International Journal of Innovations & Research Analysis (IJIRA)
(An International Quarterly Multidisciplinary Peer Reviewed Refereed Journal)
ISSN: 2583-0295 (Online), Impact Factor: (Cosmos 7.662 & I2OR 5.105), Volume 06, No. 03(II), July-September, 2026, pp 85-96

An Optuna-Optimized Dynamic Stacked Ensemble for Cardiovascular Disease Prediction

Sunanda Budihal[1*] | Dr. Sheetalrani R Kawale[2]

[1]Research Scholar, Department of Computer Science, Karnatak State Akkamahadevi Women's University, Vijayapura, Karnataka, India.
[2]Assistant Professor, Department of Computer Science, Karnatak State Akkamahadevi Women's University, Vijayapura, Karnataka, India.

[*]Corresponding Author: sunanda.kiran2003@gmail.com

Citation: Budihal, S. & Kawale, S. (2026). An Optuna-Optimized Dynamic Stacked Ensemble for Cardiovascular Disease Prediction. International Journal of Innovations & Research Analysis, 06(03(II)), 85–96. https://doi.org/10.62823/IJIRA/06.3(II).950

Abstract

As cardiovascular disease (CVD) has become a growing burden, the need for strong predictive models that can aid in early diagnosis and clinical decision making has become increasingly urgent. To overcome this challenge, in this work, a real world clinical dataset is obtained from Ayush Multi-speciality Hospital and Research Centre (AMSHRC), Vijayapur, karnataka (India) and a dynamic stacked ensemble model is proposed for prediction of cardiovascular diseases. The proposed approach combines the strengths of multiple machine learning models through adaptive ensemble learning to achieve accurate and reliable risk prediction. The framework uses dynamic top three model selection, hard-sample instability-aware weighting, stratified fivefold cross validation and Optuna-based Bayesian hyperparameter optimization to boost the predictive power and model robustness. Four state-of-the-art machine learning algorithms Xgboost, Catboost, Extra Trees and Random forest are fine-tuned and evaluated, and the top three models are joined via stacked ensemble framework. Experimental evaluation demonstrates that the proposed stacked ensemble model outperformed the individual baseline models by achieving 98.0% accuracy, 97.4 % F1-score, 99.6% ROC-AUC and 95.8% MCC on the AMSHRC clinical dataset. These findings indicate that the developed dynamic stacked ensemble model is an effective and robust approach for cardiovascular disease prediction and has the potential to support early clinical decision-making.

***Keywords**: Cardiovascular Disease Prediction (CVD), Dynamic Stacked Ensemble, Bayesian Hyperparameter Optimization, Clinical Decision Support System, Machine Learning.*

Introduction

Cardiovascular disease (CVD) is a major global health concern due to its high incidence and mortality [1], highlighting the demand for accurate and reliable early risk assessment methods. Timely and accurate prediction of cardiovascular disease is an important step for early diagnosis and effective therapeutic management.

However, accurate prediction remains difficult because of the complex interaction of demographic, clinical and physiologic factors. It is thus a promising way to intelligently predict cardiovascular diseases. With the growing availability of electronic health data, machine learning approaches such as Random Forest, ET, XGBoost and CATBoost have been utilized to predict cardiovascular diseases [3], [16-18], [20]. The approach to stacked ensemble learning utilizes the meta-learning technique to combine multiple classifiers producing better prediction capability and improved

[*] *Copyright © 2026 by Author's and Licensed by Inspira. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work properly cited.*

accuracy and robustness than individual models [5], [6]. However, several existing ensemble techniques employ static classifier ensembles without adaptive model selection and instability learning, which might.

This paper proposes a dynamic weighted stacked ensemble framework to improve prediction accuracy and model robustness using a real-world clinical dataset to address the limitations of existing methods for cardiovascular disease prediction. The proposed system consists of several synergistic components, including adaptive selection of the three best performing base learners, instability-sensitive weighting of difficult-to-classify clinical examples, Bayesian hyperparameter tuning with Optuna[14],[15], and stacked ensemble approaches[5],[6]. The framework aims to enhance the learning ability of the individual models and the collective robustness of the ensemble by integrating these strategies. The adaptive weighting approach identifies the clinically challenging cases, enabling the framework to identify the complex disease patterns.

In addition, Optuna-based optimization discovers effective hyperparameter settings for each base learner, thus enhancing predictive performance[14]. The extensive experimental analysis on AMSHRC clinical dataset shows that the proposed framework outperforms the individual machine learning models consistently, thus proving the reliability of the framework for cardiovascular disease prediction.

The arrangement of this manuscript is as follows: Section II provides a review of major academic literature on prediction of cardiovascular disease. In Section III we introduce the suggested technique and detail the general structure. Section IV presents the empirical evaluation and the comparative analysis of the results obtained. Finally, Section V concludes the main findings of this study and discusses potential future works.

Literature Review

Recent development in machine learning has brought considerable improvements in predicting heart disease based on demographic, physiological and clinical data. The recent review studies have also emphasized that the deep learning-based models are being used for the prediction of heart disease in the recent years [11]. Certain initial studies have found out that conventional machine learning models, when accompanied by proper preprocessing and feature engineering, can be used in detecting diseases. Biswas et al. [2] reported that Random Forest with feature selection provided the highest accuracy of 94.51%. Similar to the aforementioned researches, Rimal et al. [3] considered a number of classification algorithms like Logistic Regression, Support Vector Machine, K-Nearest Neighbors and Random Forest and discovered that Random Forest had the best performance with accuracy of 94%. Moreover, Teja and Rayalu [4] have proven the effectiveness of gradient boosting approach where XGBoost produced accuracy **of 93%** with selected clinical features.

Ensemble learning and explainable artificial intelligence methods have gradually been employed to improve prediction capabilities in recent studies.. El-Sofany et al. [7] enhanced the utility of heart disease prediction models in clinical decision support by integrating explainable AI with machine learning to boost their transparency and trustworthiness. The advantages of ensemble methods over single classifiers have been shown by Rohan et al. [8] and they have conducted a comprehensive evaluation of artificial intelligence approaches and achieved an accuracy of 97.3% using an ensemble-based prediction framework. Tiwari et al. [23] also performed a study on an ensemble framework for cardiovascular disease prediction, which showed the effectiveness of combining several learning models for better prediction performance. In addition, Qadeer et al. [9] showed that the accurate prediction of heart failure and better diagnostic performance are achieved thru the successful capturing of complex nonlinear relationships in clinical data by advanced machine learning and deep learning models.

Most studies so far have focused on encouraging predictive performance, but they are mostly based on the use of traditional optimization approaches and have fixed ensemble architecture, unable to deal with the changes in model behavior in different clinical datasets. Moreover, the dynamic selection of the most suitable base learners and the assignment of adaptive weights to the difficult clinical samples in ensemble construction have not been adequately tackled. We propose a dynamic weighted stacked ensemble framework to address the limitations above by employing the stacked ensemble learning, Optuna-based Bayesian hyperparameter optimization, instability-aware hard-sample weighting, and dynamic Top-three model selection. The proposed framework is to improve the accuracy of cardiovascular disease predictions, robustness of the model, and generalizability of the model on the real-world clinical dataset.

Methodology

Figure 1 illustrates the conceptual workflow of the developed system for predicting cardiovascular disease. The methodology consists of consecutive steps of clinical data acquisition and preprocessing, feature engineering, hard-sample instability weighting and class balancing, Bayesian hyperparameter optimization and five-fold stratified cross-validation. First, we consider four optimized tree-based classifiers XGBoost, CatBoost, Random Forest and Extra Trees as candidate base learners. The best three classifiers are dynamically chosen according to the validation performance and used for ensemble construction. The out-of-fold prediction probabilities are converted to meta-features, and fed to an optimized Level-1 meta-learner for the final classification of cardiovascular disease. The framework aggregates complementary model predictions to enhance the robustness and reliability of cardiovascular disease prediction.

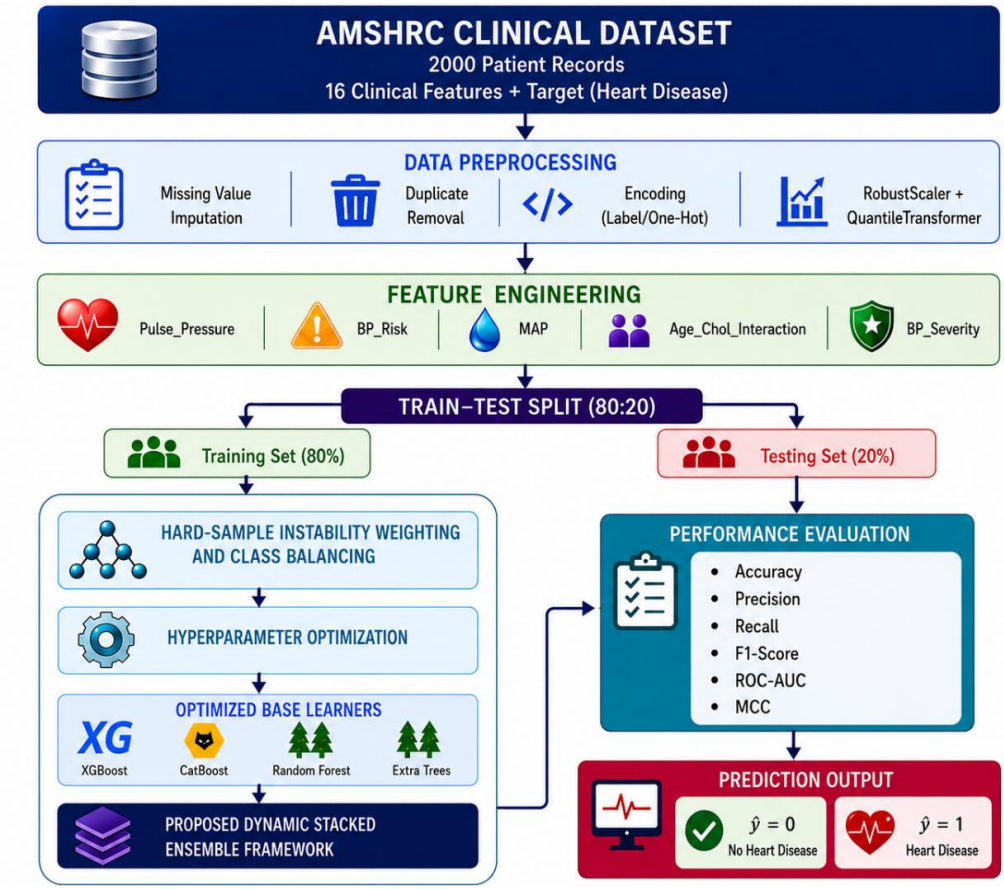

Figure 1: Workflow of the proposed system for predicting cardiovascular disease

- **Description of Data Set**

The experiments were performed using clinical records obtained from Ayush Multi Specialty Hospital and Research Centre (AMSHRC), Vijayapur, India. The data set includes 2,000 anonymized patient records collected retrospectively from individuals ranging in age from 20 to 95 years old, and consists of 779 cardiovascular disease and 1,221 normal cases.The clinical dataset used in this study has also been employed in our previous research on cardiovascular disease prediction using conventional machine learning approaches [10],[19]. The aim of the present work is different. Here, the focus is on evaluating a dynamic weighted stacked ensemble that integrates adaptive model selection, instability-aware sample weighting. Using an identical clinical dataset allows the methodological improvements introduced in this study to be assessed without the influence of changes in the underlying data. Only anonymized records were included in the analysis. No information capable of identifying

individual patients was available during data processing or model development. The study was carried out in accordance with the ethical guidelines governing the use of retrospective clinical data.

- **Preprocessing**

Prior to model training, the clinical dataset was preprocessed to improve data quality and ensure consistency. Duplicate patient records were eliminated to retain a single record for each individual. Missing values were denoted by symbols like "?" or "NA" or empty values and were detected and handled accordingly. Continuous clinical attributes, including age, blood pressure, pulse rate(PR), oxygen saturation(spo2), body weight, and cholesterol level, were converted into numerical format to facilitate machine learning analysis. Median imputation was applied to missing numerical values to minimize the impact of outliers on the data distribution., where as categorical attributes were used the most frequently occurring category (mode). One-Hot Encoding was applied to the categorical variables. Then, class balancing was used [12], [13] to reduce the effect of class imbalance on the learning of the model. To prevent data leakage and ensure an unbiased evaluation, all preprocessing steps were fitted exclusively on the training data, and the learned transformations were subsequently applied to the test set.

- **Feature Engineering**

Feature engineering was performed to enrich the clinical dataset by generating additional variables from the original clinical attributes. The derived features were designed to represent clinically meaningful relationships among independent risk factors associated with cardiovascular disease. By incorporating these interactions into the learning process, the dataset became more informative, enabling the model to identify complex patterns and improve predictive accuracy and robustness .

Pulse Pressure, an significant cardiovascular indicator allied with arterial stiffness, was figured as the variance between systolic and diastolic blood pressure:

$$Pulse\ Pressure = BP_{systolic} - BP_{diastolic}$$

As the mean of the overall arterial pressure across a cardiac cycle the Mean Arterial Pressure (MAP) was calculated as:

$$MAP = BP_{diastolic} + \frac{(BP_{systolic} - BP_{diastolic})}{3}$$

Therefore, an indicator of blood pressure risk, based on measured blood pressure (systolic) and hypertension status: was developed.

$$BP_{Risk} = BP_{systolic} \times Hypertension$$

Similarly, but also the joint effects of interactions among the clinical variables were produced;

$$Age_Chol_Interaction = Age \times Cholesterol$$

$$BP_Severity = BP_{systolic} \times Age$$

The engineered feature set provides better predictive insights that help in classifying cardiovascular diseases by integrating clinically relevant associations among physiological variables.To explore the association between the clinical variables prior to building the model, a correlation analysis was conducted.

A correlation matrix based on the samples generated helped to determine which attributes had high correlation content and aided to focus on the contribution of engineered attributes to data structure. These are further graphed in figure 2.

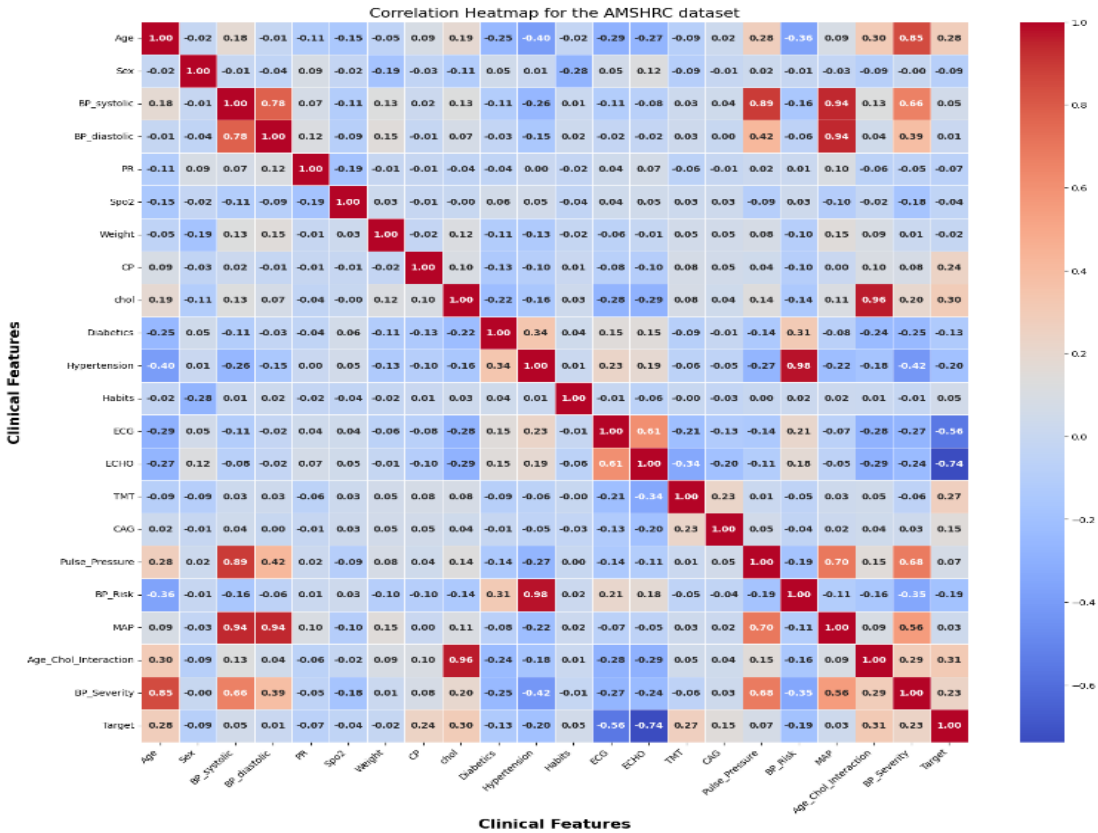

Fig. 2: Heatmap of the correlations between demographic, clinical, diagnostic, and engineered variables applied to cardiovascular disease prediction on the AMSHRC data set

- **Hard Sample Instability Weighting strategy**

A hard-sample instability weighting strategy was adopted to improve the learning from equivocal cardiovascular cases. To estimate prediction variability, Random Forest models were trained with several random seeds and standard deviation of prediction probabilities for each training instance was calculated. Difficult cases were identified and assigned increased training weights for samples with instability scores at or above the upper quartile (Q75). Then, the weighted samples were used for hyperparameter tuning and stacked ensemble training, which enabled the models to focus on uncertain cases better while preserving the original data distribution.

- **Class Imbalance Management and Weighted Class Balancing**

In medical data, the imbalance in the classes is a universal issue as there are more samples without disease than those with the disease. The class imbalance may skew the performance of classifiers to the majority class samples[12, 13].

Therefore, techniques to address class imbalance were incorporated into the developed framework. In case of the XGBoost model, the value of the scale_pos_weight parameter was computed based on the ratio between negative and positive samples:

$$scale_pos_weight = \frac{N_{negative}}{N_{positive}}$$

where N_{negative} and N_{positive} are the number of negative and positive instances in the training data set, respectively. Moreover, balanced class weight strategy was applied CatBoost to improve the learning capacity of the minority- class instances.

- **Hyperparameter Optimization**

To determine suitable model configurations, hyper-parameters were optimised using the Optuna optimisation framework [14]. To find appropriate configurations, stratified cross-validation was used. Predefined search spaces for the number of estimators, tree depth, learning rate, sampling ratios and minimum samples for splitting were created according to the specifications of each classifier.

- **Machine Learning Models**

The proposed stacked ensemble with complementary tree-based classifiers captures the heterogeneous pattern of the clinical data set. XGBoost, CatBoost, Random Forest and Extra Trees were selected as candidate base learners, as they apply different learning and tree construction strategies. This diversity enables the ensemble to learn non-linear relationships between demographic, physiological, lifestyle and diagnostic features, and mitigates the reliance on a single classifier in order to promote robustness of the prediction. The complementary predictions are then used to guide the next meta learning stage.

- **XGBoost**

XGBoost [16] was selected as a potential classifier due to its excellent performance on structured clinical datasets. Hyperparameters of the model are optimized by means of Optuna to achieve high performance with a low degree of overfitting. Then, the classifier is ranked according to its Accuracy, MCC, and ROC-AUC. Prediction probabilities of XGBoost are used for meta-features in case it is selected among the best-performing classifiers.

The objective function of XGBoost is given by:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

where $\ell(y_i, \hat{y}_i)$ $is\ \ loss\ function\ for\ the\ i-th\ sample$, the regularization function on the k-th base learner is $\Omega(f_k)$ to decrease the overfitting and limit the model

- **CatBoost**

CatBoost[18] was employed to evaluate its capability for learning complex interactions among clinical variables. Following Bayesian hyperparameter optimization, its performance was compared with the remaining classifiers. If selected during the dynamic ranking stage, CatBoost contributed prediction probabilities to the meta-learning layer of the proposed framework.

CatBoost's prediction function can be expressed as:

$$\hat{y}_i = \sum_{m=1}^{M} f_m(\mathrm{x}_i)$$

where $f_m(\mathrm{x}_i)$ represents the decision tree in the mth iteration, the feature vector xi represents the input features vector, M represents the number of boosting iterations and $\hat{y}_i$represents the predicted output for the ith patient sample.

The training

- **Random Forest**

Random Forest [17] is an example of ensemble learning in which multiple DTs are created and the final result is the prediction mode of the classification trees (or the average of the regression trees) of the individual trees. More powerful than a single DT as it reduces overfitting and averages multiple trees. The equation below is used to measure the RF output.

$$\hat{y} = majority_{vote}\left(T_1(x), T_2(x) \ldots T_n(x)\right)$$

where $T_i(x)$denotes the prediction of the i-th decision tree, and $\hat{y}$denotes the final predicted output obtained through majority voting

- **Extra Trees**

The Extra Tree Classifier algorithm[20] is an algorithm of classification task-specific decision tree ensemble. It's split selection at each node is different from a traditional decision tree. Extra Tree

Classifier randomly chooses the split and not the best possible split out of all possible splits. This stochasticity leads to the robustness of the model and reduces overfitting.

The final prediction of the Extra Trees classifier is based on the majority voting of the randomly built decision trees:

$$\hat{y} = \text{mode}\{T_m(x)\}_{m=1}^{M}$$

where $T_m(x)$ is the prediction of the m-th randomized decision tree, Mis the number of the total trees and $\hat{y}$ the final predicted class

- **Proposed Stacked Ensemble Approach**

To leverage the synergetic prediction power of improved classifiers , a two-level dynamic stacked ensemble architecture was developed[5],[6]. At the first step four algorithms were evaluated as Level-0 foundational learners: XGBoost[16], CatBoost[18], Random Forest[17] and Extra Trees[20]. To select the three most appropriate base learners, a dynamic selection process was conducted resulting in the selection of CatBoost, Extra Trees and XGBoost, while Random Forest was excluded due to its relatively poor predictive performance. Hard-sample instability weighting and class balancing were integrated during model training to better deal with difficult and imbalanced samples. We generated out-of-fold probability predictions in the meta-feature generation process using five-fold stratified cross-validation to prevent information leakage [22]. The meta-feature representation was constructed from the probability outputs of the dynamically selected base learners, their mean, standard deviation and the specified clinical attributes. We also considered potential meta-learners at Level-1 like XGBoost and Extra Trees with their hyperparameters optimized using Optuna-based Bayesian optimization and stratified cross-validation [14],[15]. The meta-learner with the greatest ROC-AUC was used for the final prediction. The same preprocessing and feature engineering pipeline was used on unseen patient records at inference time.. The selected level-0 models produced probability estimates that were transformed into meta-features and fed to the improved level-1 learner to produce the final cardiovascular disease classification.

- **Performance Evaluation Metrics**

The performance of the proposed framework was compared with the standard classification metrics derived from the confusion matrix. The evaluation included Accuracy (ACC), Precision (PREC), Recall (REC), F1-score, and the Receiver Operating Characteristic (ROC) curve with the corresponding Area Under the Curve (AUC) [21]. These measures together form a well-rounded evaluation of the model's classification accuracy, reliability, sensitivity and discrimination ability. The mathematical formulations of the evaluation metrics are presented in as shown below.

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

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

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

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

where TP, TN, FP, and FN denote the numbers of true positives, true negatives, false positives, and false negatives, respectively. The ROC curve shows the relationship between the true positive rate (TPR) and false positive rate (FPR) for various thresholds, while the AUC provides a measure of the overall performance of the classifier.

- **Stratified Cross-Validation Strategy**

To obtain a reliable estimate of the proposed framework's predictive performance, a stratified 5 cross-validation strategy was employed [22]. The dataset was randomly divided into five stratified folds where the distribution of cardiovascular disease and non-disease cases remains the same in each fold. During each iteration, four folds were used for training and the remaining fold was used for validation. This procedure was repeated until each fold had served as the validation set once. The entire process

was then repeated five times using different random data partitions to improve the robustness and stability of the performance estimates.

- **Experimental Setup**

All computational procedures were executed within the Python environment using Scikit-learn and the XGBoost, CatBoost and Optuna software libraries. Once data was prepared, the patient records were partitioned into a training and independent evaluation set in a stratified manner to preserve the class proportions seen in the original data set.. Models were built using Bayesian tuning along with stratified cross-validation. The efficiency of the developed method was assessed using classification performance evaluation as discussed in next Section

Results and Discussion

- **General Performance Analysis**

Table 4 and Figure 3 compares the predictive capability of the evaluated classifiers. The proposed dynamic stacked ensemble achieved the best overall performance across all major evaluation metrics, including 98.00% accuracy, 98.7% precision, 97.4% F1-score, 95.8%MCC, and 99.6 %ROC–AUC. Among the standalone models, XGBoost , catboost and Extra Trees showed the best performance, with accuracies of 96.8%,97.50%, and 97.00%, respectively.

However, the proposed system proved to be superior to all the baseline systems by successfully fusing the best of several learners in a stacked ensemble system.

Table 4: Predictive Performance of Individual and Ensemble Models

Model	Acc. (%)	Precision (%)	Recall (%)	F1 (%)	AUC (%)	MCC (%)
XGB	96.8	98.6	92.9	95.7	99.6	93.2
CAT	97.5	98.7	94.9	96.7	99.6	94.8
RF	96.5	98.0	92.9	95.4	99.5	92.7
ET	97.0	98.0	94.2	96.1	99.6	93.7
Stacked	**98.0**	**98.7**	**96.2**	**97.4**	**99.6**	**95.8**

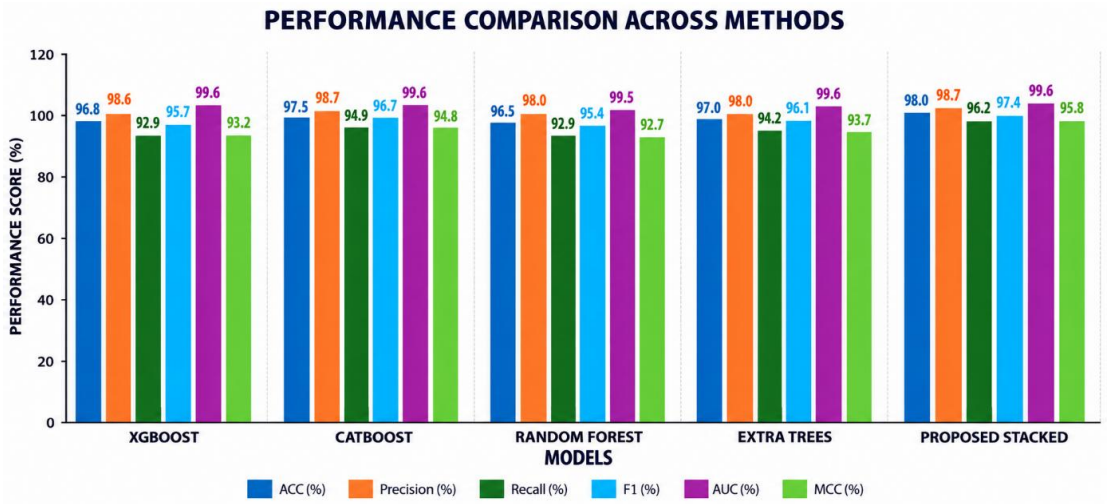

Figure 3: Predictive Performance of Individual and Stack Models.

- **Training and Testing Performance of the Proposed STACKED Model**

The training and testing accuracy of the STACKED model proposed were analyzed by using an 80:20 split of train-test, where there were 1,600 training samples and 400 independent testing samples. The training accuracy was found to be 97.38% and testing accuracy was 98.00% as illustrated in Figure. 4. The close resemblance of training and testing accuracies shows that the proposed model has consistent prediction accuracy for new data as well without any considerable difference between training and testing.

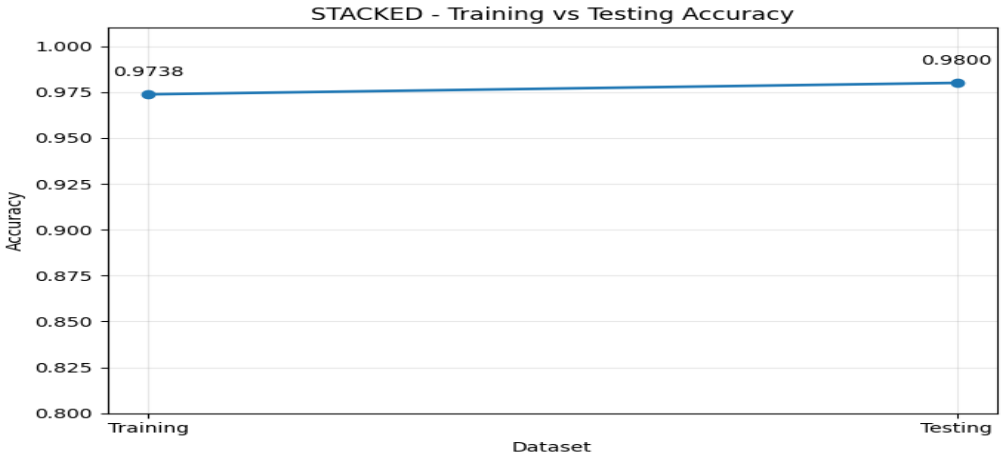

Figure 4 : The training and testing accuracy of the proposed STACKED model

- **Confusion Matrix and ROC–PR Curve Analysis**

The confusion matrices of the proposed dynamic stacked ensemble framework and baseline classifiers are provided in Figure 5. As can be seen from the figure, the predictive framework performed the best in terms of balancing sensitivity and specificity, as it correctly predicted 242 non-diseased cases and 150 diseased cases with merely 2 false positives and 6 false negatives. From the clinical point of view, reducing the number of false negative forecasts is vital since failure to detect the cardiovascular disease can lead to the postponement of its diagnostics and treatment, which in turn results in possible negative effects on patients. Meanwhile, a low number of false positives will help reduce the number of unnecessary tests and related expenditures. Consequently, the current work showed better classification performance and was able to provide accurate predictions on cardiovascular disease diagnostics.

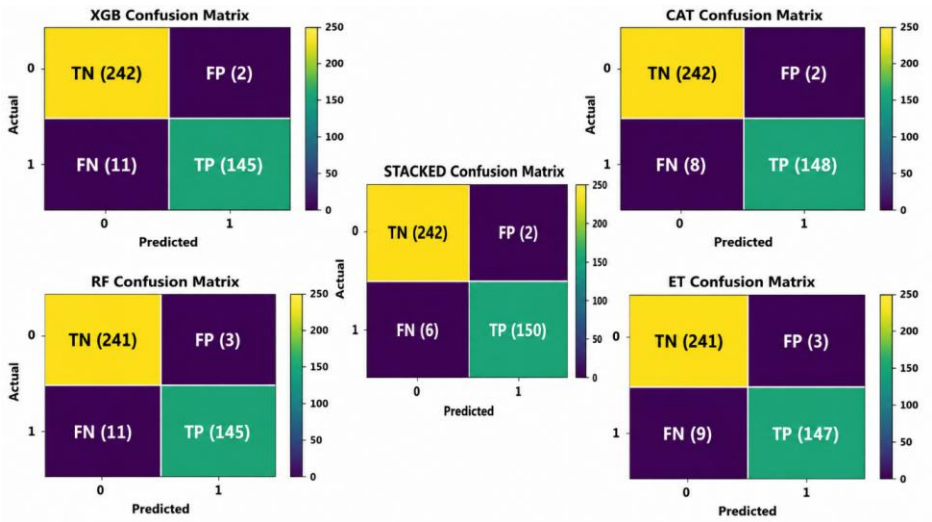

Figure 5: Confusion matrices of the base line classifiers and the proposed dynamic stacked ensemble framework.

Figure 6 presents the ROC curves for the proposed dynamic weighted stacked ensemble and the baseline models. . It is apparent that all models provide high discriminative capacity, with AUC scores varying from 0.9955 to 0.9964. The Extra Trees method provides the best AUC score equal to 0.9964, whereas the proposed stacked model shows the AUC of 0.9963, followed by XGBoost and CatBoost with 0.9962 AUC scores. These results demonstrate its effectiveness and reliability for cardiovascular disease prediction on the AMSHRC clinical dataset.

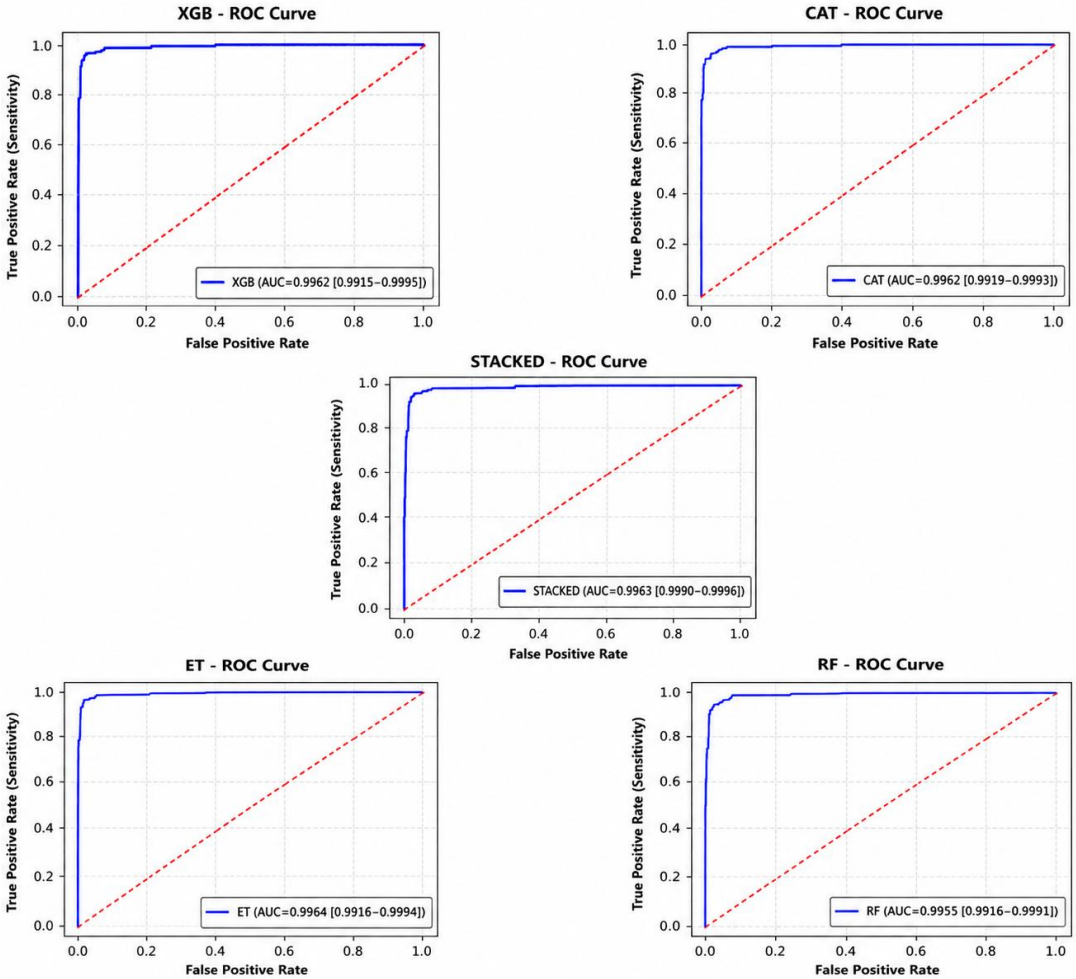

Figure 6: ROC curves of the baseline classifiers and the proposed dynamic stacked ensemble framework.

- **Comparative Discussion with Existing and Proposed Studies**

Figure 7 provides the comparison between the proposed framework and existing research works for prediction of cardiovascular diseases. Existing studies have obtained various accuracy results using various datasets, where El-Sofany et al. [7] reported 97.57% accuracy using a combined Cleveland and private heart-disease dataset, Tiwari et al. [23] achieved 92.34% accuracy using the IEEE DataPort Heart Disease Dataset.In comparison, the proposed dynamic weighted stacked ensemble framework achieved an accuracy of 98.00% and an F1-score of 97.40% when evaluated on the real-world AMSHRC clinical dataset. These results demonstrate the effectiveness of the proposed strategy. However, the comparison studies should be considered with careful consideration, as each study is based on different datasets, preprocessing techniques, feature engineering methods, model settings, and evaluation strategies. Overall, the results demonstrate that the proposed framework offers competitive and reliable predictive performance, suggesting that it is a promising approach for cardiovascular disease prediction using real-world clinical data.

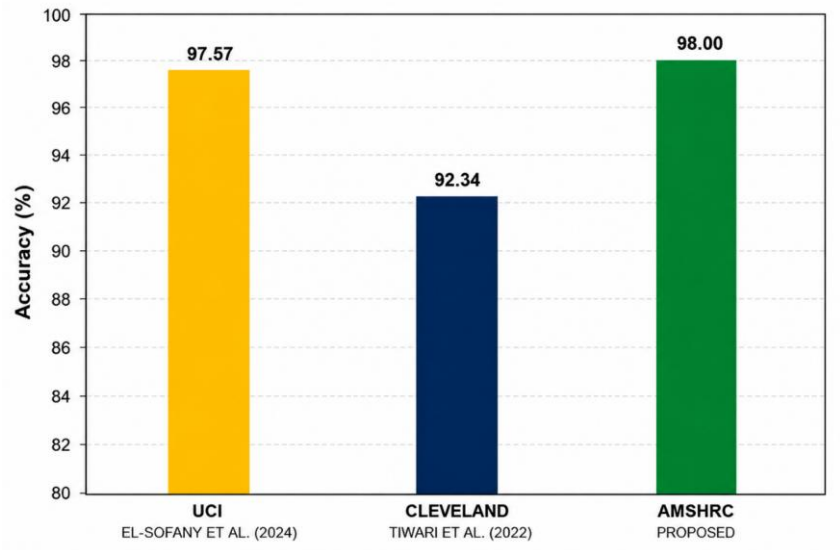

Fig. 7: Comparison of the proposed framework with existing cardiovascular disease prediction approaches based on accuracy.

Discussion

The proposed dynamic stacked ensemble model achieved 98.0% accuracy, 97.4% F1-score, 99.6% ROC-AUC, and 95.8% MCC, indicating its superior predictive performance in comparison to the individual machine learning models. The proposed framework enhanced accuracy to 98.0%, while CatBoost achieved the highest baseline accuracy of 97.50%. The results show that the ensemble strategy is able to effectively fuse the prediction ability of the base learners selected. The framework is limited by the computational expenses of model optimization and ensemble training, and the use of retrospective data from one healthcare institution. The framework's generalizability and robustness will be further evaluated thru multicenter validation and multimodal clinical data in future research.

Conclusion

A dynamic weighted stacked ensemble framework was developed to improve cardiovascular disease prediction using a real-world clinical dataset collected from AMSHRC . Our dynamic stacked ensemble framework employs instability-aware hard-sample weighting, Optuna-based Bayesian hyperparameter optimization, dynamic Top-3 model selection, and stacked ensemble learning. Experimental evaluation demonstrated that the proposed model consistently outperformed the individual machine learning models, achieving 98.00% accuracy, 97.4% F1-score, 99.6% ROC-AUC and 95.8% MCC.The proposed STACKED model had 97.38% training accuracy and 98.00% testing accuracy in the 80-20 split between the training set and the test set, demonstrating reliable performance across unseen clinical datasets. The results show that the proposed framework is feasible, reliable and efficient approach for predicting cardiovascular disease and can be used to clinical decision making with intelligence. For future study, the framework will be tested with a larger multi-centre clinical dataset, multimodal patient data added, and more advanced explainable artificial intelligence (XAI) techniques explored to improve the clinical utility and transparency of the model.

References

1. World Health Organization, “Cardiovascular diseases (CVDs),” World Health Organization, Jul. 31, 2025.
2. N. Biswas, V. Kuppili, D. R. Edla, S. Lipare, A. K. Singh, and S. Badhoni, “Machine learning-based model to predict heart disease in early stage employing different feature selection techniques,” *Biomed. Res. Int.*, vol. 2023, Art. no. 6864343, 2023, doi: 10.1155/2023/6864343.
3. Y. Rimal, A. Sharma, S. Pokharel, R. Rijal, P. Khanal, and B. Kandel, “Comparative analysis of heart disease prediction using logistic regression, SVM, KNN, and random forest with cross-validation,” *Sci. Rep.*, vol. 15, Art. no. 13444, 2025, doi: 10.1038/s41598-025-93675-1.

4. M. D. Teja and G. M. Rayalu, “Optimizing heart disease diagnosis with advanced machine learning models: A comparison of predictive performance,” *BMC Cardiovasc. Disord.*, vol. 25, Art. no. 212, 2025, doi: 10.1186/s12872-025-04627-6.
5. D. H. Wolpert, “Stacked generalization,” *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992, doi: 10.1016/S0893-6080(05)80023-1.
6. O. Sagi and L. Rokach, “Ensemble learning: A survey,” *WIREs Data Mining Knowl. Discov.*, vol. 8, no. 4, Art. no. e1249, 2018, doi: 10.1002/widm.1249.
7. H. El-Sofany, B. Bouallegue, and Y. M. Abd El-Latif, “A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method,” *Sci. Rep.*, vol. 14, Art. no. 23277, 2024, doi: 10.1038/s41598-024-74656-2.
8. D. Rohan, G. P. Reddy, Y. V. P. Kumar, K. P. Prakash, and Ch. P. Reddy, “An extensive experimental analysis for heart disease prediction using artificial intelligence techniques,” *Sci. Rep.*, vol. 15, Art. no. 6132, 2025, doi: 10.1038/s41598-025-90530-1.
9. M. Qadeer, R. Ayaz, and M. I. Thohir, “Heart failure prediction through a comparative study of machine learning and deep learning models,” *Eng. Proc.*, vol. 107, Art. no. 61, 2025, doi: 10.3390/engproc2025107061.
10. S. Budihal, S. R. Kawale, and N. Agarwal, “A machine learning framework for heart disease prediction with feature selection and hyperparameter optimization across benchmark and real-world datasets,” *Cureus J. Comput. Sci.*, vol. 3, no. 1, 2026, doi: 10.7759/s44389-026-00150-7.
11. C. Zhou *et al.*, “A comprehensive review of deep learning-based models for heart disease prediction,” *Artif. Intell. Rev.*, vol. 57, Art. no. 263, 2024, doi: 10.1007/s10462-024-10899-9.
12. H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.
13. B. Krawczyk, “Learning from imbalanced data: Open challenges and future directions,” *Prog. Artif. Intell.*, vol. 5, no. 4, pp. 221–232, 2016, doi: 10.1007/s13748-016-0094-0.
14. T. Akiba, S. Sano, T. Yanase, T. Ohta, and T. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Anchorage, AK, USA, 2019, pp. 2623–2631, doi: 10.1145/3292500.3330701.
15. J. Snoek, H. Larochelle, and R. P. Adams, “Practical Bayesian optimization of machine learning algorithms,” in *Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 2951–2959.
16. T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
17. L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
18. Hamid M, Rahman A, Islam M. Fine-tuned CatBoost machine learning approach for early detection of cardiovascular disease through predictive model IEEE Access, 2025, 13: 99876-99888. Doi : 10.1038/s41598-025
19. S. Budihal, S. R. Kawale, N. Agarwal, and B. Halalli, "Feature-group-aware cardiovascular disease prediction using gated neural fusion and CatBoost-based clinical feature analysis," Cureus J. Comput. Sci., vol. 3, no. 1, 2026, doi: 10.7759/s44389-026-00068-0.
20. P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees,” *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, 2006, doi: 10.1007/s10994-006-6226-1.
21. D. M. W. Powers, “Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation,” *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.
22. R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *Proc. 14th Int. Joint Conf. Artif. Intell.*, Montreal, QC, Canada, 1995, pp. 1137–1145.
23. A. Tiwari, A. Chugh, and A. Sharma, “Ensemble framework for cardiovascular disease prediction,” *Comput. Biol. Med.*, vol. 146, Art. no. 105624, 2022, doi: 10.1016/j.compbiomed.2022.105624.

❑❑❑