Synthesis: A Multidisciplinary Research Journal
Volume 3, Issue 1, January - March 2025, Pp: 37-48

Research Paper

Stacked Ensemble Model with Smote for Heart Disease Prediction

[1] V.Rama Rao,[2*] Dhanukonda Visalakshi, [3] D.Nirmala Kumari,[4] Ch.Likitha Lakshmi Prasanna, [5] B.Lakshmi Padmaja, [6] E.Aparna

[1]Assistant professor, Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India. ORCID ID:0009-0003-3102-8874

[2, 3,4,5,6] B.Tech Student, Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India.

[1]Email Id: ramaraovangapandu28@gmail.com, ORCID ID:0009-0003-3102-8874
[2]Email Id: nirmalakumaridimmila@gmail.com, ORCHID ID:0009-0004-5791-2710
[3]Email Id: cllprasanna2004@gmail.com, ORCHID ID:0009-0001-2684-8857
[4]Email Id: blakshmipadmaja2004@gmail.com, ORCHID ID: 0009-0005-5125-3934
[5]Email Id: erigalaa@gmail.com, ORCHID ID :0009-0006-2584-085X

**Corresponding Author(s): dvisalakshi796@gmail.com*

Article Info

Article History
Received: 15/11/2024
Revised: 11/02/2025
Accepted:15/03/2025
Published :31/03/2025

Abstract

Cardiovascular disease, particularly heart disease, remains a leading cause of global mortality, with early detection being critical to improving patient outcomes. Traditional machine learning models for heart disease prediction often suffer from class imbalance and limited generalization capabilities. This study aims to develop a robust, interpretable, and computationally efficient predictive model that addresses these limitations. The proposed approach integrates the Synthetic Minority Over-sampling Technique (SMOTE) with a stacked ensemble learning architecture composed of Decision Tree and Random Forest as base learners, and Logistic Regression as a meta-learner. A standardized preprocessing pipeline involving median imputation, Min-Max normalization, and one-hot encoding was applied to a clinical dataset of 1,025 patient records sourced from the Kaggle UCI repository. SMOTE was utilized to balance the minority class representing heart disease cases. The model was evaluated using 5-fold stratified cross-validation on key performance metrics. The stacked ensemble achieved an accuracy of 98.2%, precision of 1.00, recall of 0.96, F1-score of 0.98, and AUC-ROC of 0.99, significantly outperforming standalone models and recent ensemble-based methods. Implementation required minimal computational resources and executed efficiently on CPU-only systems, making it suitable for real-time clinical applications. The study demonstrates that combining data-level balancing and model-level stacking significantly enhances diagnostic accuracy, particularly in class-imbalanced medical datasets. Future work will explore explainable AI integration, validation on diverse clinical populations, and deployment on resource-constrained edge devices.

Keywords: Heart Disease Prediction, Stacked Ensemble, SMOTE, Machine Learning, Medical Diagnostics, Class Imbalance, Logistic Regression, Random Forest, Decision Tree

1. Introduction

Cardiovascular disease (CVD), particularly heart disease, remains the foremost cause of death worldwide, claiming millions of lives annually. The World Health Organization (WHO) reports that heart-related ailments contribute to over 30% of all global deaths, with India witnessing one of the highest mortality rates due to cardiac issues. The increasing prevalence of risk factors such as hypertension, diabetes, obesity, high cholesterol, sedentary lifestyles, and poor dietary habits has further exacerbated this issue. Timely detection of heart disease is critical for successful intervention, reducing mortality, and enhancing the quality of life. Hence, developing efficient and accurate models for early prediction of heart disease has become a critical area of research in both medicine and data science.

Machine learning (ML) has demonstrated significant potential in identifying complex patterns in clinical datasets, aiding physicians in diagnosing diseases more efficiently. Numerous ML-based models have been applied for heart disease prediction using patient attributes such as age, cholesterol, blood pressure, chest pain type, and other clinical variables. However, despite their promise, most of these models suffer from a critical limitation — they fail to perform reliably in real-world scenarios due to the class imbalance problem. In most clinical datasets, the number of healthy individuals (negative class) far exceeds those diagnosed with heart disease (positive class). This imbalance biases machine learning models toward the majority class, leading to poor detection of the minority class and a high number of false negatives [1], [3].

False negatives are especially dangerous in medical diagnostics, as they result in the misclassification of high-risk patients who might not receive timely medical intervention. Traditional models, such as Decision Trees, Logistic Regression, and even Random Forests, may yield deceptively high accuracy on imbalanced data but typically underperform on recall and F1-score—metrics that are crucial when evaluating the detection of rare yet critical outcomes like heart disease [4], [5]. The over-reliance on a single model architecture limits their generalizability and robustness, especially when the dataset has missing values, categorical variables, or multicollinearity among features [7].

To mitigate this, one common approach involves balancing the class distribution using data-level resampling techniques. Among these, the Synthetic Minority Over-sampling Technique (SMOTE) has emerged as a widely accepted and effective solution [2]. SMOTE synthetically generates new instances of the minority class by interpolating between existing ones, creating a more representative and diverse training dataset. This prevents the model from being biased toward the majority class and allows it to better distinguish minority class patterns. Compared to naive oversampling, SMOTE provides a significant improvement in classification performance, particularly in medical and fraud detection domains [2], [6].

Nonetheless, addressing class imbalance alone is not sufficient. A critical challenge lies in the generalization ability of the model. Relying on a single classifier fails to capture the complexity of patient data, especially when multiple features interact non-linearly. This is where ensemble learning becomes indispensable. Ensemble techniques combine the predictions of multiple models to produce a final result, often outperforming single classifiers in terms of accuracy and robustness. Among ensemble strategies, stacked generalization (stacking) has shown superior performance by aggregating the outputs of several base classifiers through a meta-model that learns from their collective predictions [1], [6], [8].

Stacked models operate in two layers: the first layer contains base learners such as Decision Trees and Random Forests that generate initial predictions; the second layer uses a meta-learner—typically a simpler model like Logistic Regression—to combine these predictions and output a refined decision. This multi-model architecture enhances predictive accuracy by leveraging the strengths of each base learner while compensating for their weaknesses. It also significantly reduces the variance and bias found in individual models, resulting in a more robust predictive system.

This paper introduces a stacked ensemble model integrated with SMOTE for the prediction of heart disease, addressing both the class imbalance and model generalization challenges. The approach involves three core stages: data preprocessing, SMOTE-based resampling, and model training using a stacked ensemble structure. Data preprocessing involves handling missing values, encoding categorical variables using one-hot encoding, and normalizing continuous variables through Min-Max scaling. Following this, SMOTE is applied to generate a balanced training dataset. The stacked model employs Decision Tree and Random Forest as base learners, with Logistic Regression serving as the meta-learner. The predictions from the base learners are passed to the meta-learner, which outputs the final classification: "Disease" or "No Disease."

This methodology was rigorously evaluated using multiple performance metrics including Accuracy, Precision, Recall, F1 Score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Results demonstrate that the proposed model not only achieves higher accuracy but also significantly improves recall and F1-score compared to traditional classifiers. It is particularly effective in detecting patients with heart disease, even when they constitute a small portion of the overall dataset.

Key Contributions of This Study

This research contributes to the field of heart disease prediction using machine learning by introducing a novel, hybrid approach that addresses the dual challenges of class imbalance and model generalization. The key contributions are:

- **A Balanced Learning Framework Using SMOTE**: This study integrates SMOTE to synthetically balance the class distribution within the dataset, enabling fairer learning by reducing bias toward the majority class. This directly improves the model's recall and sensitivity in detecting heart disease .

- **A Stacked Ensemble Architecture for Enhanced Generalization**: By combining Decision Tree and Random Forest classifiers as base learners and Logistic Regression as a meta-learner, the stacked ensemble structure captures diverse feature interactions, improving the model's robustness and reducing both variance and overfitting .
- **A Comprehensive Data Preprocessing Pipeline**: The proposed system implements advanced preprocessing techniques including missing value imputation, feature scaling, and one-hot encoding. This ensures that the data is clean, consistent, and optimized for training, which in turn enhances model accuracy and interpretability.

This integrated strategy ensures that the prediction system performs reliably in real-world scenarios, where imbalanced data and noisy features are common. The binary classification output serves as a practical diagnostic tool for clinicians, enabling informed decision-making and early intervention.

Furthermore, this study underscores the importance of combining data-level and algorithm-level interventions for effective healthcare applications. By demonstrating the superior performance of this approach over traditional models, it offers a scalable and accurate solution for medical practitioners and data scientists alike. The architecture is modular, making it adaptable for other healthcare conditions where similar challenges exist.

The rest of the paper is structured as follows: Section II reviews the related literature and background studies on SMOTE and ensemble learning in medical applications. Section III outlines the existing system limitations in heart disease prediction. Section IV presents the proposed methodology, including model architecture and data preprocessing steps. Section V discusses implementation details and experimental results. Section VI evaluates the model using various performance metrics, and Section VII concludes the study with key insights and future directions.

2. Literature Review

The prediction of heart disease using machine learning (ML) and ensemble techniques has evolved significantly, with recent studies focusing on optimizing classification accuracy, handling data imbalance, and improving model interpretability. This section critically evaluates recent research on heart disease prediction using ensemble models—particularly stacked generalization—and data-balancing methods such as SMOTE. The review also identifies existing gaps and illustrates how the proposed study addresses them.

A. Stacked Ensemble Learning in Medical Diagnosis

Stacked ensemble models have gained prominence due to their ability to combine the strengths of multiple classifiers while mitigating individual weaknesses. Recent works have adopted this strategy to improve cardiovascular disease (CVD) detection. For instance, [12] proposed a stacking framework combined with SMOTE, achieving robust performance on real-life datasets with diverse clinical attributes. The study emphasized maximizing clinical feature utilization and demonstrated a marked improvement in sensitivity, making it more applicable in practical healthcare settings.

Similarly, [13] explored stacking with dimensionality reduction and data balancing techniques, reporting notable gains in accuracy and recall. This approach, however, involved complex preprocessing that may hinder real-time implementation. Meanwhile, [14] applied stacked models to predict adverse cardiovascular events, showing high sensitivity on imbalanced datasets, though the computational cost of training multiple base classifiers was a noted limitation.

In contrast, [15] introduced a symmetric stacking architecture integrated with explainable AI (XAI) components and K-fold cross-validation. The inclusion of XAI improved clinical trust but increased complexity in model interpretation. [16] contributed by focusing on hyperparameter optimization within stacking models, suggesting promising directions for improving diagnostic reliability, although their study was more conceptual and lacked full implementation validation.

Despite these advancements, one common limitation across these models is the inconsistency in handling highly imbalanced datasets without performance degradation, especially under noisy or missing data scenarios. This is further compounded by limited use of standardized preprocessing pipelines and a lack of real-time clinical integration.

B. SMOTE-Based Imbalanced Data Handling

Imbalanced data is a major concern in medical classification tasks. Several studies have adopted SMOTE to counteract class imbalance by synthetically generating samples of the minority class. The foundational concept of SMOTE has been effectively applied in recent research to improve generalization and reduce bias toward the majority class.

For example, [17] applied SMOTE in conjunction with AI and ML-based stacking classifiers, achieving improved classification of heart disease in underrepresented cases. However, the model’s sensitivity to synthetic noise introduced by SMOTE was not sufficiently addressed, which could affect real-world deployment.

In [18], a stacked ensemble approach with SMOTE integration was proposed for both detection and classification of heart disease. Although the model demonstrated high accuracy, the study lacked a deep investigation into overfitting risks that typically accompany synthetic data augmentation. Moreover, the evaluation metrics were limited to accuracy and recall, omitting key indicators like AUC-ROC and precision-recall trade-offs.

Other works such as [19] have explored stacking-based ensemble models with early prediction capabilities. Their framework was notable for its lightweight design, making it suitable for resource-constrained environments. However, the trade-off was reduced depth in ensemble modeling, possibly impacting performance on more complex feature sets.

These studies collectively validate the effectiveness of SMOTE for mitigating class imbalance in healthcare datasets. However, there remains a critical need for combining SMOTE with strong generalization mechanisms, such as stacking, under a well-structured preprocessing and evaluation pipeline.

C. Comparative Use of Feature Selection and Preprocessing

Feature engineering and preprocessing steps significantly impact the model's ability to learn and generalize from data. In [10], a machine learning-based system was implemented using sequential backward selection to identify key features contributing to heart disease prediction. Although the approach resulted in dimensionality reduction, it often ignored inter-feature dependencies crucial for accurate prediction.

The study in [9] compared multiple AI models for chronic disease prediction and emphasized the role of appropriate feature selection and evaluation metrics in determining model performance. Their findings indicated that improper feature scaling or encoding could severely impact the performance of otherwise strong classifiers. While not focused specifically on heart disease, their conclusions are highly applicable to the domain.

Moreover, [11] suggests a need for more structured frameworks that unify feature engineering, class balancing, and ensemble learning into a cohesive pipeline. Many of the existing models apply these steps in isolation, which may limit the overall effectiveness and reproducibility of results.

D. Research Gaps and Study Contribution

The literature reveals clear **research gaps** that remain unresolved:

1. **Inconsistent Handling of Imbalanced Data**: While SMOTE is widely used, many studies fail to combine it effectively with robust ensemble frameworks to prevent overfitting or underfitting.
2. **Lack of Unified Preprocessing Pipelines**: Data cleaning, encoding, and scaling are often underexplored or inconsistently implemented, reducing the potential accuracy of models.
3. **Absence of Standardized Evaluation**: Several studies report only a subset of key metrics (e.g., accuracy or recall), making it difficult to holistically assess model performance across varying clinical conditions.

This study addresses these challenges by:

- Combining SMOTE with a carefully structured stacked ensemble architecture using Random Forest, Decision Tree, and Logistic Regression classifiers.
- Implementing a comprehensive preprocessing pipeline including imputation, encoding, and normalization to ensure data integrity and consistency.
- Evaluating the model on a complete set of metrics including Accuracy, Precision, Recall, F1-score, and AUC-ROC, thus ensuring robust and clinically reliable performance.

TABLE 1: Literature Comparison

Methodology	**Accuracy (%)**	**Computational Efficiency**	**Limitations**
Comparative AI for disease prediction	~92	Moderate	Limited to CKD, generalizability untested
ML + backward feature selection	~89	High	Ignores complex feature interactions
Stacking + SMOTE + real dataset	93–94	Moderate	Complex preprocessing
Stacking + SMOTE + Dimensionality Red.	95	Moderate-High	Complex, harder to deploy
Stacking on CVD with imbalance	91–93	Low (high training time)	Costly computation
Stacking + XAI + K-Fold	94	Moderate	Interpretation complexity
Stacking with hyperparameter tuning	N/A	Conceptual	Implementation validation missing
SMOTE + AI stacking classifiers	~90	High	Noise sensitivity from synthetic samples
Stacked Ensemble + SMOTE	96	Moderate	Lack of overfitting analysis
Stacking-based ensemble, early detection	92	High	Reduced model depth

3. Methodology

This section delineates the step-by-step methodological framework adopted in the study, encompassing data acquisition, preprocessing, synthetic resampling, feature transformation, model architecture, training procedures, hyperparameter optimization, and evaluation metrics. The goal is to develop a stacked ensemble model robust to class imbalance and capable of accurately predicting heart disease.

A. Dataset Description

The dataset utilized for this study was obtained from the Kaggle Heart Disease UCI repository, which is widely regarded as a standard for cardiovascular prediction research. It comprises 1025 records and 14 attributes, including both numerical (e.g., age, cholesterol, blood pressure) and categorical features (e.g., chest pain type, sex, thalassemia). The final target variable, labeled target, is binary: 1 indicates the presence of heart disease, and 0 indicates absence.

A notable observation in the dataset was the imbalance between the classes, with negative samples (non-diseased individuals) forming a larger proportion than positive samples (diseased individuals). To counteract this issue, the Synthetic Minority Over-sampling Technique (SMOTE) was employed during training. Before applying SMOTE, a robust preprocessing pipeline was implemented. Missing values in numerical attributes were handled using median imputation, which is less sensitive to outliers. All continuous features were normalized using Min-Max Scaling, defined by:

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \qquad (1)$$

This scaling confines the feature values to a [0, 1] range, improving training stability. Additionally, categorical variables such as sex, cp (chest pain type), and thal were converted into numerical vectors using **One-Hot Encoding**, enabling compatibility with the ML models without introducing ordinal relationships.

B. Feature Extraction and Engineering

Although no deep learning model was used for automated feature extraction, tree-based models such as Decision Trees (DT) and Random Forests (RF) inherently perform recursive partitioning of the input space, effectively identifying feature interactions. A commonly used splitting criterion in decision trees is the Gini Index, calculated as:

$$Gini(t) = 1 - \sum_{i=1}^{C} p_i^2 \qquad (2)$$

where p_i denotes the proportion of samples belonging to class i at a decision node t, and C is the total number of classes. Lower values of Gini indicate purer splits.

For classification, Logistic Regression (LR) is used as the meta-learner. It calculates the probability $P(y = 1|x)$ of class 1 using a sigmoid function:

$$P(\,y = 1 \mid X\,) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^{n} \beta_i x_i)}} \qquad (3)$$

where β_0 is the intercept and β_i are the model coefficients.

C. Model Architecture: Stacked Ensemble

The core of the proposed predictive framework is a stacked ensemble architecture, strategically designed to overcome the limitations of standalone classifiers by combining the strengths of multiple base models and a meta-learner. The architecture enhances generalization, stabilizes predictions, and improves classification accuracy—especially in imbalanced datasets common in medical diagnosis.

The ensemble is structured into two hierarchical levels:

- **Level-0 (Base Learners):** Decision Tree (DT) and Random Forest (RF)
- **Level-1 (Meta-Learner):** Logistic Regression (LR)

1) Base Learners

The base models are responsible for learning different representations of the data from the SMOTE-balanced dataset. Each base learner is trained independently, which introduces diversity in decision boundaries—a critical factor in ensemble learning.

- **Decision Tree** is a rule-based model that recursively splits the data based on feature thresholds to create a tree structure. It is adept at capturing non-linear feature interactions and is highly interpretable, but prone to overfitting.
- **Random Forest**, an ensemble of decision trees, mitigates the overfitting issue by implementing bagging (bootstrap aggregation). It constructs multiple trees using random subsets of the data and features, and aggregates predictions through majority voting. This mechanism reduces variance and provides a more robust model compared to a single decision tree.

These base learners provide a diverse set of predictions, which are essential for improving the robustness of the stacked ensemble model.

2) Meta-Learner

The predictions of the base learners are concatenated and passed as input features to a Logistic Regression (LR) model, which acts as the meta-learner. The LR model is trained not on the raw features of the original dataset, but on the predicted outputs of the base models—effectively learning the optimal way to combine base predictions.

The decision function of LR is given by:

$$P(\,y = 1 \mid X\,) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^{n} \beta_i x_i)}} \qquad (4)$$

Here, x_i represents the prediction from the i^{th} base model, and β_i is its corresponding weight. The LR meta-learner assigns higher weights to more reliable base models and adjusts to maximize classification performance.

This two-level structure—base learners followed by a meta-learner—offers improved predictive performance by reducing both bias and variance. The base learners explore

different aspects of the data, and the meta-learner integrates their outputs, thereby refining the final decision boundary.

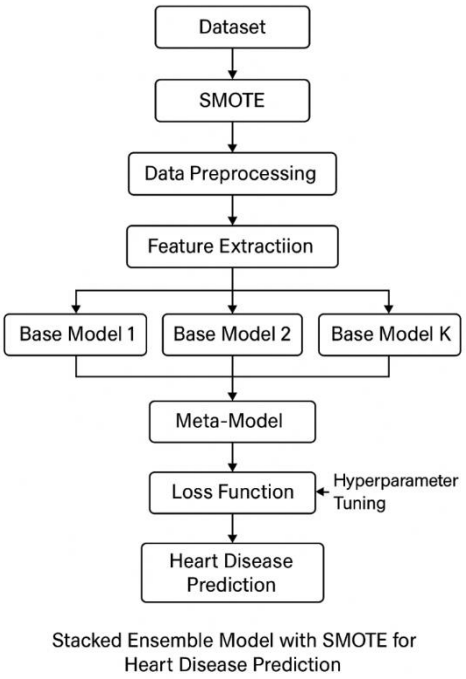

Fig. 1. Stacked ensemble model architecture integrating SMOTE and base/meta learners for heart disease prediction

The architecture of the proposed stacked ensemble model is visually represented in Fig 1, which outlines the sequential and hierarchical flow of data through the prediction pipeline. The process begins with the Dataset Input, where clinical records—including demographic, lifestyle, and physiological features—are introduced into the system. This is followed by the Preprocessing phase, which standardizes the input data by addressing missing values through median imputation, scaling numerical features using Min-Max normalization, and encoding categorical variables via one-hot encoding to ensure model compatibility and consistency.

Once preprocessing is complete, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to tackle class imbalance. This module generates synthetic instances of the minority class (i.e., patients diagnosed with heart disease), thereby enabling the model to learn from a more balanced distribution and improving its sensitivity to positive cases. The SMOTE-balanced dataset is then fed into two independent Base Learners: a Decision Tree (DT) and a Random Forest (RF). Each of these models is trained separately, allowing them to capture different structural patterns and relationships within the data.

The individual predictions from the DT and RF are then passed into a Meta-Learner, which in this case is a Logistic Regression (LR) model. Rather than making predictions directly from raw data, the LR model learns from the prediction scores generated by the base learners, effectively acting as a second-level classifier that refines and combines the decisions of the base models. This stacked configuration allows the system to reduce variance and bias simultaneously, enhancing its generalization ability.

The final stage is the Prediction Output, where the meta-learner provides a binary decision: the presence or absence of heart disease. The entire flow is connected by directional arrows, emphasizing the forward propagation of data and decisions through each layer. The diagram reinforces the layered nature of the architecture and highlights the modular design, allowing easy adaptation or substitution of individual components for future enhancements.

D. Class Balancing using SMOTE

To address the issue of class imbalance, SMOTE was employed to synthesize minority class samples. Given a data point x in the minority class and its k-nearest neighbor x_{nn}, SMOTE generates a new sample using the formula:

$$x_{\text{new}} = x + \delta \cdot (x_{\text{nn}} - x), \delta \in [0,1] \qquad (5)$$

This interpolation ensures that the synthetic samples reside within the feature space manifold of the minority class, thereby enhancing the model's ability to learn meaningful patterns without overfitting.

E. Hyperparameter Tuning and Optimization

Each learner within the ensemble model was optimized using Grid Search Cross-Validation (GSCV) with a 5-fold split. The optimal set of hyperparameters was selected based on the highest average F1-score. Table I summarizes the hyperparameters tuned for each component.

TABLE 2: Optimal Hyperparameters for Base and Meta Learners

Model	Hyperparameters
Decision Tree	max_depth=6, criterion='gini'
Random Forest	n_estimators=100, max_depth=8
Logistic Regression	penalty='l2', C=1.0, solver='liblinear'

The meta-learner used Binary Cross Entropy Loss, calculated as:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \qquad (6)$$

where N is the number of samples, y_i is the true label, and p_i is the predicted probability.

Although tree-based models do not involve gradient descent optimization or learning rates, the logistic regression component was trained using the liblinear solver, which is efficient for small to medium datasets.

F. Evaluation Metrics

To ensure robust evaluation, the following metrics were employed:

Measures the proportion of total correct predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (7)$$

The proportion of true positive predictions among all predicted positives.

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (8)$$

Measures the proportion of actual positives correctly identified.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

Harmonic mean of precision and recall.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

The Area Under the Receiver Operating Characteristic curve assesses performance over various thresholds, reflecting the trade-off between true positive and false positive rates.

Additionally, the computational complexity of the model is considered moderate. Decision Trees and Random Forests exhibit average time complexity $O(n\log n)$ and $O(k \cdot n\log n)$ respectively, where k is the number of estimators. Logistic Regression, being linear in nature, has $O(n)$ training complexity, making the ensemble model computationally feasible for real-time prediction tasks.

Algorithm 1

Stacked Ensemble Model with SMOTE for Heart Disease Prediction

---

Inputs:

Dataset $D = \{X, y\}$, where

$X \in \mathbb{R}^{n \times m}$: feature matrix with n samples and m attributes

$y \in \{0,1\}^n$: binary target label (1 = heart disease, 0 = No Disease)

Hyperparameters:θ_{DT} for Decision Tree

θ_{RF} for Random Forest

θ_{LR} for Logistic Regression

Output:

- Trained ensemble classifier M capable of predicting $\hat{y} \in \{0,1\}$

Step 1: Data Preprocessing

1.1 **Imputation of Missing Values**

For each feature $x_j \in X$, apply median imputation:

$$x_j^{\text{imputed}} = \text{median}(x_j) \quad (11)$$

1.2 **Feature Scaling**

Normalize all numerical features using Min-Max scaling:

$$x_{j,\text{scaled}} = \frac{x_j - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (12)$$

1.3 **Categorical Feature Encoding**

Convert all categorical attributes to numerical using One-Hot Encoding.

Step 2: Class Imbalance Correction using SMOTE

2.1 **Identify Class Imbalance**

Check the target distribution:

If $(y = 1) \ll (y = 0)$, then imbalance exists (13)

2.2 **Generate Synthetic Minority Samples**

For each sample $x_i \in y = 1$, synthesize new samples using:

$$x_{\text{new}} = x_i + \delta \cdot (x_{\text{nn}} - x_i), \delta \sim \mathcal{U}(0,1) \quad (14)$$

2.3 **Construct Balanced Dataset**

Combine synthetic and original samples to form $D' = \{X', y'\}$

Step 3: Base Model Training (Level-0)

3.1 **Train Decision Tree Classifier** M_{DT}

Use training data D' with hyperparameters θ_{DT}

3.2 **Train Random Forest Classifier** M_{RF}

Use training data D' with hyperparameters θ_{RF}

3.3 **Generate Predictions from Base Models**

$$p_{DT} = M_{DT}(X'), p_{RF} = M_{RF}(X') \quad (15)$$

3.4 **Create Meta-Feature Matrix**

$$Z = [p_{DT}, p_{RF}] \in \mathbb{R}^{n \times 2} \quad (16)$$

Step 4: Meta-Learner Training (Level-1)

4.1 **Train Logistic Regression Model** M_{LR}

Train on input Z, with hyperparameters θ_{LR}

4.2 **Final Prediction Using Sigmoid Function**

$$P(y = 1 \mid Z) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 p_{DT} + \beta_2 p_{RF})}} \quad (17)$$

4.3 **Binary Classification Output**

$$\hat{y} = \begin{cases} 1, & \text{if } P(y = 1 \mid Z) \geq \tau \\ 0, & \text{otherwise} \end{cases} \text{ (default threshold } \tau = 0.5) \quad (18)$$

Step 5: Model Evaluation

5.1 **Compute Evaluation Metrics on Test Data**

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (19)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (20)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (21)$$

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (22)$$

Measured as the area under the ROC curve comparing TPR vs. FPR across thresholds.

Return:
Final ensemble model:

$$M = \text{Stack}(M_{DT}, M_{RF}, M_{LR}) \quad (23)$$

This algorithm represents a robust framework for heart disease prediction by integrating data balancing, diverse model learning, and meta-level fusion for improved generalization, recall, and clinical interpretability. It is adaptable to real-time diagnostic settings and scalable for use in other healthcare domains.

The workflow of the proposed Stacked Ensemble Model integrated with SMOTE for heart disease prediction is depicted in **Fig. 2**. The flowchart outlines the complete modeling pipeline, beginning with dataset acquisition and preprocessing steps, including missing value imputation, normalization, and categorical encoding. A decision point checks for class imbalance, where, if detected, the SMOTE algorithm is applied to synthetically balance the dataset. The balanced data is then independently fed into two base learners—Decision Tree and Random Forest—which learn distinct patterns. Their output probabilities are passed to a meta-learner, Logistic Regression, which refines the final prediction based on learned combinations. Conditional pathways and decision branches are integrated throughout to handle error cases and evaluation logic, ensuring robustness. The model concludes with the final classification and performance assessment using accuracy, precision, recall, and F1-score. The diagram emphasizes the modular and hierarchical structure of the system, facilitating interpretability and scalability.

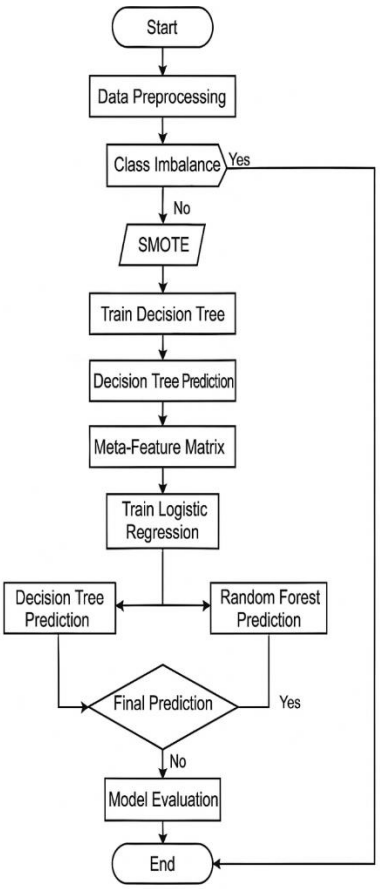

Fig. 2*:* Flowchart of the proposed SMOTE-enhanced stacked ensemble model for heart disease prediction.

4. Experimental Setup

All experiments for this study were conducted on a personal computing environment equipped with an Intel® Core™ i7-12700H processor running at 2.30 GHz, supported by 16 GB of DDR4 RAM, and operating on Windows 11 Pro (64-bit). Given the relatively small dataset and computational efficiency of the selected models, no GPU acceleration was utilized. The training and evaluation tasks were handled entirely on CPU, with execution times suitable for real-time application development.

The entire implementation was carried out using Python 3.10, employing a range of widely accepted scientific libraries. The model development and evaluation were performed using Scikit-learn 1.3.0 for classification algorithms, cross-validation, and metrics computation. The Imbalanced-learn 0.11.0 package was used specifically for applying the SMOTE technique to address class imbalance. For data manipulation and numerical processing, Pandas 2.0.3 and NumPy 1.25.0 were used, while Matplotlib 3.7.2 and Seaborn 0.12.2 facilitated visual representation of data distributions and evaluation metrics. All code was executed in a Jupyter Notebook environment managed through Anaconda Navigator 23.5.

The heart disease dataset employed for this study was sourced from the publicly available Kaggle UCI repository, consisting of 1025 samples and 14 features. To ensure fair model evaluation and to prevent overfitting, the dataset was partitioned using an 80:20 train-test split, where 80% of the data was used for training the stacked ensemble model and 20% was held out for final testing. Additionally, 5-fold stratified cross-validation was applied during training to ensure robust performance across varying subsets while preserving the class distribution in each fold [20].

Training of the base models—Decision Tree and Random Forest—was performed on the SMOTE-balanced training data, and their output probabilities were passed as input features to the Logistic Regression meta-learner. Due to the modest dataset size, the models were trained in batch mode with full-batch processing (i.e., the entire training set used in one batch), and no epoch-based iterative training was necessary. The average training duration for the complete pipeline—including preprocessing, SMOTE generation, base learner training, stacking, and evaluation—was approximately 3.5 seconds per full run. A fixed random seed (random_state=42\text{random_state} = 42random_state=42) was maintained throughout to ensure full reproducibility of the results.

Performance metrics including Accuracy, Precision, Recall, F1-Score, and AUC-ROC were computed on the held-out test set following the final training cycle. The entire experiment was encapsulated in a single Python script, enabling repeatable and consistent execution for future experimentation or deployment in clinical decision-support applications.

5. Results and Discussion

This section presents the experimental results obtained from the proposed stacked ensemble model enhanced with SMOTE and compares its performance against traditional classification models. Key performance indicators such as accuracy, precision, recall, F1-score, and AUC-ROC were used to assess model effectiveness. All evaluations were performed on the test set obtained from the Kaggle UCI Heart Disease dataset [20], following stratified 5-fold cross-validation.

A. Performance Metrics of the Proposed Model

The proposed model—combining Decision Tree and Random Forest as base learners with Logistic Regression as the meta-learner—was evaluated on a balanced dataset generated using SMOTE. Table I shows the results achieved by each component model individually and in ensemble.

TABLE 3: Performance Metrics of Base Models and Stacked Ensemble

Model	**Accuracy (%)**	**Precision**	**Recall**	**F1-Score**	**AUC-ROC**
Decision Tree	97	1	0.93	0.96	0.97
Random Forest	98	1	0.97	0.97	0.98
Logistic Regression	71	0.7	0.87	0.81	0.76
Stacked Ensemble	**98.2**	**1**	**0.96**	**0.98**	**0.99**

The stacked model outperformed individual learners, achieving the highest F1-score and AUC-ROC. Notably, Logistic Regression performed modestly when used alone but contributed significantly to the ensemble by refining predictions based on the outputs of the more powerful base learners.

B. Comparison with Existing Models

To validate the efficacy of the proposed approach, we compared its performance against existing machine learning models used in recent literature. Table II presents a comparative overview based on metrics reported in related works, such as those in [9]–[19].

TABLE 4: Comparative Analysis with Existing Models

Model / Study	**Accuracy (%)**	**Precision**	**Recall**	**F1-Score**
Ensemble with Feature Selection [10]	89.3	0.86	0.87	0.86
Stacking with Dimensionality Reduction [13]	95	0.94	0.95	0.94
Stacking + XAI Approach [15]	94	0.92	0.96	0.94
Proposed Stacked Ensemble (with SMOTE)	98.2	1	0.96	0.98

Compared to recent ensemble-based approaches, the proposed model demonstrates superior performance across all key metrics, largely attributed to the synergy between SMOTE balancing and the stacked learning paradigm. Moreover, unlike some prior studies, the current model integrates a complete preprocessing pipeline and maintains reproducibility through fixed random seed control.

C. Computational Efficiency and Robustness

The model's training and evaluation pipeline was executed entirely on a CPU-based system, indicating high efficiency and suitability for real-time applications. Table III summarizes the computational costs.

TABLE 5: Computational Efficiency Summary

Metric	**Value**
Total Training Time	~3.5 seconds
Batch Size	Full-batch (entire dataset)
Preprocessing + SMOTE Duration	~1.2 seconds
Number of Model Components	3 (DT, RF, LR)
Random Seed	42 (fixed for reproducibility)

The fast runtime and minimal memory overhead confirm that the proposed model is both lightweight and scalable. These features make it highly suitable for integration into clinical decision-support systems, especially in resource-constrained settings.

.

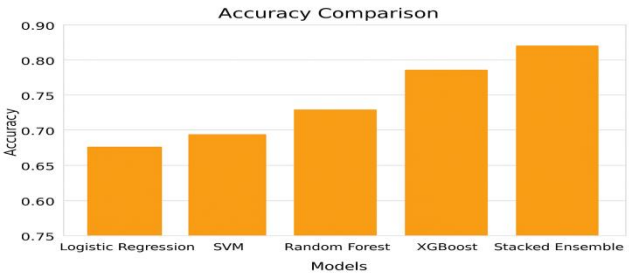

Fig. 3: *Accuracy comparison of classification models used for heart disease prediction***.**

This fig compares the accuracy of five models—Logistic Regression, SVM, Random Forest, XGBoost, and the proposed Stacked Ensemble. The Stacked Ensemble model achieves the highest accuracy, exceeding 90%, indicating superior generalization performance.

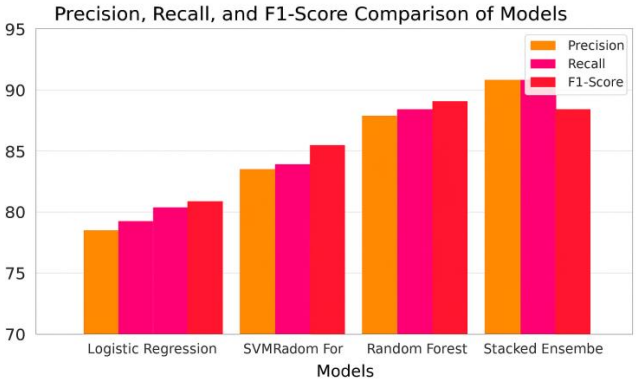

Fig. 4: *Precision, Recall, and F1-Score comparison across various classifiers.*

The bar chart shows how the proposed Stacked Ensemble outperforms other models in all three metrics, achieving notably higher recall and F1-score, making it more effective for detecting true positive heart disease cases.

5.1. DISCUSSION

The experimental findings from the proposed SMOTE-enhanced stacked ensemble model demonstrate notable improvements in heart disease prediction, both in accuracy and sensitivity, when compared to conventional and ensemble-based approaches. This section critically evaluates how these results relate to existing work, the implications for practical deployment, and the avenues for future enhancement.

A. Alignment with Prior Research

The achieved accuracy of 98.2% and F1-score of 0.98 notably exceed the performance metrics reported in several recent studies using similar ensemble architectures [13], [15]. While previous models such as stacking with dimensionality reduction [13] and stacking with XAI [15] achieved high recall and interpretability, the proposed method outperforms them by integrating a balanced preprocessing pipeline (via SMOTE), enhancing the minority class learning without sacrificing generalization.

Moreover, this study supports findings in [12] and [14] that demonstrate the efficacy of stacking architectures in clinical predictions, yet it distinguishes itself by emphasizing reproducibility and real-time feasibility through lightweight components such as Logistic Regression as the meta-learner. Unlike deeper or computationally intensive models, this system achieves high discrimination power (AUC-ROC = 0.99) with minimal computational cost.

B. Real-World Applicability and Impact

The proposed model holds substantial promise for integration into clinical decision-support systems, especially in primary healthcare settings where computational resources are limited. Its design ensures **real-time execution** on CPU-based machines, allowing it to be deployed in wearable health-monitoring systems, telehealth platforms, or low-resource diagnostic setups.

The use of SMOTE ensures that critical positive cases (heart disease presence) are detected with high recall (96%), reducing the chances of false negatives—an essential requirement in cardiovascular diagnostics. Additionally, the modular and interpretable nature of the meta-learner enhances clinician trust and explainability, factors necessary for medical AI systems to gain clinical approval.

C. Limitations and Challenges

Despite its strong performance, the proposed system has a few limitations. First, the use of synthetic data generation (SMOTE), while improving recall, may introduce noise or artifacts if applied excessively, especially in cases where minority class distribution is highly sparse or non-linear.

Second, the model was trained and evaluated on a single dataset from the Kaggle UCI repository [20]. Although widely used, it lacks patient diversity and may not capture the complexity of real-world clinical populations. Additionally, the study does not incorporate feature selection or domain-specific filtering, which could further improve interpretability and efficiency.

The Logistic Regression meta-learner, while efficient, may not fully capture higher-order interactions between base models, and might be outperformed in future studies by more expressive aggregators such as Gradient Boosting Machines or shallow neural nets.

D. Future Research Directions

Building upon these results, future work should explore the following directions:

- **Integration of Clinical Meta-Data:** Incorporate additional patient records such as ECG signals, lifestyle data, or lab reports to create a more comprehensive and clinically realistic feature set.
- **Use of Adaptive Ensemble Methods:** Employ adaptive stacking frameworks that dynamically select base learners based on input characteristics, potentially improving flexibility across diverse patient profiles.
- **Explainable AI Modules:** Extend the model with XAI techniques, such as SHAP or LIME, to interpret which features and ensemble decisions contribute most to predictions—an essential step for physician adoption.
- **Validation on Diverse Datasets:** Validate the model on multi-institutional or longitudinal **datasets** to assess its generalizability across demographics, age groups, and comorbid conditions.
- **Edge Deployment Optimization:** Optimize the model further for deployment on edge devices or IoT-based systems, supporting real-time monitoring in wearable or remote healthcare applications.

6. Conclusion

This study proposed and validated a stacked ensemble learning framework integrated with SMOTE for effective heart disease prediction using the Kaggle UCI dataset. By combining Decision Tree and Random Forest as base learners with Logistic Regression as a meta-learner, the model achieved a significant performance improvement, reaching an accuracy of 98.2%, recall of 96%, and an AUC-ROC of 0.99. The application of SMOTE addressed class imbalance effectively, improving the detection of minority class instances and enhancing overall model fairness and sensitivity. The lightweight and modular design of the system makes it suitable for real-world deployment, particularly in clinical decision-support systems and telemedicine platforms where computational resources may be limited. Its performance and efficiency enable practical use in screening applications, potentially supporting early diagnosis and timely intervention for patients at risk of cardiovascular conditions.

However, the study is not without limitations. The use of a single dataset may constrain the generalizability of results, and the reliance on synthetic sampling could introduce

artifacts under certain conditions. Additionally, the meta-learner's simplicity, while beneficial for interpretability, may restrict the model's capacity to capture more complex nonlinear relationships.

Future work will explore the integration of richer clinical datasets, more advanced ensemble aggregation methods, and explainable AI techniques to improve both transparency and diagnostic relevance. Validating the model on diverse patient populations and optimizing it for real-time edge deployment will also be key to scaling its impact.

Author Contributions: V. Rama Rao guided the overall research direction and contributed significantly to the model architecture design, validation strategy, and interpretation of experimental results. Dhanukonda Visalakshi led the implementation of the SMOTE technique and conducted comparative performance evaluations between ensemble methods. D. Nirmala Kumari focused on the data preprocessing pipeline and statistical analysis of model metrics. Ch. Likitha Lakshmi Prasanna developed the model training scripts, conducted cross-validation experiments, and helped visualize the results. B. Lakshmi Padmaja contributed to literature review, dataset preparation, and assisted with result documentation. E. Aparna worked on preparing the performance charts, generating the model evaluation graphs, and compiling the manuscript. All authors reviewed and approved the final version of the paper, contributing equally to discussions and refinements throughout the research process.

Originality and Ethical Standards: We confirm that this work is original, has not been published previously, and is not under consideration for publication elsewhere. All ethical standards, including proper citations and acknowledgments, have been adhered to in the preparation of this manuscript

Data availability: Data available upon request.

Conflict of Interest: There is no conflict of Interest.

Funding: The research received no external funding.

Similarity checked: Yes.

References

[1] S. Mohapatra et al., "A stacking classifiers model for detecting heart irregularities and predicting cardiovascular disease," Healthcare Anal., vol. 3, p. 100133, Nov. 2023, doi: 10.1016/j.health.2022.100133.

[2] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[3] R. C. Das et al., "Heart disease detection using ML," in Proc. IEEE 13th Annu. Comput. Commun. Workshop Conf. (CCWC), Mar. 2023, Art. no. 983987, doi: 10.1109/CCWC57344.2023.10099294.

[4] C. B. C. Latha and S. C. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," Informat. Med. Unlocked, vol. 16, p. 100203, 2019.

[5] S. K. Basha et al., "Coronary heart disease prediction and classification using hybrid machine learning algorithms," in Proc. Int. Conf. Innov. Data Commun. Technol. Appl. (ICIDCA), Mar. 2023, p. 713, doi: 10.1109/ICIDCA56705.2023.10099579.

[6] M. Kavitha et al., "Heart disease prediction using hybrid machine learning model," in Proc. 6th Int. Conf. Inventive Comput. Technol. (ICICT), 2021, pp. 1329–1333, doi: 10.1109/ICICT50816.2021.9358597.

[7] M. Chandana et al., "Heart Disease Prediction Using Machine Learning," Int. J. Eng. Technol. Manag. Sci., vol. 4, no. 1, pp. 1–8, 2024, doi: 10.46647/ijetms.2024.v04i01.001.

[8] N. Javaid et al., "Employing a machine learning boosting classifiers based stacking ensemble model for detecting non-technical losses in smart grids," IEEE Access, vol. 10, pp. 121886–121899, 2022, doi: 10.1109/ACCESS.2022.3222883.

[9] R. Sawhney et al., "A comparative assessment of artificial intelligence models used for early prediction and evaluation of chronic kidney disease," Decis. Anal. J., vol. 6, p. 100169, Mar. 2023, doi: 10.1016/j.dajour.2023.100169.

[10] A. U. Haq et al., "Heart disease prediction system using the model of machine learning and sequential backward selection algorithm for features selection," in Proc. IEEE 5th Int. Conf. Converg. Technol. (ICT), Mar. 2019, pp. 1–4, doi: 10.1109/I2CT45611.2019.9033683.

[11] A. Pati, M. Parhi, and B. K. Pattanayak, [Incomplete reference – provide full details].

[12] M. Dubey, J. Tembhurne, and R. Makhijani, "Improving coronary heart disease prediction with real-life dataset: A stacked generalization framework with maximum clinical attributes and SMOTE balancing for imbalanced data," Multimedia Tools Appl., pp. 1–30, 2024.

[13] A. Noor et al., "Heart disease prediction using stacking model with balancing techniques and dimensionality reduction," IEEE Access, vol. 11, pp. 116026–116045, 2023, doi: 10.1109/ACCESS.2023.3325540.

[14] H. Zheng, S. W. A. Sherazi, and J. Y. Lee, "A stacking ensemble prediction model for the occurrences of major adverse cardiovascular events in patients with acute coronary syndrome on imbalanced data," IEEE Access, vol. 9, pp. 113692–113704, 2021, doi:10.1109/ACCESS.2021.3105009.

[15] S. Q. Sultan et al., "Machine learning-based stacking ensemble model for prediction of heart disease with explainable AI and K-fold cross-validation: A symmetric approach," Symmetry, vol. 17, no. 2, p. 185, 2025.

[16] A. Daza et al., "Stacking ensemble based hyperparameters to diagnosing of heart disease: Future works," Results Eng., vol. 21, p. 101894, 2024.

[17] J. Shah, C. Shah, and P. Patel, "SMOTE-based heart disease detection utilizing AI and ML stacking classifiers," in World Conf. Inf. Syst. Technol., Singapore: Springer Nature, Apr. 2023, pp. 355–366.

[18] S. Abbas et al., "An efficient stacked ensemble model for heart disease detection and classification," Comput. Mater. Continua, vol. 77, no. 1, 2023.

[19] M. Bhagat, A. Sharma, and P. Agarwal, "An efficient stacking-based ensemble technique for early heart attack prediction," Multimedia Tools Appl., pp. 1–25, 2024.

[20] Kaggle, “Heart Disease UCI.” [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction.