomness in the split selection process, which helps reduce variance and overfitting [20].

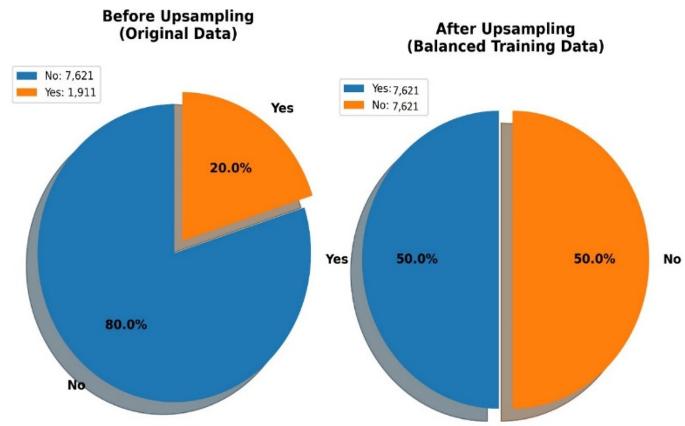

Fig. 2. Distribution before and after the upsampling technique.

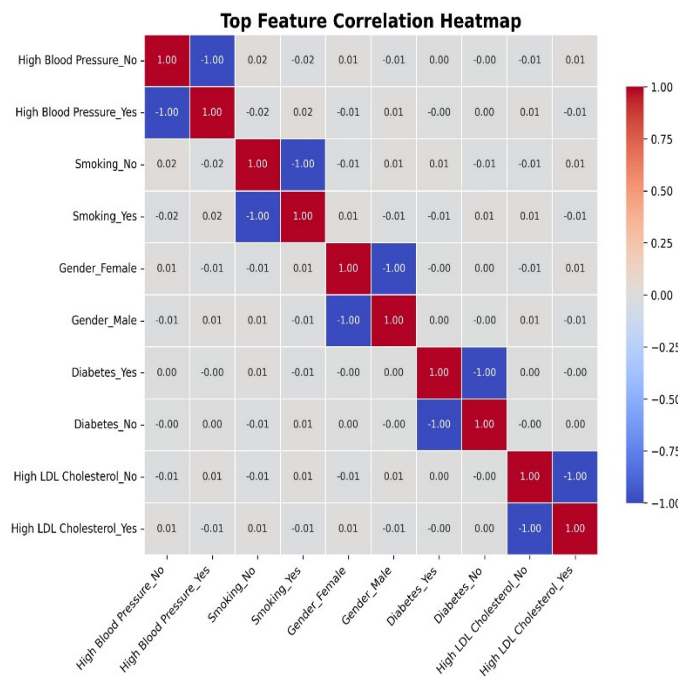

Fig. 3. Correlation heatmap of variables.

C. Proposed Model

This study uses a soft voting ensemble classifier to improve heart disease prediction accuracy through the combination of three different ML systems, including MLP, SVM and Extra Trees, as shown in Figure 4. The system aggregates class probability outputs from the individual classifiers using a soft voting ensemble method. The final prediction uses the predicted probabilities to determine the class with the highest combined confidence, which improves heart disease prediction performance through enhanced model robustness and generalization capability [21].

D. Hyperparameter Optimization of Machine Learning Models

This study performed model tuning through grid search to identify optimal parameters for each distinct model [22]. The results of hyperparameter optimization for the three ML models used in the proposed framework are presented in Table I.

TABLE I. OPTIMAL HYPERPARAMETERS OF THE ML MODELS

Model	Parameters
MLP	activation = 'relu', alpha = 0.0001, learning_rate = adaptive
Extra Trees	max_depth = 20, min_samples_leaf = 3, n_estimators = 200
SVM	svm__C = 10, svm__gamma = 0.1, svm__kernel = rbf

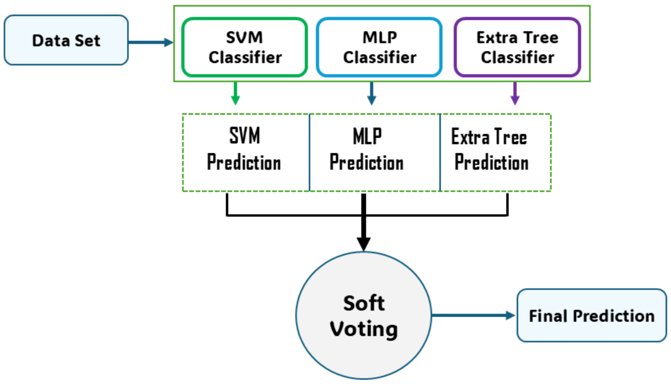

Fig. 4. Proposed soft voting ensemble model for heart disease prediction.

E. Evaluation Metrics

Multiple evaluations were conducted to assess model performance. Accuracy, precision, recall, F1-score, and ROC-AUC were employed to evaluate the classifier for all ML models. In addition, ROC curves were used for each classifier to illustrate the trade-off between the true positive rate and the false positive rate across different thresholds [22].

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

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

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

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

III. RESULTS

This section presents the complete assessment of the soft voting classification system, which operates based on an ensemble learning approach. The proposed ensemble model integrates three diverse base learners, namely SVM, MLP, and Extra Trees, combined through soft voting to effectively leverage their complementary strengths and improve overall predictive performance.

This study employed cross-validation (k = 5) to assess the efficacy of the model and mitigate the risk of overfitting. A publicly available dataset sourced from Kaggle was employed for all experiments. The dataset was divided into training (80%) and testing (20%) subsets using stratified sampling. Experiments were conducted in a controlled software environment (Python 3.10, scikit-learn 1.3) on Kaggle notebooks.

Furthermore, oversampling was applied exclusively to training folds to prevent data leakage. The total training time for the proposed model was less than five minutes. The evaluation of results was conducted using different standard classification metrics to assess diagnostic performance. The ROC curves for all models are shown in Figures 5, 6, 7, and 8 for the binary classification task. Figure 9 demonstrates how the models compare to each other. Table II shows the results of all models. Finally, the comparison of our proposed model with previous work is illustrated in Table III.

TABLE II. PERFORMANCE OF THE PROPOSED MODEL AND BASE CLASSIFIERS

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC (%)	Improvement over best base (%)
SVM	94	91.63	97.04	94.26	98.56	-
MLP	88.98	84.02	96.25	89.72	92.07	-
Extra Trees	97.34	99.18	95.47	97.29	98.68	-
Proposed model	98.03	99.19	96.85	98.01	98.69	+0.71 accuracy / +0.01 precision / +0.72 F1-score / +0.01 ROC-AUC

TABLE III. COMPARISON OF THE PROPOSED MODEL WITH PRIOR STUDIES

Study	Method	Dataset used	Accuracy (%)	F1-score (%)
[8]	Naïve Bayes	Heart disease dataset, Kaggle	90	93.20
[9]	SVM	Heart disease dataset, UC Irvine Machine Learning Repository	83	82
[11]	CART	Heart disease dataset, IEEE Dataport	87.25	86.3
Proposed model	Soft voting (SVM + MLP + Extra Trees)	Heart Disease Data for Health Research, Kaggle	98.03	98.01

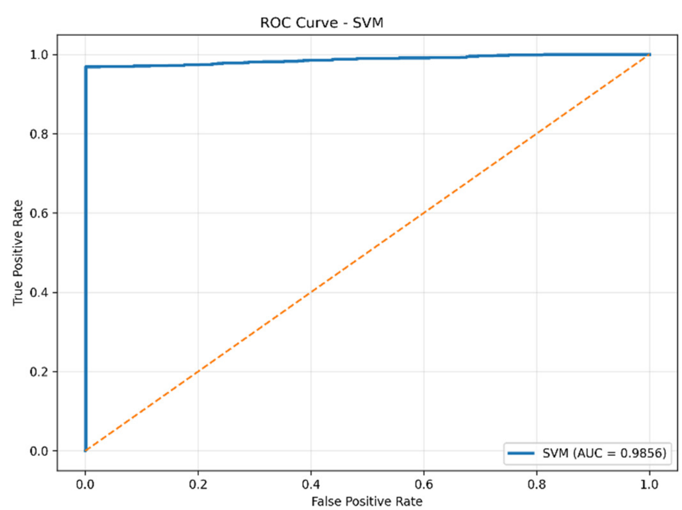

Fig. 5. ROC curve of the SVM classifier.

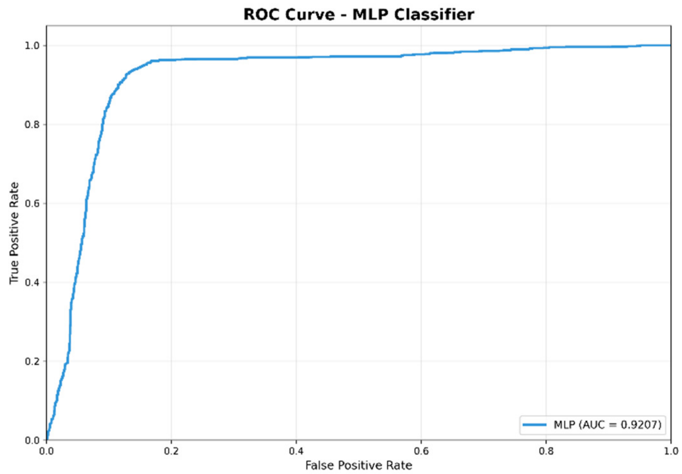

Fig. 6. ROC curve of the MLP classifier.

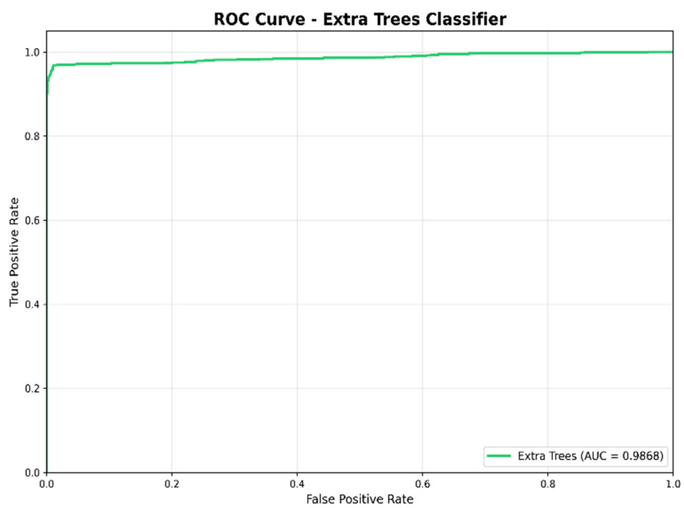

Fig. 7. ROC curve of the Extra Trees classifier.

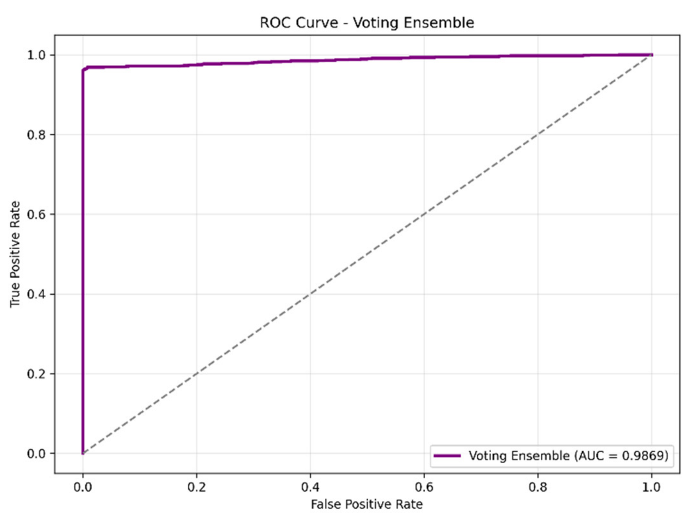

Fig. 8. ROC curve of the soft voting classifier.

IV. DISCUSSION

The proposed soft voting ensemble model outperforms all individual base classifiers across all evaluation metrics, demonstrating its effectiveness for heart disease prediction. Among standalone models, SVM achieved strong performance with 94% accuracy and 97.04% recall. MLP showed comparatively lower performance, with 88.98% accuracy and an F1-score of 89.72%. In contrast, the Extra Trees classifier exhibited superior performance among individual models, achieving an ROC-AUC of 98.68%, 99.18% precision, and 97.34% accuracy. The proposed soft voting ensemble model performed best, with 98.03% accuracy, 99.19% precision, 96.85% recall, 98.01% F1-score, and 98.69% ROC-AUC.

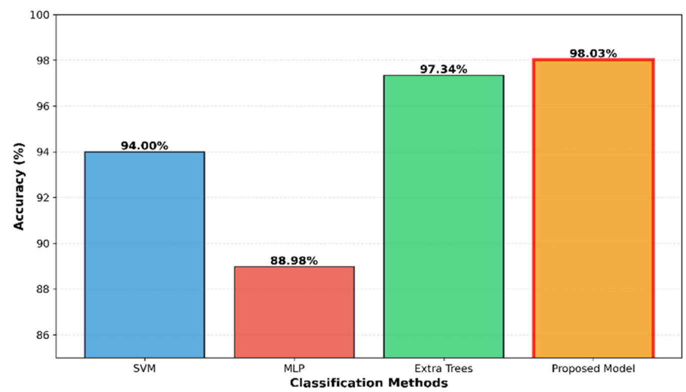

Fig. 9. Accuracy comparison of the different models.

Although the ROC-AUC values are similar, indicating comparable discriminative capability, variations in accuracy arise due to threshold dependence and sensitivity to class imbalance. For instance, MLP achieves higher recall but lower precision, reducing overall accuracy, whereas Extra Trees and the ensemble maintain a better precision–recall balance. This highlights the importance of using multiple evaluation metrics for comprehensive performance assessment.

Table III shows performance metrics from previous studies along with the performance metrics of the proposed model. Several ML approaches have been applied in the literature for heart disease prediction. For instance, authors in [8] employed the Naïve Bayes algorithm on a heart disease dataset obtained from Kaggle, achieving an accuracy of 90% and an F1-score of 93.20%. Similarly, authors in [9] utilized SVM with the widely used heart disease dataset from the UCI Machine Learning Repository, reporting an accuracy of 83% and an F1-score of 82%. In another study, authors in [11] applied the CART model on a dataset obtained from IEEE Dataport, achieving an accuracy of 87.25% and an F1-score of 86.3%.

In contrast to these individual models, our proposed approach integrates multiple classifiers (SVM, MLP, and Extra Trees) within a soft voting ensemble framework. The model was evaluated using the Heart Disease Data for Health Research dataset available on Kaggle. Experimental results demonstrate that the proposed ensemble model achieved an accuracy of 98.03% and an F1-score of 98.01%, significantly outperforming the previously reported methods. All metrics improved, indicating that the ensemble efficiently combines the complementary strengths of SVM, MLP, and Extra Trees while mitigating their individual limitations. These results demonstrate that soft voting ensembles are more accurate and clinically relevant than single classifiers, making the proposed model a promising choice for intelligent heart disease diagnosis and risk assessment systems.

V. CONCLUSION

This study developed a soft voting ensemble classifier, which achieved outstanding predictive accuracy for heart disease identification using a public dataset. The ensemble achieved an accuracy of 98.03% and an F1-score of 98.01% through the combination of Support Vector Machine (SVM), Multilayer Perceptron (MLP) and Extra Trees base learners. The proposed model shows better performance than the base classifiers, indicating that soft voting with probability-based aggregation brings benefits to binary classification of clinical data. The results demonstrate that soft voting ensemble methods provide effective medical prediction solutions that handle complex and diverse medical data.

Further research should focus on developing improved methods that enhance the interpretability of ensemble modeling techniques. The combination of local explanation methods with feature importance analysis, providing global insights, will enable better risk assessment and increase clinician confidence.

DECLARATION OF COMPETING INTERESTS

Not applicable to this work.

ACKNOWLEDGMENT

Not applicable to this work.

DATA AVAILABILITY

The dataset used in this study, Heart Disease Data for Health Research, is publicly available at [13].

REFERENCES

[1] M. Rahardi, B. P. Asaddulloh, A. Aminuddin, F. F. Abdulloh, I. Saifudin, and F. P. Kusumawijaya, "Optimizing Machine Learning Models for Class Imbalance in Heart Disease Prediction," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 23599–23604, June 2025, https://doi.org/10.48084/etasr.10407.

[2] V. V. R. Karna, V. R. Karna, V. Janamala, V. N. K. R. Devana, V. R. S. Ch, and A. B. Tummala, "A Comprehensive Review on Heart Disease Risk Prediction using Machine Learning and Deep Learning Algorithms," *Archives of Computational Methods in Engineering*, vol. 32, no. 3, pp. 1763–1795, Apr. 2025, https://doi.org/10.1007/s11831-024-10194-4.

[3] A. Dhankhar, S. Juneja, A. Juneja, and V. Bali, "Kernel Parameter Tuning to Tweak the Performance of Classifiers for Identification of Heart Diseases," *International Journal of E-Health and Medical Communications*, vol. 12, no. 4, pp. 1–16, 2021, https://doi.org/10.4018/IJEHMC.20210701.oa1.

[4] S. Parsa, P. Shah, R. Doijad, and F. Rodriguez, "Artificial Intelligence in Ischemic Heart Disease Prevention," *Current Cardiology Reports*, vol. 27, no. 1, Feb. 2025, Art. no. 44, https://doi.org/10.1007/s11886-025-02203-0.

[5] M. Ferdowsi, C.-H. Goh, H. Liu, G. Tse, J. M. Ho Hui, and X. Wang, "Clinical Application of Artificial Intelligence in the Diagnosis, Prediction, and Classification of Coronary Heart Disease," *Cardiovascular Innovations and Applications*, vol. 10, no. 1, Mar. 2025, Art. no. 976, https://doi.org/10.15212/CVIA.2025.0009.

[6] A. Arya, M. Sehgal, N. Bhatia, S. Juneja, and D. Koundal, "Heart disease prediction with machine learning and virtual reality: from future perspective," in *Extended Reality for Healthcare Systems*, S. Khan, M. Alam, S. A. Banday, and M. S. Usta, Eds. Cambridge, MA, USA: Academic Press, 2023, pp. 209–228, https://doi.org/10.1016/B978-0-323-98381-5.00011-8.

[7] R. Saini *et al.*, "Firefly algorithm and DNN for improved contactless heart rate measurement from videos," *Scientific Reports*, vol. 16, no. 1, Jan. 2026, Art. no. 2778, https://doi.org/10.1038/s41598-025-32633-3.

[8] A. R. Ilyas, S. Javaid, and I. L. Kharisma, "Heart Disease Prediction Using ML," *Engineering Proceedings*, vol. 107, no. 1, Oct. 2025, Art. no. 124, https://doi.org/10.3390/engproc2025107124.

[9] M. D. Teja and G. M. Rayalu, "Optimizing heart disease diagnosis with advanced machine learning models: a comparison of predictive performance," *BMC Cardiovascular Disorders*, vol. 25, no. 1, Mar. 2025, Art. no. 212, https://doi.org/10.1186/s12872-025-04627-6.

[10] R. Reátegui, C. Tandazo-Malla, R. Suárez, and L. Ramírez-Cerna, "Cardiovascular risk prediction via ensemble machine learning and oversampling methods," *Scientific Reports*, vol. 15, no. 1, Dec. 2025, Art. no. 43576, https://doi.org/10.1038/s41598-025-30895-5.

[11] M. Ozcan and S. Peker, "A classification and regression tree algorithm for heart disease modeling and prediction," *Healthcare Analytics*, vol. 3, Nov. 2023, Art. no. 100130, https://doi.org/10.1016/j.health.2022.100130.

[12] N. Biswas *et al.*, "Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques," *BioMed Research International*, vol. 2023, no. 1, May 2023, Art. no. 6864343, https://doi.org/10.1155/2023/6864343.

[13] "Heart Disease Data for Health Research." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/oktayrdeki/heart-disease.

[14] M. H. N. Le *et al.*, "Protective predictors of cardiovascular disease: an explainable AI approach," *Public Health*, vol. 250, Jan. 2026, Art. no. 106050, https://doi.org/10.1016/j.puhe.2025.106050.

[15] A. Kumar, S. Pal, A. Singh, and A. P. Singh, "Comparative study of supervised machine learning techniques in heart disease prediction: A review," *AIP Conference Proceedings*, vol. 3224, no. 1, Feb. 2025, Art. no. 020052, https://doi.org/10.1063/5.0247130.

[16] H. A. Al-Alshaikh *et al.*, "Comprehensive evaluation and performance analysis of machine learning in heart disease prediction," *Scientific Reports*, vol. 14, no. 1, Apr. 2024, Art. no. 7819, https://doi.org/10.1038/s41598-024-58489-7.

[17] M. Imani, A. Beikmohammadi, and H. R. Arabnia, "Comprehensive Analysis of Random Forest and XGBoost Performance with SMOTE, ADASYN, and GNUS Under Varying Imbalance Levels," *Technologies*, vol. 13, no. 3, Feb. 2025, Art. no. 88, https://doi.org/10.3390/technologies13030088.

[18] J. Beinecke and D. Heider, "Gaussian noise up-sampling is better suited than SMOTE and ADASYN for clinical decision making," *BioData Mining*, vol. 14, no. 1, Nov. 2021, Art. no. 49, https://doi.org/10.1186/s13040-021-00283-6.

[19] M. Ahmadi, M. Khashei, and N. Bakhtiarvand, "Enhancing air quality classification using a novel discrete learning-based multilayer perceptron model (DMLP)," *International Journal of Environmental Science and Technology*, vol. 22, no. 5, pp. 3051–3062, Mar. 2025, https://doi.org/10.1007/s13762-024-06017-5.

[20] A. M. Elshewey, E. Selem, and A. H. Abed, "Improved CKD classification based on explainable artificial intelligence with extra trees and BBFS," *Scientific Reports*, vol. 15, no. 1, May 2025, Art. no. 17861, https://doi.org/10.1038/s41598-025-02355-7.

[21] Z. Abbas, S. Kim, N. Lee, S. A. W. Kazmi, and S. W. Lee, "A robust ensemble framework for anticancer peptide classification using multi-model voting approach," *Computers in Biology and Medicine*, vol. 188, Apr. 2025, Art. no. 109750, https://doi.org/10.1016/j.compbiomed.2025.109750.

[22] H. Khoshvaght, R. R. Permala, A. Razmjou, and M. Khiadani, "A critical review on selecting performance evaluation metrics for supervised machine learning models in wastewater quality prediction," *Journal of Environmental Chemical Engineering*, vol. 13, no. 6, Dec. 2025, Art. no. 119675, https://doi.org/10.1016/j.jece.2025.119675.