The Scientific Temper (2025) Vol. 16 (10): 4870-4878

E-ISSN: 2231-6396, ISSN: 0976-8653

Doi: 10.58414/SCIENTIFICTEMPER.2025.16.10.05

https://scientifictemper.com/

RESEARCH ARTICLE

Clean Balance-Ensemble CHD: A Balanced Ensemble Learning Framework for Accurate Coronary Heart Disease Prediction

Merlin Sofia S[1], D. Ravindran[2], G. Arockia Sahaya Sheela[3]

Abstract

Coronary Heart Disease (CHD) is still one of the leading causes of death worldwide, which necessitates early and reliable prediction methods to support timely medical interventions. Traditional machine learning approaches frequently struggle with noisy and imbalanced datasets which leading to biased predictions and reduced diagnostic reliability. To address these limitations, this paper proposes the CleanBalance-EnsembleCHD algorithm that combines data cleaning, balancing, and ensemble learning to improve prediction accuracy. The goal is to reduce noise, handle imbalance, and combine the strengths of multiple classifiers to detect CHDs more effectively. For noise reduction, the methodology employs Edited Nearest Neighbor (ENN) and Iterative Partitioning Filter (IPF), if imbalance persists Synthetic Minority Oversampling Technique (SMOTE) used. Five classifiers namely Rotation Forest, LogitBoost, Multilayer Perceptron, Logistic Model Trees (LMT), and Random Forest were trained, with the best models chosen for weighted soft-voting ensemble integration. The experimental evaluation on a CHD dataset with an initial class imbalance (maj/min ratio: 1.038, Gini index: 0.4998) revealed significant improvements. After ENN and IPF cleaning, the dataset was reduced from 1011 to 853 balanced instances (class counts: {1.0=414, 0.0=439}). Individual classifiers performed well, with accuracies of 97.36% (Rotation Forest), 94.72% (LogitBoost), 96.04% (Multilayer Perceptron), 97.95% (LMT), and 98.53% (Random Forest). After that, the top three models chosen Random Forest, LMT, and Rotation Forest were combined into an ensemble that outperformed all individual models on the test set, with Accuracy: 99.42%, F1-score: 0.9939, and MCC: 0.9884. These findings show that CleanBalance-EnsembleCHD provides superior predictive reliability leading to noise-resistant and balanced decision-making. Finally, the proposed framework provides a powerful and interpretable solution for early CHD detection using the potential to help clinicians with risk assessment and medical decision support.

Keywords: Coronary Heart Disease (CHD) Prediction, Balanced Ensemble Learning, Preprocessing, Noise Reduction, Prediction Accuracy.

[1]Research Scholar, Department of Computer Science, St. Joseph's College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.

[2]Associate Professor, Department of Computer Science, St. Joseph's College (Autonomous),Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.

[3]Assistant Professor, Department of Computer Science, St. Joseph's College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.

***Corresponding Author:** Merlin Sofia S, Research Scholar, Department of Computer Science, St. Joseph's College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India, E-Mail: Merlinsofia13@gmail.com

How to cite this article: Sofia, M.S., Ravindran, D., Sheela, G.A.S. (2025). Clean Balance-Ensemble CHD: A Balanced Ensemble Learning Framework for Accurate Coronary Heart Disease Prediction. The Scientific Temper, **16**(10): 4870-4878.

Doi: 10.58414/SCIENTIFICTEMPER.2025.16.10.05

Source of support: Nil

Conflict of interest: None.

Introduction

Coronary heart disease (CHD) is a leading cause of death worldwide and greatly burdens healthcare systems (Hammoud, A., Karaki, A., Tafreshi, R., Abdulla, S., & Wahid, M.F. 2024). Early detection of CHD can help save lives, but the complexity of risk factors makes it difficult (Song, J. 2024). Machine learning makes it possible to detect hidden patterns in medical data and make accurate CHD predictions possible (Dubey, M., Tembhurne, J., & Makhijani, R. 2024). But quality, imbalance, noise, and duplicates affect the performance of the prediction model (Al-Ssulami, A.M., Alsorori, R.S., Azmi, A.M., & Aboalsamh, H. 2023). Traditional ML models in CHD prediction are weakened by noise and unbalanced data, misclassifying a minority of cases. Thus, an intelligent framework is needed that combines the strengths of classifiers, overcoming quality and imbalance (Wanyonyi, M., Morris, Z.N., Musyoka, F.M., & Kitavi, D.M. 2025).

Many studies have used machine learning methods such as SVM, KNN, decision trees, bagging, and boosting for CHD prediction. Over/undersampling methods were

Received: 13/09/2025 **Accepted:** 05/10/2025 **Published:** 16/10/2025

used to overcome the imbalance, and feature selection and dimensionality reduction were used for performance improvement. Despite improvements, oversampling causes artificial noise and undersampling causes valuable data loss. Many models fail to remove noise/overlapping records, resulting in overfitting and poor generalization. Relying on a single classifier limits robustness, as no single algorithm performs consistently well across all datasets; thus, a unified strategy that combines quality, balance improvement, and multiple classifiers is necessary.

To address these shortcomings, the study introduces CleanBalance-EnsembleCHD, a new CHD framework that integrates data cleaning, balancing, and ensemble learning. Reliable models with noise reduction are generated by the Edited Nearest Neighbor (ENN) and Iterative Partitioning Filter (IPF); if there is imbalance, the class is balanced by the Synthetic Minority Oversampling Technique (SMOTE). The cleaned data is trained on classifiers such as RotationForest, LogitBoost, MultilayerPerceptron (MLP), Logistic Model Trees (LMT), and RandomForest; Weighted Soft-Voting Ensemble Learning is used to combine the best models and increase prediction accuracy and reliability.

The framework begins with data analysis, imbalance detection, and Gini index computation. It systematically improves dataset quality by using ENN and IPF, subsequent to SMOTE as needed. After creating a balanced dataset, stratified sampling is used to ensure fair train-test splits. The framework takes advantage of the diversity of classifiers by selecting the best-performing models for Ensemble Integration. The final ensemble significantly outperformed the individual models.

The key contributions of this work are:

- Proposing CleanBalance-EnsembleCHD algorithm which is an integrated pipeline combining noise removal, balancing, and ensemble learning for CHD prediction.
- Employing ENN and IPF to clean noisy and overlapping records while preserving representative samples.
- Using stratified sampling and balanced dataset preparation to avoid bias.
- Training various classifiers with building a weighted soft-voting ensemble using the top-performing models.
- Showing superior performance with high accuracy, F1-score, with MCC which guarantees reliability in medical prediction contexts.

The goal of this study is to design a noise-resilient, and also balanced ensemble framework that guarantees robust with accurate prediction of coronary heart disease.

The specific objectives contain:

- To clean and balance the CHD dataset for unbiased learning.
- To assess multiple classifiers and detect the most effective ones.
- To construct a weighted ensemble model that utilizes classifier complementarity.
- To validate the proposed approach utilizing rigorous performance metrics.

CleanBalance-EnsembleCHD is unique in that it combines ENN-IPF-SMOTE preprocessing with ensemble incorporation of various sophisticated classifiers to ensure data quality and predictive robustness. Unlike previous work, it systematically handles noise, overlap, and imbalance before model training leading to more accurate outcomes.

The proposed framework can be utilized in clinical decision support systems, hospitals, and health monitoring platforms to assist physicians in early CHD detection and risk assessment. Its adaptability also makes it applicable to other medical domains where data imbalance and noise are common. The remainder of this paper is organized as follows: Section 2 presents the related works of the proposed methodology in detail. Section 3 describes dataset preparation and preprocessing and the details about the CleanBalance-EnsembleCHD algorithm. Section 4reports experimental results and performance analysis. Section 5 concludes the study and shows future research directions.

Literature Review

Recently, research is underway to improve the accuracy and reliability of diagnosis by using machine learning to predict CHD. Khdair, H.S., & Dasari, N. (2021) examined several ML algorithms in CHD prediction and emphasized the importance of accurate handling of clinical data.

Ayon, S.I., Islam, M.M., & Hossain, M.R. (2020) demonstrated in a computational intelligence comparison that ensemble methods perform better than individual classifiers in capturing complex patterns in CHD data. Poonkuzhali, R., Pavithra, S., Kumar, K.S., & Nallusamy, C. (2024) demonstrated the effectiveness of feature selection using ML approaches to predict CHD in large experiments.

AmoghVarshith, 1.P., Kumar, S., Harika, D., D.Haritha, & Priyanka, P. (2024) compared several methods and emphasized the benefits of ensemble strategies in improving accuracy in predicting CHD outcomes. Srinija, P., & Pranay, V. (2025) evaluated classifiers on benchmark data and emphasized that model performance is affected by preprocessing and equalization techniques.

Guo, H. (2023) compared five ML models and found that the mixed and ensemble approaches provided higher accuracy than single models. Rasheed, M., Khan, M.A., Elmitwally, N.S., Issa, G.F., Ghazal, T.M., Alrababah, H., & Mago, B. (2022) proposed an ML framework for predicting CHD in a cyber resilience context and emphasized that data preprocessing is important for robust performance.

Keshav Srivastava, D., Choubey, K., & Choubey, D.K. (2020) emphasized the important role of data mining techniques with ML models to improve prediction accuracy. Omkari, D.Y., & Shaik, K. (2024) combined multiple classifiers in a TLV

Table 1: Summary Table

Authors	*Methods*	*Key Findings*	*Limitations*
Khdair & Dasari [6]	Applied multiple machine learning techniques for CHD prediction	Showed potential of ML in analyzing complex medical datasets	Did not address class imbalance and interpretability
Ayon et al. [7]	Comparative study of computational intelligence techniques	Demonstrated variation in accuracy across techniques	Lack of unified ensemble approach
Poonkuzhali et al. [8]	Supervised ML algorithms	Showed adaptability to patient datasets	Limited exploration of hybrid models
AmoghVarshith et al. [9]	Multiple ML algorithms	Improved prediction efficiency	Scalability issues with larger datasets
Srinija & Pranay [10]	Machine learning-based CHD prediction	Reliable detection in small datasets	Poor generalization to diverse populations
Guo [11]	Comparative study with five ML models	Identified trade-offs between complexity and performance	Limited scope of models tested
Rasheed et al. [12]	Systematic evaluation of ML methods	Highlighted improved generalization	Lacked ensemble integration
Srivastava et al. [13]	Machine learning and data mining	Hybrid methods boosted prediction	Data preprocessing challenges not fully solved
Omkari & Shaik [14]	Two-layered voting (ensemble) framework	Achieved higher performance via ensemble	Increased computational complexity
Mohd et al. [15]	Survey of ML techniques	Provided comprehensive overview of heart disease ML methods	No experimental validation
Kavitha et al. [16]	Hybrid ML model	Achieved superior prediction accuracy	Complexity and interpretability issues remain

framework for CHD prediction and demonstrated that multi-layered ensembles improved performance and consistency over individual ensembles.

Mohd, N., Sharma, J., & Upadhyay, D. (2021) identified challenges such as class imbalance and noise in CHD prediction, revealing that they can affect generalization. Kavitha, D.M., Gnaneswar, G., Dinesh, R., Sai, Y.R., & Suraj, R. (2021) developed a hybrid ML model that combines multiple classifiers and showed that it significantly reduces error rates.

Although current research confirms the strength of ML in CHD prediction, noisy data, class imbalance, and generalization deficiencies remain ongoing issues. These limitations highlight the need for a robust framework that integrates data cleaning, balancing, and ensemble learning of CleanBalance-EnsembleCHD (Table 1).

Materials and Methods

This section explains the procedural background of the CleanBalance-EnsembleCHD framework for CHD prediction. To overcome noise, class imbalance, and sample heterogeneity issues, CleanBalance-EnsembleCHD will integrate preprocessing, data cleaning, equalization, and ensemble modeling steps.

This algorithm follows a structured pipeline that combines data preparation, noise reduction via an Edited Nearest Neighbors (ENN)/Iterative Partitioning Filter (IPF), class balancing *via* a Synthetic Minority Oversampling Technique (SMOTE), and model training with a Weighted Soft-Voting group.

This design combines noisy data removal, misclassification correction, class imbalance correction, and ensemble learning to ensure fairness and robustness by integrating multiple classifiers. This creates a reliable predictive framework across multiple populations; Algorithm 1 shows CleanBalance-EnsembleCHD.

The CleanBalance-EnsembleCHD algorithm predicts coronary heart disease (CHD) by combining data cleaning, balancing, addition to ensemble learning in a structured workflow. It starts by examining class distribution and measuring imbalances with the imbalance ratio and Gini Index. To improve data quality, noisy and overlapping samples are removed using Edited Nearest Neighbors (ENN), and also instances that are consistently misclassified are filtered out using the Iterative Partitioning Filter (IPF). The cleaned dataset is then balanced with SMOTE to produce synthetic samples for the minority class. After dividing the balanced data into training and testing sets, various classifiers such as RotationForest, LogitBoost, MultilayerPerceptron, Logistic Model Trees, and also RandomForest are trained and tested using cross-validation. The best-performing models, as determined by F1 scores, are combined using a weighted soft voting ensemble. Finally, the ensemble is tested on unseen data, as well as its performance is assessed using accuracy, precision, recall, F1-score, and MCC. The

Algorithm 1: CleanBalance-EnsembleCHD
Input: PFHD dataset (Preprocessed & Feature-selected)
Output: Best-performing ensemble model for CHD prediction

1: Analyze class distribution → detect imbalance
2: IR ← N_majority / N_minority
3: Gini ← 1 - Σ (p_i)^2

4: D_ENN ← Apply ENN(D) // remove noisy/overlapping samples
5: BaseModel ← Train(RandomForest, D_ENN)
6: D_IPF ← RemoveMisclassified(D_ENN, BaseModel) // IPF filtering

7: D_balanced ← ApplySMOTE(D_IPF)
8: (D_train, D_test) ← StratifiedSplit(D_balanced)

9: Models ← {RotationForest, LogitBoost, MLP, LMT, RandomForest}
10: For each model m in Models do
Train m on D_train
Evaluate m using 5-fold CV → F1_m, MCC_m
End For

11: TopModels ← Select models with highest F1 scores
12: Ensemble ← WeightedSoftVoting(TopModels, weights = F1_m)

13: Predictions ← Ensemble(D_test)
14: Evaluate Predictions using {Accuracy, Precision, Recall, F1, MCC}

15: BestModel ← Select configuration with highest balanced performance
16: Save(BestModel)

best configuration is saved for integration into the full CHD prediction framework. Figure 1 shows the flow diagram of CleanBalance-EnsembleCHD algorithm.

The flow diagram of the CleanBalance-EnsembleCHD algorithm depicts the entire pipeline for predicting CHD. It cleans noisy data, balances classes, trains diverse classifiers, as well as combines the best models using weighted soft voting to deliver accurate with reliable CHD predictions.

Dataset Preparation

The experimental dataset utilized in this study is known as the Preprocessed and Feature-Selected Heart Disease (PFHD) dataset. It was created by combining five benchmark CHD datasets commonly used in machine learning research: Cleveland, Hungarian, Long Beach, Switzerland, and Statlog. Each dataset comes from various clinical settings which ensures a diverse range of patient demographics and medical profiles. Eq. (1) shows Dataset Representation.

$$D = \{(x_i, y_i) \mid x_i \in \mathbb{R}^n, y_i \in \{0,1\}, i=1,2,\ldots,N\} \quad (1)$$

where:
D: dataset
x_i: n-dimensional feature vector of patient i
y_i: binary target class (0 = no CHD, 1 = CHD)
N: total number of patient records

Eq. (1) formally defines the dataset as a set of feature vectors with corresponding class labels. Each patient record contains multiple clinical features, while the class label indicates whether the patient has CHD or not. This mathematical representation serves as the foundation for all future preprocessing, with modeling steps. This unified dataset improves generalizability by combining diverse patient populations, making the proposed framework more robust to dataset-specific biases. The dataset preparation is a critical step in ensuring that the CHD prediction framework uses reliable with representative data. Raw clinical datasets frequently contain incomplete records, irregular patterns, as well as irrelevant features, which can skew the learning process. Consequently, a structured preprocessing pipeline is applied prior to model training.

First, missing data are handled and incomplete entries are filled in to ensure consistency and fairness across all patient samples. Then, as outliers are removed, the learning algorithm avoids false conclusions and creates a dataset that accurately reflects the underlying population. After that, normalization is done to equalize the different ranges of

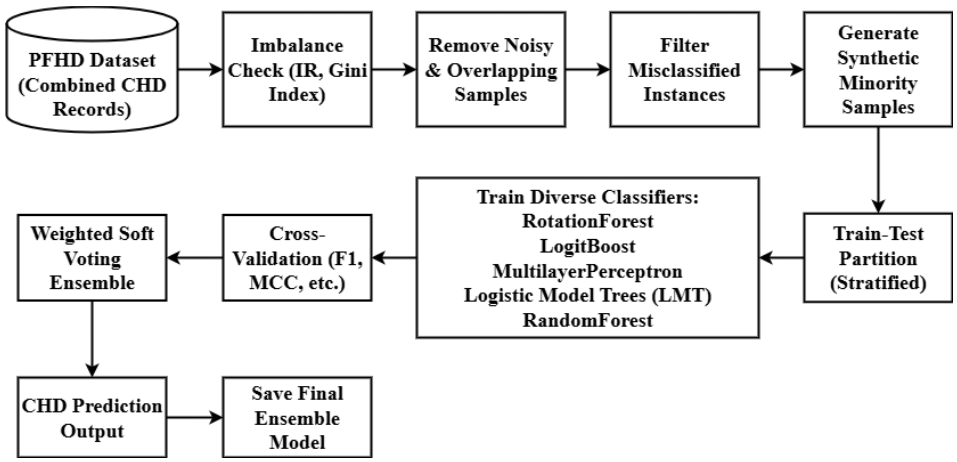

Figure 1: Flow diagram of CleanBalance-EnsembleCHD algorithm

medical features (fat, blood pressure, age), otherwise some variables will be more influential than others.

By equalizing features, each attribute contributes equally; feature selection identifies key variables in CHD prediction and eliminates unnecessary ones. This step focuses on clinical factors, structures and balances the data, and creates the foundation for accurate and generalizable CHD prediction.

Class Imbalance Analysis

For sample reliability, the class distribution was examined and the skewness of the data was estimated through the majority-minority ratio. Eq. (2) shows Imbalance Ratio.

$$IR = N\_majority / N\_minority \quad (2)$$

Where:
IR = Imbalance Ratio
N_majority = Number of majority class samples
N_minority = Number of minority class samples
The imbalance ratio measures how skewed the dataset is by comparing the number of majority and minority class samples. A higher imbalance ratio indicates that one class significantly outperforms the other that may bias the classifier to predict the majority class more frequently. Impurity in class distribution was additionally measured to determine dataset bias. The Gini Index was used to measure inequality between class labels. Eq. (3) shows Gini Index.

$$Gini = 1 - \Sigma (p\_i)^2 \quad (3)$$

Where:
Gini = Gini Index value
p_i = Probability of class i in the dataset
Σ = Summation across all classes

The Gini Index measures the impurity or inequality in class distribution. If all patients belong to the same class, the Gini value is zero, indicating perfect purity. A higher Gini value reflects a more balanced but potentially noisy class distribution, making it useful for assessing dataset fairness.

Noise and Overlap Reduction

The dataset may contain noisy or overlapping records that reduces classifier performance. To improve quality, Edited Nearest Neighbors (ENN) was used to remove instances that cause excessive overlap between classes. This filtering ensures that only reliable and representative data points are used for training. Equation (4) depicts the quality measure for samples.

$$Q(x) = d(x, y\_same) - d(x, y\_diff) \quad (4)$$

Where:
Q(x) = Quality measure for a sample x
d(x, y_same) = Distance between x and nearest neighbor of the same class
d(x, y_diff) = Distance between x and nearest neighbor of a different class

This quality metric compares how close a data point is to its nearest neighbor in the same class versus one from the opposite class. A positive value indicates that the sample is more similar to its own class which implies reliability, whereas negative or small values may indicate noisy or overlapping data. ENN systematically removes potentially misleading samples, increasing the dataset's robustness.

Iterative Partition Filtering

Iteratively removing instances that were consistently misclassified by a trained model helped to refine the dataset even more. This filtering step ensured which borderline or mislabeled data did not reduce model performance. Eq. (5) shows Misclassification Error.

$$M\_error = |y\_true - y\_pred| \quad (5)$$

Where:
M_error = Misclassification error for a sample
y_true = True class label
y_pred = Predicted class label

Eq. (5) quantifies the difference between a sample's true and predicted labels. A value of zero indicates that the prediction is correct, whereas a nonzero value indicates an error. Iteratively filtering out such misclassified records makes the dataset cleaner and more robust.

Data Balancing

Following cleaning, synthetic samples for the minority class were created to balance class representation. This step ensured that classifiers were not biased toward the majority class. Eq. (6) demonstrates Synthetic Sample Generation.

$$x\_new = x\_minority + \lambda * (x\_neighbor - x\_minority) \quad (6)$$

Where:
x_new = Generated synthetic minority sample
x_minority = Existing minority sample
x_neighbor = Nearest neighbor of the minority sample
λ = Random number between 0 and 1

Eq. (6) describes how to generate new synthetic samples for the minority class. Realistic synthetic examples are created by interpolating between an existing minority sample and one of its neighbors with applying a random scaling factor. This balances the dataset and lowers bias towards the majority class.

Train-Test Partitioning

To evaluate model performance, the balanced dataset was divided into two subsets: training and testing. To keep relative class proportions consistent across splits which a stratified approach was used. Eq. (7) indicates Train-Test Partitionin.

$$D = D\_train \cup D\_test,\ D\_train \cap D\_test = \emptyset \quad (7)$$

Where:
D = Complete dataset
D_train = Training subset

D_test = Testing subset
∅ = Empty set, indicating no overlap between splits
This expression ensures that the dataset is split into two disjoint subsetsshows training data used for model learning, and testing data used for evaluation. The non-overlapping condition guarantees that the evaluation remains unbiased by preventing data leakage.

Model Training

Multiple classifiers were trained on the processed training set to capture a variety of decision boundaries. RotationForest, LogitBoost, MLP, LMT, and RandomForest were selected for their diversity, as they offer different learning methods and strengths:

Rotation forest

Randomly rotating features and learning from a different subset of features improves accuracy and reduces correlation.

LogitBoost

Reduces logistic loss, engages weak learners, and provides accurate probabilities by handling complex boundaries.

MLP

A neural classifier that learns nonlinear relationships with hidden layers and captures complex medical data.

LMT

Combines decision trees and logistic regression to provide clarity and precision to medical data.

RandomForest

Combines multiple random decision trees, preventing overfitting and handling high-dimensional data and feature importance.

Eq. (8) demonstrates the classifier prediction function.

$$h_m(x) = f_m(W_m, x) \tag{8}$$

Where:
h_m(x) = Prediction of classifier m
f_m = Mapping function for classifier m
W_m = Learned parameters of classifier m
x = Input feature vector

Eq. (8) processes each classifier input and predicts the class label. Combining models such as RotationForest, LogitBoost, MLP, LMT, RandomForest, etc., the various strategies combine to provide robust and generalized prediction.

Model Evaluation

Each classifier was evaluated using multiple metrics to ensure balanced performance across both classes. Key metrics included accuracy, precision, recall, F1-score, and MCC. Eq. (9–13) shows Evaluation Metrics.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \tag{9}$$

$$\text{Precision} = TP / (TP + FP) \tag{10}$$

$$\text{Recall} = TP / (TP + FN) \tag{11}$$

$$F1 = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \tag{12}$$

$$MCC = (TP*TN - FP*FN) / \sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)} \tag{13}$$

Where:
TP = True Positives
FP = False Positives
FN = False Negatives
TN = True Negatives
These equations Eq. (9–13) define standard evaluation metrics. Accuracy measures overall correctness, Precision focuses on how many positive predictions are correct, recall captures how many actual positives are identified, F1-score balances Precision and Recall, and MCC provides a correlation-based measure of classification quality even in imbalanced settings. Together, these metrics provide a comprehensive evaluation of model performance.

Ensemble Construction

Top-performing classifiers were combined using a weighted soft voting ensemble. Each model's prediction contributed to the final decision in proportion to its F1-score. Eq. (14) shows Weighted Soft Voting.

$$P_{final}(c) = \Sigma (w_m * P_m(c)) / \Sigma w_m \tag{14}$$

Where:
P_final(c) = Final probability for class c
P_m(c) = Predicted probability of class c by model m
w_m = Weight of model m based on its F1-score
Σ = Summation across selected models

This ensemble method combines the predictions of multiple classifiers by weighting each contribution according to its performance (measured by F1-score). The final decision is made based on the aggregated weighted probabilities, leveraging the strengths of diverse models while minimizing individual weaknesses.

Final Model Selection

The ensemble model's predictions were compared against actual labels on the testing set. The final configuration was selected based on achieving the best balance across all evaluation metrics. Eq. (15) shows Final Model Scoring.

$$Score_{final} = \alpha*Accuracy + \beta*F1 + \gamma*MCC \tag{15}$$

Where:
Score_final = Final evaluation score

Accuracy = Proportion of correctly classified samples
F1 = F1-score of the model
MCC = Matthews Correlation Coefficient
α, β, γ = Weighting factors for balanced evaluation

Eq. (15) defines the overall evaluation criterion for selecting the final model. By assigning weights to Accuracy, F1-score, and MCC, the score ensures that the chosen model performs well across different performance dimensions rather than excelling in only one metric. This balanced scoring system leads to a more reliable model.

Discussion

Experimental Setup

The proposed CleanBalance-EnsembleCHD algorithm was experimentally evaluated in the Java programming environment with the WEKA tool. The dataset was cleaned, balanced, and feature selected prior to training and testing the classifiers. Several machine learning models were trained and compared to the proposed ensemble approach. The performance was evaluated using Accuracy, Precision, Recall, and F1-score.

Results

The CleanBalance-EnsembleCHD algorithm was tested on a CHD dataset with the following initial properties: class counts (0.0=515, 1.0=496), imbalance ratio (1.038), and Gini index (0.4998). After applying cleaning methods, ENN reduced the dataset from 1011 to 853 instances (counts: {1.0=414, 0.0=439}), while IPF maintained the balanced distribution. After applying SMOTE, the dataset was divided into 682 training and 171 testing samples.

Classifier performance (5-fold CV): Rotation Forest (Acc: 97.36%, F1: 0.9731, MCC: 0.9473), LogitBoost (Acc: 94.72%, F1: 0.9459, MCC: 0.8944), Multilayer Perceptron (Acc: 96.04%, F1: 0.9596, MCC: 0.9210), LMT (Acc: 97.95%, F1: 0.9788, MCC: 0.9589), and Random Forest (Acc: 98.53%, F1: 0.9848, MCC: 0.9707). The final ensemble included Random Forest, LMT, and Rotation Forest.

Ensemble results: On the test set (171 instances), the ensemble achieved 99.42% accuracy, 0.9939 F1-score, and 0.9884 MCC, correctly classifying 170 of 171 cases. Other evaluation metrics include the Kappa statistic (0.9883), MAE (0.0282), RMSE (0.0904), and 100% coverage at the 0.95 confidence level.

Other classifiers: SVM (Acc: 96.48%, F1: 0.9648), KNN (Acc: 91.78%, F1: 0.9179), and REPTree (Acc: 93.54%, F1: 0.9354) all outperformed baseline models. Table 2 compares the results obtained by Kadhim, M. A., & Radhi, A. M. (2023) and the proposed CleanBalance-EnsembleCHD algorithm.

The results show that CleanBalance-EnsembleCHD consistently outperformed the baseline models developed by Kadhim, M. A., & Radhi, A. M. (2023). For example, Random Forest's accuracy increased from 94.9% to 97.94%, while SVM's increased from 89.0% to 96.48%. This improvement was achieved by cleaning and balancing the data, and combining multiple models together.

Discussion

Figure 2 compares the accuracy of Kadhim, M. A., & Radhi, A. M. (2023) and CleanBalance-EnsembleCHD; the ensemble performed better, while RandomForest achieved 97.94% accuracy. The improvement is due to improved generalization with cleaned balanced data.

Figure 3 illustrates that the CleanBalance-EnsembleCHD outperformed the baseline using precision scores. The Random Forest model provides the highest precision especially 0.98. This improvement demonstrates that the CleanBalance-EnsembleCHD effectively reduced false positives. Therefore, it makes predictions more reliable for clinical decision-making.

Figure 4 illustrates recall performance. The CleanBalance-EnsembleCHD improved recall for all classifiers which results more positive CHD cases being correctly identified. The development of CleanBalance-EnsembleCHD demonstrates its ability to correctly identify a minority of CHD cases and reduce missed diagnoses.

Figure 5 shows that CleanBalance-EnsembleCHD showed the best balance (RandomForest 0.97) in F1 value, which helps to accurately identify CHD cases and reduce misclassifications. Experimental results confirm the robustness and reliability of CleanBalance-EnsembleCHD, a combination of data cleaning, balancing, and ensemble learning, showing improvements in accuracy, recall, and F1 value, confirming its robustness and reliability in CHD prediction.

Future work and Conclusion

This study introduced the CleanBalance-EnsembleCHD framework, which improves CHD prediction by cleaning

Table 2: Performance Comparison

Classifiers	*Kadhim et al. [17]*				*CleanBalance-EnsembleCHD Algorithm*			
	Accuracy (in %)	*Precision*	*Recall*	*F1-Score*	*Accuracy (in %)*	*Precision*	*Recall*	*F1-Score*
SVM	89.0	0.87	0.93	0.90	96.48	0.96	0.96	0.96
KNN	88.6	0.87	0.92	0.89	91.78	0.91	0.91	0.91
DT	89.9	0.93	0.86	0.87	93.54	0.93	0.93	0.93
RF	94.9	0.94	0.96	0.95	97.94	0.98	0.97	0.97

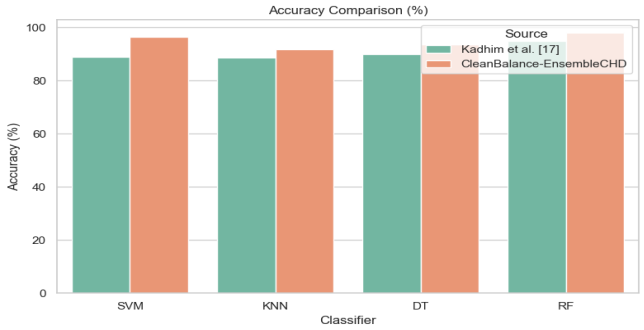

Figure 2: Accuracy Comparison

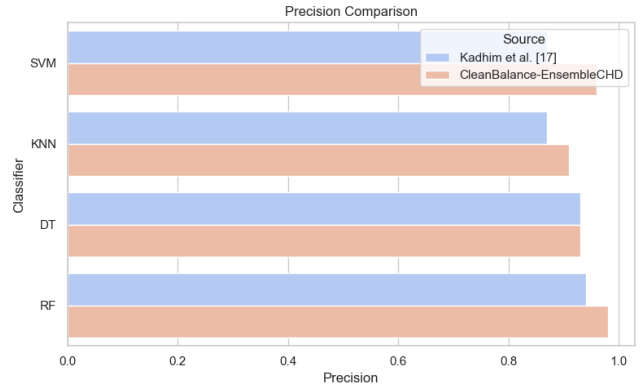

Figure 3: Precision Comparison

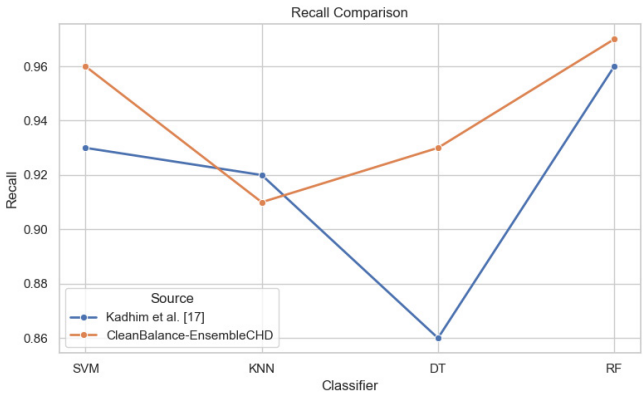

Figure 4: Recall Comparison

data, balancing classes, and integrating ensemble learning with multiple classifiers. The best results were obtained by combining classifiers such as ENN, IPF, SMOTE, and Rotation Forest, LogitBoost, MLP, LMT with Random Forest. The weighted ensemble provided the highest accuracy in MCC and F1-score.

Limitations

Although the framework shows strong prediction in five datasets, its applicability to highly heterogeneous clinical data and computational challenges in low-resource environments have yet to be validated.

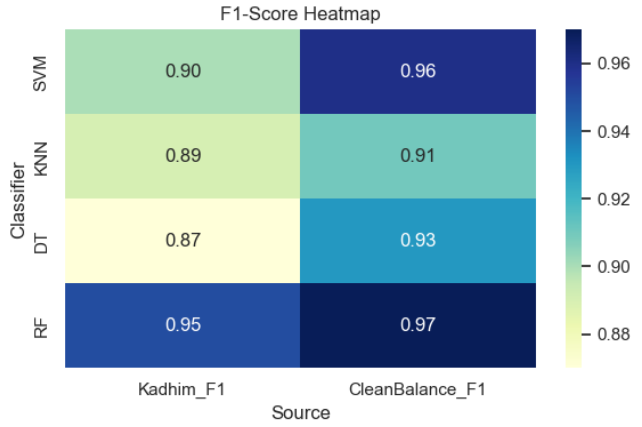

Figure 5: F1-Score Comparison

Future Works

Future research could focus on extending the framework to real clinical data by adding deep learning models and improving reliability and interpretation through interpretable AI.

References

Hammoud, A., Karaki, A., Tafreshi, R., Abdulla, S., & Wahid, M.F. (2024). Coronary Heart Disease Prediction: A Comparative Study of Machine Learning Algorithms. Journal of Advances in Information Technology.

Song, J. (2024). Coronary Heart Disease Prediction from Common Risk Factors. Highlights in Science, Engineering and Technology.

Dubey, M., Tembhurne, J., & Makhijani, R. (2024). Improving coronary heart disease prediction with real-life dataset: a stacked generalization framework with maximum clinical attributes and SMOTE balancing for imbalanced data. Multim. Tools Appl., 83, 85139-85168.

Al-Ssulami, A.M., Alsorori, R.S., Azmi, A.M., & Aboalsamh, H. (2023). Improving Coronary Heart Disease Prediction Through Machine Learning and an Innovative Data Augmentation Technique. Cognitive Computation, 15, 1687 - 1702.

Wanyonyi, M., Morris, Z.N., Musyoka, F.M., & Kitavi, D.M. (2025). Enhanced machine learning and hybrid ensemble approaches for coronary heart disease prediction. medRxiv.

Khdair, H.S., & Dasari, N. (2021). Exploring Machine Learning Techniques for Coronary Heart Disease Prediction. International Journal of Advanced Computer Science and Applications.

Ayon, S.I., Islam, M.M., & Hossain, M.R. (2020). Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques. IETE Journal of Research, 68, 2488 - 2507.

Poonkuzhali, R., Pavithra, S., Kumar, K.S., & Nallusamy, C. (2024). Heart Disease Prediction Using Machine Learning Techniques. 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), 1-7.

AmoghVarshith, 1.P., Kumar, S., Harika, D., D.Haritha, & Priyanka, P. (2024). Heart Disease Prediction Using Machine Learning Algorithms. 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS), 1, 1-5.

Srinija, P., & Pranay, V. (2025). Heart Disease Prediction Using Machine Learning. International Journal of Research Publication and Reviews.

Guo, H. (2023). Comparative Study on Coronary Heart Disease Prediction Using Five Machine Learning Models. Proceedings of the 1st International Conference on Data Analysis and Machine Learning.

Rasheed, M., Khan, M.A., Elmitwally, N.S., Issa, G.F., Ghazal, T.M., Alrababah, H., & Mago, B. (2022). Heart Disease Prediction Using Machine Learning Method. 2022 International Conference on Cyber Resilience (ICCR), 1-6.

Keshav Srivastava, D., Choubey, K., & Choubey, D.K. (2020). Heart Disease Prediction using Machine Learning and Data Mining. International Journal of Recent Technology and Engineering.

Omkari, D.Y., & Shaik, K. (2024). An Integrated Two-Layered Voting (TLV) Framework for Coronary Artery Disease Prediction Using Machine Learning Classifiers. IEEE Access, 12, 56275-56290.

Mohd, N., Sharma, J., & Upadhyay, D. (2021). A SURVEY: HEART DISEASE PREDICTION USING MACHINE LEARNING TECHNIQUES. Webology.

Kavitha, D.M., Gnaneswar, G., Dinesh, R., Sai, Y.R., & Suraj, R. (2021). Heart Disease Prediction using Hybrid machine Learning Model. 2021 6th International Conference on Inventive Computation Technologies (ICICT), 1329-1333.

Kadhim, M. A., & Radhi, A. M. (2023). Heart disease classification using optimized Machine learning algorithms. Iraqi Journal For Computer Science and Mathematics, 4(2), 31-42.