Proceedings of Engineering and Technology Innovation, vol. 30, 2025, pp. 66-78

Efficient Model for Early Prediction of Heart Disease Using Ensemble Technique

Ankush Hutke*, Jyoti Deshmukh

Department of Computer Engineering, MCT's Rajiv Gandhi Institute of Technology, University of Mumbai, Maharashtra, India

Received 25 January 2025; received in revised form 17 April 2025; accepted 29 April 2025

DOI: https://doi.org/10.46604/peti.2024.14787

Abstract

The growing global burden of cardiovascular diseases has created an urgent need for advanced early-detection devices that revolutionize preventive cardiology. This research presents a novel two-stage ensemble (TSE) learning framework that outperforms traditional machine learning methods by integrating multiple complex algorithms, including random forest, adaptive boosting, gradient boosting machine, light gradient boosting machine, and extra trees classifier, into a highly accurate predictive system in stage 1. The approach incorporates a sophisticated preprocessing pipeline with feature scaling and the synthetic minority oversampling technique SMOTE to address the class imbalance and ensure robust input data quality. The model optimizes a meta-learner for enhanced predictions by leveraging meta-features derived from various classifiers. The developed TSE model, utilizing the CatBoost classifier in stage 2, achieved average accuracies of 92.5% and 90.19% on the Cleveland and Statlog datasets, respectively. This comprehensive ensemble framework significantly advances clinical decision support for early detection and intervention in cardiovascular disease.

Keywords: heart disease prediction, ensemble learning, SMOTE, CatBoost classifier

1. Introduction

One of the most urgent medical issues of the twenty-first century is the widespread impact of cardiovascular disease on global public health, which calls for innovative early identification and intervention methods. With an estimated 17.9 million fatalities yearly and over 31% of all global deaths, cardiovascular diseases remain the primary cause of death globally, according to the comprehensive global health assessment published by the World Health Organization. This alarming statistic underscores the critical need for developing advanced early detection technologies capable of identifying potential heart disease cases before they progress to more severe stages. Although useful, traditional diagnostic techniques often rely heavily on a combination of clinical testing, medical imaging procedures, and physician expertise. However, they can be time-consuming, resource-intensive, and occasionally subject to interpretive variability. Emerging sophisticated machine learning technologies have opened new horizons in medical diagnosis by providing the possibility for more accurate, objective diagnostic tools that may complement current diagnostic procedures.

Artificial intelligence and machine learning technologies are increasingly integrated into the healthcare industry have sparked a revolution in patient care and medical diagnostics [1]. Particularly suitable for complex medical diagnosis applications, these cutting-edge computational techniques have remarkable capabilities in pattern recognition, risk prediction, and data analysis [2]. The field of disease diagnosis and prognosis has been profoundly transformed by machine learning

* Corresponding author. E-mail address: ankush.hutke@mctrgit.ac.in

algorithms' ability to process and analyze large patient datasets, detect subtle patterns and connections, and deliver accurate forecasts [3]. In medical applications, recent advances in ensemble learning techniques, which integrate multiple machine learning methods to create reliable and accurate prediction systems, have shown great promise. However, existing methods face challenges such as selecting and integrating base models, determining feature importance for cardiovascular risk factors, and balancing computational efficiency with prediction accuracy, limiting their effectiveness in clinical settings.

Along with notable progress in understanding cardiovascular disease pathophysiology and risk factors, machine-learning technologies have evolved rapidly [4]. This confluence of medical knowledge and computational capabilities has opened unprecedented opportunities for the development of more accurate and complex diagnostic instruments. Although useful, conventional machine learning methods may suffer from constraints like sensitivity to noise in medical data, difficulty handling missing values, and challenges incorporating domain knowledge into the modeling process [5]. Through the combination of many complementary models, ensemble learning techniques have become a potent answer to these challenges, providing enhanced robustness, accuracy, and reliability. Still, the effective use of ensemble learning in medical diagnosis requires a comprehensive evaluation of various factors, including model selection, feature engineering, and clinical domain expertise integration.

The goal of this article is to offer a thorough framework that solves the difficulties related to cardiovascular disease prediction. This study introduces numerous novel components that collectively advance the field of medical diagnosis systems using a thorough examination of current literature and rigorous evaluation of clinical needs [6]. The main objective of the work is to build a sophisticated ensemble learning architecture that ensures interpretability and clinical relevance in addition to optimizing prediction accuracy. These are the important aspects often overlooked in current methods.

This research contributes the following significant insights:

(1) Developing a robust two-stage stacking ensemble framework to enhance the prediction of cardiovascular disease, leveraging meta-features generated by a diverse set of base classifiers and a highly optimized meta-learner to improve predictive accuracy.
(2) Incorporating a comprehensive data preprocessing pipeline, including the synthetic minority oversampling technique (SMOTE) for addressing class imbalance and feature scaling, ensuring the reliability and quality of the input data for effective learning.
(3) Integrating an efficient hyperparameter optimization framework using grid search cross-validation (GridSearchCV) for the meta-learner, balancing computational efficiency with high prediction performance, making it adaptable to real-world scenarios.
(4) Providing a rigorous evaluation pipeline with cross-validation (CV) and classification metrics, guaranteeing the model's robustness and applicability across diverse patient groups and clinical environments.

2. Related Work

The application of data mining and machine learning methods for predicting heart disease has progressed substantially in recent years, focusing primarily on early diagnosis and prevention. This progress has been made possible by several important studies in the field [7]. This article reviews these research methods, conclusions, and limitations. A study by Bhatt et al. [8] in 2023 used several machine learning methods, including eXtreme Gradient Boosting (XGBoost), random forest, decision trees, and multilayer perceptron (MLP). There were about 59,000 rows and 11 attributes in their dataset. With a CV accuracy of 87.28%, the MLP algorithm produced the best outcomes in terms of area under the ROC curve, F1-score, precision, and recall.

Similar to this, Ogundepo and Yahya [9] used two separate datasets to perform an estimated analysis of risk factors for heart disease in 2023. Data from the Cleveland dataset was used in the training of the classification models and their effectiveness was evaluated using the Statlog dataset, providing a robust validation of their performance. They achieved notable performance metrics, including 85% accuracy and a log loss of 38%. Using exploratory analysis, they discovered notable relationships between certain clinical factors and the occurrence of heart disease. Notably, the support vector machine (SVM) model outperformed the other nine models for classification that were trained using the Cleveland dataset, suggesting its potential as a valuable tool for predicting and preventing heart disease. To ensure the reliability of their findings, the researchers employed a 10-fold CV approach on the Statlog dataset, which revealed consistent performance across the various models.

Table 1 Summary of Literature Survey

Author	Methodology	Dataset	Algorithms	Results
Bhatt et al.	Applied various machine learning algorithms for heart disease prediction.	~59,000 rows, 11 attributes	Random Forest, Decision Tree, MLP, XGBoost	MLP: 87.28% Accuracy, with Good Recall, Precision, AUC, and F1-Score.
Ogundepo & Yahya	Predictive analysis using the Cleveland and Statlog datasets. Explored relationships between cardiac disease and bio-clinical variables.	Cleveland and Statlog datasets	SVM, 10 Classification Models	SVM: Accuracy of 85%, Specificity of 88%, Precision of 87%, Log Loss of 38%. Consistent Results on the Statlog Dataset
M. Zeng	Analyzed the dataset of heart disease patients with various machine learning algorithms.	Randomly selected patient dataset	Naïve Bayes, Decision Tree, Random Forest, SVM, and Logistic Regression	Random Forest: 85.01% Accuracy, Best-Performing Algorithm.
A. Khan et al.	Examined patient data from cardiac illness using several machine-learning techniques.	Randomly selected patient dataset	Naïve Bayes, Decision Tree, Random Forest, SVM, and Logistic Regression	Random Forest: 85.01% Accuracy, Best-Performing Algorithm.
Shah et al.	Used the Cleveland dataset using KNN, naïve Bayes, decision tree, and Random Forest models.	Cleveland dataset	Decision Tree, Naïve Bayes, KNN, Random Forest	KNN: 86.885% Accuracy, Best-Performing Algorithm.
A. Garg et al.	Conducted a comparison between KNN and Random Forest models using the Kaggle Heart Disease dataset.	Kaggle heart disease dataset	KNN, Random Forest	KNN: 86.885% Accuracy, Higher Accuracy Than Random Forest.
V. Shorewala	Performed feature selection using LASSO for analysis and utilized conventional classifiers for modeling.	Not specified	Decision Trees, Neural Networks, LASSO (feature selection), Dense Neural Network	Dense Neural Network: 73.9% Accuracy, F1-Score 72.0%.
G. Ahamad et. al.	Used six ML algorithms to predict heart disease with GridSearchCV for hyperparameter tuning.	UCI Kaggle Cleveland, Hungary, Switzerland, Long Beach V	Logistic Regression, KNN, SVM, Decision Tree, Random Forest, XGBoost	Logistic Regression:85.71%, KNN: 87.91%, SVM: 84.62%, Decision Tree: 81.32%, Random Forest: 84.62%, and XGBoost: 79.12%.

Another study by Zeng [10] from 2023 used 11 clinical features to predict cardiac disease using decision trees, k-Nearest Neighbors (KNN), SVM, and XGBoost algorithms. The SVM algorithm fared better than the other models, obtaining 90.7% F1-score, 88.8% accuracy, and 89.3% recall. The study's focus was on highlighting the significance of exercise-induced ST depression and angina as key characteristics linked to heart disease. Furthermore, Khan et al. [11] used a variety of machine learning methods, such as logistic regression, random forest, naïve Bayes, SVM, and decision trees, to analyze a dataset of randomly chosen heart disease patients. The random forest algorithm beat the other algorithms and showed the potential of machine learning for the analysis and prediction of cardiovascular disease with a maximum accuracy of 85.01%.

Machine learning approaches have also been used in other studies to predict cardiac disease. Shah et al. [12] used the Cleveland dataset using KNN, naïve Bayes, decision tree, and random forest models. They discovered that KNN produced the highest accurate prediction, with an accuracy of 86.885%. The Kaggle heart disease dataset was used in a study by Garg et al. [13] to evaluate the effectiveness of the KNN and random forest methods. The results indicated that KNN outperformed random forest, achieving an accuracy of 86.885%. V. A feature analysis was performed by Shorewala [14] to forecast cardiac disease, using the LASSO algorithm for selecting significant features. Numerous conventional classifiers, including neural networks and 92 decision trees, were used; the dense neural network performed the best, obtaining an F1-score of 72.0% and a 93% testing accuracy of 73.9%. Ahamad et al. [15] used six ML algorithms to predict heart disease with GridSearchCV to adjust hyperparameters. A summary of the different methods used by various researchers is given in Table 1.

Ensemble learning has been widely used for disease prediction, with stacking demonstrating superior performance over other ensemble techniques. Comparative studies have shown that multi-level stacking outperforms classical stacking, bagging, and boosting, highlighting its effectiveness in reducing bias and variance while enhancing predictive accuracy. This motivates the present study's focus on a two-stage stacking ensemble for enhanced disease prediction [16-17].

3. Materials and Methods

The effectiveness of any machine learning-based diagnostic model strongly depends on the quality, diversity, and relevance of the data used during training and evaluation. To ensure robustness and broad applicability, this study begins with careful selection of clinically relevant datasets, followed by comprehensive preprocessing steps. These foundational stages are critical for developing a reliable and generalizable predictive model.

3.1 Dataset Description and Preprocessing

This work guarantees strong validation and generalizability of the suggested method by using many extensive datasets. Renowned for their thoroughness and clinical importance, the Cleveland Heart Disease Dataset and Statlog Dataset are the main data sources. Comprising 303 patient records with 14 unique features, the Cleveland dataset includes a broad spectrum of cardiovascular risk factors and clinical tests. Comprising 294 patient records with a comparable attribute structure, the Statlog dataset offers significant geographical variation and helps fully validate the model's generalizability across several patient groups.

This study adopts a thoughtfully crafted preprocessing approach that significantly enhances data quality and model reliability. Missing values are handled using a hybrid imputation method that blends clinical expertise with statistical techniques, ensuring both accuracy and relevance. The process also includes effective outlier detection and normalization to maintain consistency across features. A hybrid feature selection algorithm [18] is then applied, helping identify the most informative variables for heart disease prediction. As a result, both the model's performance and interpretability are improved. The final selected features are: 'cp', 'trestbps', 'thalach', 'ca', and 'thal'.

3.2 Proposed Ensemble Architecture

Incorporating many new components that together improve prediction accuracy and model interpretability, the proposed ensemble architecture marks major progress in medical diagnosis systems. The proposed methodology outlines a two-stage ensemble (TSE) learning framework for classification tasks, utilizing multiple machine learning models as base learners and a meta-learner to aggregate their predictions for improved accuracy and generalization. Fig. 1 presents a flowchart diagram depicting the sequential steps for predicting the likelihood of heart disease. It focuses on leveraging ensemble learning algorithms and refining the model through hyperparameter tuning.

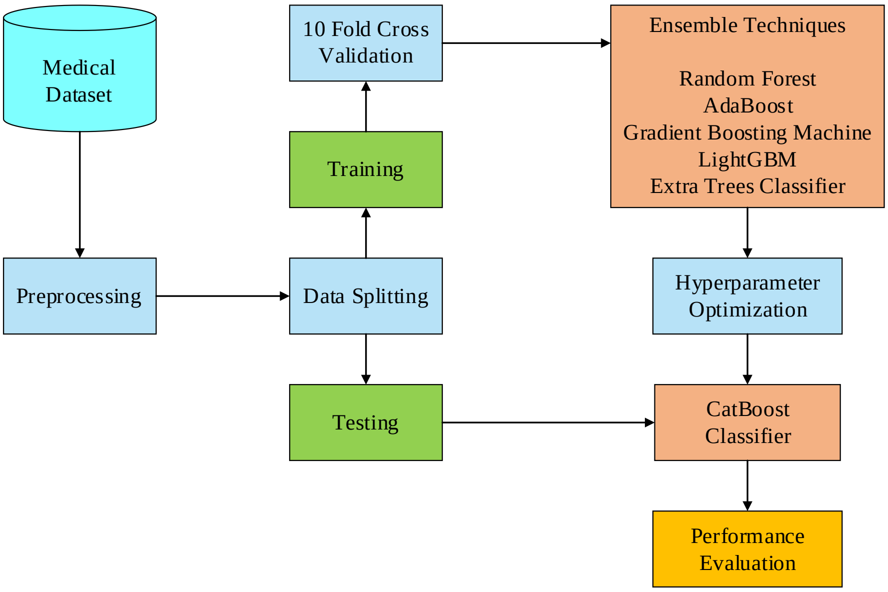

Fig. 1 Detailed Algorithm Flowchart for Heart Disease Prediction Ensemble Model

3.2.1. Stage 1: Base Models Training

To build a robust and accurate predictive system, we employ a diverse set of ensemble-based machine learning algorithms as base classifiers. The selected models—random forest, AdaBoost, gradient boosting machine, LightGBM, and extra trees classifiers—represent a variety of ensemble strategies designed to enhance generalization and reduce model bias. Each algorithm's unique mechanisms offer complementary strengths, allowing the system to capture complex data patterns more effectively. The following sections will provide a detailed overview of each algorithm, including configuration choices and performance outcomes.

(1) Random Forest Classifier

An ensemble learning method called the random forest algorithm combines several decision trees to create a strong predictive model. Recursively dividing the data according to the most informative features produces each decision tree in the random forest. The algorithm evaluates qualities at each tree node using metrics such as entropy or Gini impurity and selects the attribute that provides the largest reduction in uncertainty.

The formula for the random forest prediction can be computed by

$$\hat{y} = \frac{1}{n_{trees}} \sum_{i=1}^{n_{trees}} f_i(x) \tag{1}$$

where $f_i(x)$ is the prediction of the i-th tree.

(2) Adaptive Boosting (AdaBoost) Classifier

AdaBoost is a powerful machine learning meta technique used to enhance performance. It is a powerful technique for improving the performance of machine learning models. By combining the outputs of multiple weak learners, boosting algorithms can create a strong and accurate classifier. AdaBoost is considered adaptive because each successive weak learner focuses more on cases misclassified by previous learners. AdaBoost's strength lies in its ability to leverage the collective wisdom of many weak learners. While individual learners may not be highly accurate, their combined power, guided by AdaBoost's weighting scheme, leads to a final classifier that is surprisingly accurate and robust. This approach can sometimes be less prone to overfitting than other methods. AdaBoost combines weak classifiers iteratively to create a strong classifier by adjusting the weights of misclassified instances.

The weight of misclassified instances can be expressed as

$$w_i^{(t+1)} = w_i^t * \exp(\alpha_t * 1(y_i \neq \hat{y})) \tag{2}$$

where

$$\alpha_t = (1/2) * \ln[(1 - e_t) / e_t] \tag{3}$$

e_t is the error rate of the t^{th} weak classifier, and $1(y_i \neq \hat{y})$ is the indicator function, which equals 1 if $y_i \neq \hat{y}$ and 0 otherwise.

(3) Gradient Boosting Machine

It is an effective machine learning method that combines several weak learners (such as decision trees) to create an ensemble of strong learners that can accurately predict outcomes. Gradient boosting machine is an iterative and adaptive boosting algorithm that uses gradient descent to reduce the loss function. With each iteration, the model refines and improves on its previous prediction. Additionally, it is a powerful algorithm for solving regression, classification, and ranking problems. It is also popular for its flexibility, scalability, and robustness. Gradient boosting machine builds an additive model by sequentially fitting weak models to the residual errors of the previous models.

The residual errors of the previous models can be represented by

$$r_i^{(m)} = y_i - F^{(m-1)}(x_i) \tag{4}$$

where $F^{(m-1)}(x_i)$ is the prediction of the ensemble up to iteration m−1.

The update rule can be calculated by

$$F^{(m)}(x) = F^{(m-1)}(x) + \eta \cdot h_m(x) - F^{(m-1)}(x_i) \tag{5}$$

where $h_m(x)$ is the m^{th} weak learner and η is the learning rate.

Loss Minimization: Optimizes a differentiable loss function $L(y, \hat{y}(X_i))$ using gradient descent.

(4) Light Gradient Boosting Machine (LightGBM)

It is a high-performance gradient boosting framework designed for speed and efficiency. It is based on the principles of gradient boosting but introduces novel enhancements, such as histogram-based learning and leaf-wise tree growth, which significantly accelerate training and improve model performance. Unlike traditional boosting methods that grow trees level-wise, LightGBM grows trees leaf-wise, enabling it to capture complex patterns more efficiently.

(5) Extra Trees Classifier

It is an ensemble learning method that aggregates multiple decision trees to form a powerful predictive model. Unlike traditional methods, it adds a layer of randomization during tree construction by randomly selecting splitting thresholds for each feature. This enhances diversity among the trees and helps to reduce overfitting. Predictions are made through majority voting or averaging, similar to a random forest. This approach makes the extra trees classifier particularly effective for managing noisy data, outliers, and high-dimensional feature spaces.

3.2.2. Meta-Features Generation

Each base model is trained on resampled training data. Predictions are collected from all base models on both the training and testing datasets. These predictions are used to create a new feature space (meta-features), where each column represents the predictions of a base model. These predictions serve as inputs (meta-features) for the next stage, where the meta-learner will use them to make the final predictions. This intermediate representation allows the meta-learner to understand how different base models behave across instances and to learn higher-level patterns. This layered learning approach enhances generalization and contributes to improved overall performance of the ensemble model.

3.2.3. Stage 2: Meta-Learner

Meta-Learner Selection: The meta-learner aggregates the predictions from the base models to make the final classification. In this study, a CatBoost Classifier is employed for its exceptional performance and ability to handle categorical data efficiently. A standout feature of CatBoost is its automatic processing of categorical variables, avoiding the requirement for manual encoding. It combines one-hot encoding with target encoding techniques to manage categorical features with high cardinality effectively.

Algorithm 1 TSE (D, X, Y)

```
Input: - Training data D, with features X and labels Y
   - Selected features F
   - Base classifiers h1, h2, ..., hm
   - Meta-classifier g
   - Hyperparameter grid for g
         - Test data D_test, with features X_test and labels Y_test
Output: Final ensemble model f
Begin
// Preprocessing Phase
1.   Select features F from X.
2.   Scale features in X and X_test using StandardScaler.
3.   Address class imbalance in D using SMOTE, creating D_resampled with features X_resampled and labels
     Y_resampled.
// Level 0 Training Phase
For i = 1 to m:
    Train hi on D_resampled.
    For each (x, y) in D_resampled:
        Append hi(x) to M_train.
    For each X_test in X_test:
            Append h_i(X_test) to M_test
            // Level 1 Training Phase
Perform hyperparameter tuning for g using GridSearchCV on (M_train, Y_resampled).
Train the best g on (M_train, Y_resampled).
// Ensemble Model Prediction
For each new input x:
      For i = 1 to m:
          Append hi(x) to the meta-feature vector m_x.
             y_ensemble = g(m_x).
Return y_ensemble as the predicted output.
End
```

Hyperparameter Tuning: Hyperparameter tuning is performed on the meta-learner using GridSearchCV with a 10-fold CV to identify the optimal combination of parameters. The parameters tuned include the depth of the tree, the number of iterations, the learning rate, L2 regularization, and random state.

CV: The tuned meta-learner undergoes further validation using 10-fold CV, and accuracy scores are recorded for each fold to ensure generalization. The pseudocode of the TSE is shown in Algorithm 1.

3.3 Implementation Details and Technical Architecture

The suggested ensemble architecture is a complex combination of many software components and algorithms, meticulously coordinated to guarantee the best performance and dependability. Python 3.8 is used in the core system, which makes use of powerful scientific computing libraries like NumPy 1.21.0 for efficient numerical computations and Pandas 1.3.0 for data manipulation and preprocessing.

The architecture incorporates scikit-learn 0.24.2 for implementing base classifiers, random forest, LightGBM, AdaBoost, gradient boosting, and extra trees classifier, along with utility functions like StandardScaler for feature scaling and GridSearchCV for hyperparameter optimization. To address the class imbalance, the SMOTE algorithm from the imbalanced-learn library (version 0.8.0) is used, ensuring an equal distribution of target classes in the training data. CatBoost 1.0.0 serves as the meta-classifier in the second stage of the ensemble, offering robust handling of the meta-features generated by the base classifiers. The entire implementation is validated on a cardiovascular disease dataset, with a two-stage stacking approach that effectively combines the advantages of several models to achieve superior predictive performance and generalizability.

To further assess the feasibility of the proposed ensemble model for real-world applications, computational performance metrics are evaluated. The training time for both the base models and the meta-learner remains within a reasonable range, with the meta-learner requiring more time due to hyperparameter optimization. The model demonstrates moderate CPU and memory usage, indicating that it can be executed on standard computing hardware without significant resource demands. Additionally, GPU memory consumption is minimal, showing that the approach does not heavily rely on specialized GPU acceleration. These results confirm the computational practicality of the proposed model, supporting its potential deployment in medical settings without the need for high-end infrastructure.

The novelty of the proposed TSE lies in its carefully designed combination of diverse ensemble classifiers in the base layer and an optimized meta-learning strategy. Unlike conventional stacking methods that primarily use simple classifiers, the TSE framework integrates bagging-based (random forest, extra trees classifier) and boosting-based (AdaBoost, gradient boosting machine, LightGBM) models during the first stage. This hybridization ensures a balance between variance reduction and predictive accuracy. The use of CatBoost as a meta-learner further distinguishes this approach, as it effectively handles categorical data, prevents overfitting, and employs ordered boosting to refine final predictions. The method also includes systematic hyperparameter tuning via GridSearchCV for both base and meta-learners, a step often overlooked in traditional stacking. Additionally, hybrid imputation for missing data and an analysis of computational performance (CPU/GPU usage, execution time) enhance its practical relevance. These innovations together result in a robust, scalable, and interpretable model, setting it apart from standard stacking techniques.

4. Results and Discussion

In the experimental setup for classifier models, the CSV file containing the dataset is loaded to begin the process. The dataset is pre-processed by selecting specific features related to heart disease. The dataset is divided into training and testing sets, with 80% going to training and 20% to testing, ensuring a randomized division using a fixed random seed for reproducibility. The effectiveness of the proposed method in classification is evaluated using random forest, AdaBoost, gradient boosting machine, LightGBM, and extra trees classifier.

In the experimental setup, Python 3.11 is used to implement the proposed method, with experiments conducted in the interactive Google Colab environment. For preprocessing, feature scaling is applied using the `StandardScaler` to standardize the data. SMOTE is used to tackle class imbalance within a dataset. For the classifiers, random forest employs 200 estimators with a maximum depth of 15 and uses the 'gini' criterion. AdaBoost and gradient boosting machine both use 200 estimators, with gradient boosting machine further specifying a maximum depth of 7 and a learning rate of 0.1. LightGBM employs a histogram-based learning strategy with 200 estimators, a maximum depth of 15, and a learning rate of 0.1, optimizing node splits efficiently through leaf-wise tree growth. Extra trees classifier uses 200 estimators with a maximum depth of 15.

Table 2 Performance of Classification Models on the Cleveland Dataset

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	83.61	84	84	84
AdaBoost	78.69	79	79	79
Gradient Boosting Machine	80.33	81	81	80
LightGBM	81.97	82	82	82
Extra Trees classifier	83.61	84	84	84
TSE	92.5	92.1	92.7	92.4

The classifiers are trained on the resampled training data. Their resulting predictions are then pooled and utilized as meta-features for Stage 2 of the ensemble learning approach. After training, the models are evaluated on the test set, with accuracy and classification reports printed for each model to assess their performance. Finally, the meta-features and corresponding labels are saved for future stages of model training. The optimal hyperparameters for CatBoost are determined as follows: the tree depth is set to 4 to control model complexity, ensuring it is not too deep while still capturing important relationships in the data. The number of iterations is set to 500, enabling the model to go through a sufficient number of boosting rounds for optimal training. The regularization parameter, l2_leaf_reg, is configured to 3 to reduce the risk of overfitting and enhance the model's generalization ability. The learning rate is set to 0.1 to balance the update size during training, ensuring stable learning without overshooting. Finally, a fixed random seed of 123 is used for reproducibility across experiments.

The accuracy, precision, Recall, F1_score, and ROC are used to evaluate each model using ten-fold CV. Table 2 presents the performance indicators of various classification models evaluated on the Cleveland dataset, highlighting Accuracy, Precision, Recall, and F1-Score, all expressed as percentages. These metrics evaluate the effectiveness of each model in predicting the target outcomes.

(1) Random Forest Classifier: Achieves an accuracy of 83.61% and balanced precision and recall at 84%, leading to an F1-score of 84%. This improvement reflects its ensemble approach, which mitigates overfitting by averaging multiple decision trees.

(2) AdaBoost Classifier: Records an accuracy of 78.69%, precision, recall, and F1-score at 79%. While less accurate than random forest and decision tree, its consistent performance indicates moderate robustness in handling the dataset.

(3) Gradient Boosting Machine Classifier: Matches the Decision Tree in accuracy at 80.33% with precision, recall, and F1-score also at 81%. Its performance indicates the potential underutilization of boosting capabilities in this specific dataset.

(4) LightGBM Classifier: Outperforms the gradient boosting machine classifier with an accuracy of 81.97% and balanced precision and recall at 82%, leading to an F1-score of 82%.

(5) Extra Trees Classifier: Performs on par with an accuracy of 83.61% and other metrics consistently at 84%. This demonstrates its effectiveness as another ensemble method, likely benefiting from randomness in both feature and sample selection.

(6) TSE: Excels among all models with the highest accuracy of 92.5%, precision at 92.1%, recall at 92.7%, and F1-score at 92.4%. This indicates that the TSE effectively combines predictions from multiple base models to capture patterns more accurately, achieving superior generalization.

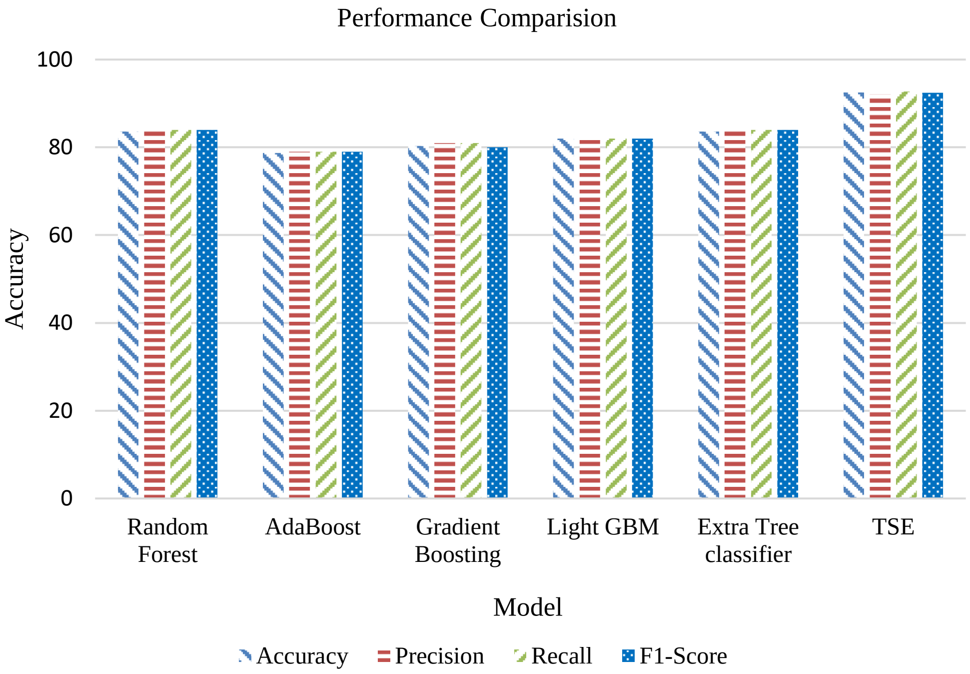

Fig. 2 Comparative Analysis of Ensemble Methods on the Cleveland Dataset

Fig. 2 illustrates the performance comparison of various models on the Cleveland heart disease dataset. The results clearly show that the proposed TSE method outperforms all baseline models, achieving the highest classification accuracy of 92.5%. The superior accuracy of TSE highlights the advantage of combining diverse base learners with an optimized meta-learner.

Table 3 summarizes the performance of different classification models on the Statlog dataset, evaluating their effectiveness using four key metrics: Accuracy, Precision, Recall, and F1-Score. Each metric reflects a specific aspect of the model's predictive capability.

(1) Random Forest: With an accuracy of 83.33%, random forest outperformed most individual models. Its precision (83%), recall (82%), and F1-score (82%) are consistently high, demonstrating its robustness in handling diverse data through ensemble decision trees.

(2) AdaBoost: This model delivered an accuracy of 79.63%, with balanced F1-score values, recall, and precision of 79% each. While its performance is slightly lower than random forest, its adaptive boosting strategy allows for improved predictions by focusing on difficult-to-classify instances.

(3) Gradient Boosting Machine: The model achieved an accuracy of 70.37%, with F1-scores, recall, and precision of 69% each. Its lower performance compared to AdaBoost may indicate challenges in capturing the underlying patterns in this dataset.

(4) LightGBM: This classifier achieved an accuracy of 75.93%, with precision, recall, and F1-score of 75%, 76%, and 75%, respectively. The relatively lower performance may stem from its sensitivity to overfitting, especially in smaller datasets or when hyperparameters are not optimally tuned.

(5) Extra Trees Classifier: This model showed an accuracy of 77.78%, with precision (77%), recall (75%), and F1-score (76%) reflecting a solid balance. By introducing additional randomness during tree construction, it generalizes better than single-tree models like Decision Trees.

(6) TSE: The model performed the best among all approaches, achieving the highest accuracy of 90.19%. It also showed strong precision (90%), recall (90%), and F1-score (90%). These results suggest that combining base models effectively amplifies their strengths, providing superior predictive performance.

Fig. 3 summarizes the results of the performance evaluation for the different models on the classification task for the Statlog heart disease dataset. TSE outperforms all other models with the highest accuracy of 90.19%. The TSE model also shows consistently higher values across other performance metrics, such as precision, recall, and F1-score, indicating its balanced predictive capability.

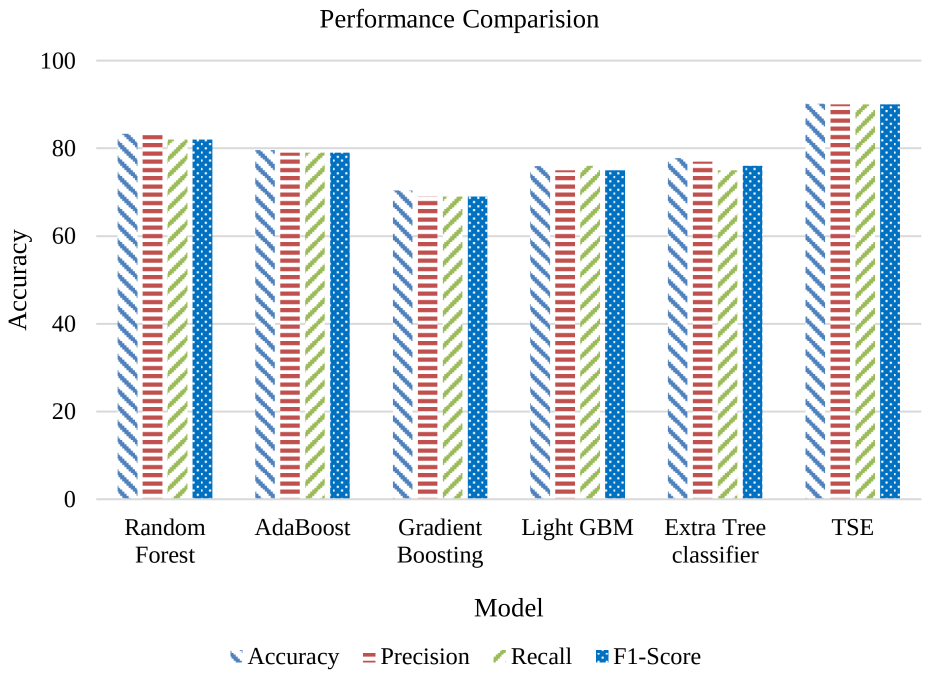

Fig. 3 Comparative Analysis of Ensemble Methods on Statlog Dataset

Deep learning models, such as convolutional neural networks and long short-term memory networks, have demonstrated strong predictive capabilities in medical diagnosis tasks. However, these models often require large datasets, significant computational resources, and lack interpretability, which is a critical factor in clinical applications. In contrast, the ensemble learning approach used in this study offers a balance between accuracy, interpretability, and computational efficiency. Given the need for explainability in medical decision-making, ensemble models provide a more suitable alternative. Future work may explore hybrid models combining deep learning with ensemble techniques to further enhance performance.

Table 3 Performance of Classification Models on the Statlog Dataset

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	83.33	83	82	82
AdaBoost	79.63	79	79	79
Gradient Boosting Machine	70.37	69	69	69
LightGBM	75.93	75	76	75
Extra trees classifier	77.78	77	75	76
TSE	90.19	90	90	90

5. Conclusions

This study presents a machine learning-based approach to assist the healthcare sector in the early detection of heart disease, aiming to enable timely interventions and informed decision-making. The primary objective is to develop a model that enhances predictive accuracy in heart disease detection. To achieve this, the performance of five base classifiers, random forest, AdaBoost, gradient boosting machine, LightGBM, and extra trees classifier, is analyzed and integrated into a novel TSE model. The key findings of this study are summarized as follows:

(1) The TSE model utilizes the outputs of the five base classifiers to generate a comprehensive meta-feature matrix, providing a robust foundation for further prediction refinement.

(2) A hyperparameter-tuned CatBoost algorithm is employed as a meta-learner to process the meta-feature matrix and deliver final predictions with improved accuracy and consistency.

The proposed TSE model demonstrated superior performance compared to individual classifiers, achieving an accuracy of 92.5% on the Cleveland dataset and 90.19% on the Statlog dataset. Moreover, the model achieved balanced precision, recall, and F1-scores, emphasizing its effectiveness in classification tasks. These findings underscore the significance of ensemble learning strategies in improving model generalization and predictive performance. Future research could focus on incorporating real-time medical datasets to enhance the model's applicability in clinical settings, as well as exploring advanced ensemble architectures or deep learning approaches for further optimization. Overall, this study highlights the potential of ensemble learning to enhance machine learning models for critical healthcare applications, offering a promising direction for the early detection and management of heart disease.

Conflicts of Interest

The authors declare no conflict of interest.

Statement of Ethical Approval

(a) Statement of human rights

For this type of study, a statement of human rights is not required.

(b) Statement on the welfare of animals

For this type of study, a statement on the welfare of animals is not required.

Statement of Informed Consent

For this type of study, informed consent is not required.

References

[1] C. Zhou, et al., “A Comprehensive Review of Deep Learning-Based Models for Heart Disease Prediction,” Artificial Intelligence Review, vol. 57, article no. 263, 2024.

[2] J. Shafi, et al., “Prediction of Heart Abnormalities Using Deep Learning Model and Wearable-devices in Smart Health Homes,” Multimedia Tools and Applications, vol. 81, no.1, pp. 543-557, 2022.

[3] P. Bizimana, Z. Zhang, M. Asim, and A. El-Latif, “An Effective Machine Learning-Based Model for Early Heart Disease Prediction,” BioMed Research International, vol. 2023, no. 1, article no. 3531420, 2023.

[4] K. M. Almustafa, “Prediction of Heart Disease and Classifiers’ Sensitivity Analysis,” BMC Bioinformatics, vol. 21, article no. 278, 2020.

[5] S. P. Barfungpa, H. K. M. Sarma, L. Samantaray, “An Intelligent Heart Disease Prediction System Using Hybrid Deep Dense Aquila Network,” Biomedical Signal Processing and Control, vol. 84, article no. 104742, 2023.

[6] A. Rehman, et. al., “HCDP-DELM: Heterogeneous Chronic Disease Prediction with Temporal Perspective Enabled Deep Extreme Learning Machine,” Knowledge-Based Systems, vol. 284, article no. 111316, 2024.

[7] A. Hutke and J. Deshmukh, "A Systematic Review of Machine Learning Approaches and Missing Data Imputation Techniques for Predicting Heart Disease," Proceedings of 2023 International Conference on Advanced Computing Technologies and Applications (ICATA), pp. 1-5, 2023.

[8] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," Algorithms, vol. 16, no. 2, article no. 88, 2023.

[9] E. A. Ogundepo and W. B. Yahya, "Performance Analysis of Supervised Classification Models on Heart Disease Prediction," Innovations in Systems and Software Engineering, vol. 19, pp. 129–144, 2023.

[10] M. Zeng, "The Prediction of Heart Failure Based on Four Machine Learning Algorithms," Highlights in Science, Engineering and Technology, vol. 39, pp. 1377–1382, 2023.

[11] A. Khan, M. Qureshi, M. Daniyal, and K. Tawiah, "A Novel Study on Machine Learning Algorithm-Based Cardiovascular Disease Prediction," Health & Social Care in the Community, vol. 2023, no. 1, article no.1406060, 2023.

[12] D. Shah, S. Patel, and S. K. Bharti, "Heart Disease Prediction Using Machine Learning Techniques," SN Computer Science, vol. 1, article no. 345, 2020.

[13] A. Garg, B. Sharma, and R. Khan, "Heart Disease Prediction Using Machine Learning Techniques," IOP Conference Series: Materials Science and Engineering, vol. 1022, article no. 012046, 2021.

[14] V. Shorewala, "Early Detection of Coronary Heart Disease Using Ensemble Techniques," Informatics in Medicine Unlocked, vol. 26, article no. 100655, 2021.

[15] G. Ahamad, et. al., "Influence of Optimal Hyperparameters on the Performance of Machine Learning Algorithms for Predicting Heart Disease," Processes, vol. 11, no. 3, article no. 734, 2023.

[16] P. Mahajan, S. Uddin, F. Hajati, M. A. Moni, "Ensemble Learning for Disease Prediction: A Review," Healthcare, vol. 11, no. 12, article no. 1808, 2023.

[17] P. Mahajan, S. Uddin, F. Hajati, M. A. Moni, "A Comparative Evaluation of Machine Learning Ensemble Approaches for Disease Prediction Using Multiple Datasets," Health and Technology, vol. 14, pp. 597–613, 2024.

[18] A. Hutke and J. Deshmukh, "A Novel Hybrid Approach for Feature Selection in Cardiovascular Risk Assessment", Advances in Technology Innovation, vol. 9, no. 4, pp. 319-331, 2024.