SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I6P126

Volume 12 Issue 6, 326-339, June 2025

Original Article

An Optimized Ensemble Learning Framework using Boosted Random Subspace SVMs and Stacked Generalization for Cardiovascular Disease Diagnosis

J. Raghunath[1], S. Kiran[2]

[1,2]*Department of CSE, YSR Engineering College of YVU, Andhra Pradesh, India.*

[2]*Corresponding Author : rkirans125@gmail.com*

Received: 14 April 2025 Revised: 15 May 2025 Accepted: 17 June 2025 Published: 27 June 2025

Abstract *- Cardiovascular disease causes most deaths worldwide, so health organizations seek to produce dependable automated medical diagnosis tools. Our proposed method combines a Support Vector Machine with boosted random subspace and stacked generalization to make more accurate CVD diagnosis predictions. The framework begins by normalizing data as part of preprocessing to normalize feature scales from clinical data. Several SVM-based learners gain training via the diverse feature subsets that the Random Subspace Method (RSM) generates. The learners achieve optimized kernel parameters by performing a grid search optimization. The voting scheme in bootstrap aggregation methods improves diversity while controlling overfitting to generate predictions. The model generalization requires stacked generalization that integrates base learner outputs into a second-level logistic regression prediction system. The assessment method involves checking accuracy rates together with precision values and recall rates and includes F1-score and Area Under the Receiver Operating Characteristics Curve (AUC) measurements. Experimental benchmark results validate that the ensemble model reaches an accuracy rate of 96.39%, surpassing standard single classifiers together with standard ensemble techniques in predicting heart disease, thus proving its clinical value for cardiovascular assistance. The proposed diagnostic framework demonstrates strength and expandability for medical diagnosis procedures that require outstanding interpretive capabilities along with specific prediction accuracy.*

Keywords *- Accuracy, Boosted SVM, Cardio Vascular Disease, F1-score, Precision, Random subspace, Recall.*

1. Introduction

Cardiovascular Disease (CVD) represents different heart and blood vessel disorders that encompass coronary artery disease together with heart failure, arrhythmias and stroke as major conditions. The development of such diseases results from either atherosclerosis-induced fat accumulation in arteries or blood clot formation that reduces vital organ blood supply [1]. The main dangers for CVD development consist of hypertension alongside diabetes mellitus as well as tobacco usage together with obesity and insufficient exercise. Doctors use clinical evaluation and medical history together with physical examinations in addition to tests such as electrocardiograms (ECG), echocardiography, cardiac biomarkers, stress tests and imaging techniques like CT or MRI for diagnosing CVD. The medical staff needs to identify CVD early because timely diagnosis enables proper treatment approaches and stops health complications from developing.

Cardiovascular disease leads to more human deaths than any other health condition worldwide each year through heart attacks and strokes combined [2]. Most CVD-related deaths take place across low- and middle-income nations, where these health problems generate 80% of total fatalities. Stroke and ischemic heart disease cause most deaths from cardiovascular diseases. Although doctors have made medical progress to fight CVD, doctors treat more patients who die because of heart disease. According to research, more worldwide focus on early cardiovascular risk factor control and disease detection is needed to manage heart disease challenges globally [3]. Detecting CVDs at their first stages helps stop disease progression and improves patient outcomes. Clinical data sets have natural barriers including their difficulty to analyze and their high level of specificity along with measurement errors. Ensemble-based ML tools have become popular because they demonstrate clear potential to build dependable automated diagnostic methods [4, 5].

1.1. Research Challenges

Multiple obstacles exist in Cardiovascular Disease (CVD) detection, affecting the speed of diagnosis and

appropriate treatment delivery process. Early-stage CVD shows minimal symptoms in most cases, which delays early diagnosis until serious medical conditions already happen [6]. The assessment of patients becomes challenging because patient symptoms show diverse patterns that depend on their age and sex combined with any coexisting medical conditions. Early CVD detection remains challenging because low-resource areas struggle to obtain echo and cardiac imaging diagnostic instruments. Data interpretation from ECGs demands specialist expertise that health facilities may lack completely. The combination of multiple clinical data types presents technical barriers, particularly during computerized detection that utilizes artificial intelligence and machine learning programming. Better diagnostic techniques are essential to provide both accuracy and accessibility, as well as flexibility in various healthcare facilities.

1.2. Problem Statement

Cardiovascular diseases, or CVD, continue to be the cause of maximum mortality in the world, and this burden is very high in those developing parts of the world where the accessibility of highly developed medical infrastructure is low. The identification of CVD at an early stage is important, but it is usually retarded owing to the non-specific or faint nature of the symptoms found in its early stages hence the late drug intervention. The standard diagnosis tools, which include Electrocardiograms (ECG) and cardiac imaging, need to be interpreted by specialists and are often unavailable and impractical to use in low-resource care environments. This increases the loss of opportunities for early intervention and deteriorates patient outcomes.

The increasing inequality in access to diagnosis can be evidence of the necessity to develop a cost-effective, automated, and scalable diagnostic system capable of providing clinicians with the tools and operating efficiently in various healthcare settings. The latest technologies, such as Artificial Intelligence (AI) and Machine Learning (ML), provide effective coping strategies, as they allow for automatizing the analysis of clinical data, better diagnosis, and reducing human-made mistakes [7]. More specifically, Support Vector Machines (SVMs) have exhibited good generalization behaviour and robustness in the presence of high-dimensional vectors and, therefore, could serve well in medical diagnosis. Nevertheless, the SVM models [8] themselves tend to reduce their predictive power when analysing heterogeneous, imbalanced, or dirty clinical data, which are common issues facing real-life clinical use [9].

To address these shortfalls, ensemble learning methods have been suggested as the new alternative to combine several base classifiers to achieve higher levels of accuracy in classification variations and serve to meet greater confidence in the predictions. As such, it is evident that a research gap exists to generate and test smarter, ensemble-oriented models of risk because of the profound impact better CVD diagnostics integrated with SVM and other methods can have in making accurate and meaningful diagnosis possible that can be reached sooner and more conveniently than currently is possible, at least in resource-poor environments. The solution to this issue can lead to a considerable decrease in CVD-related deaths and enhancement of the health outcomes of the world population [10].

This work proposes an optimal Ensemble Learning framework that combines the random subspace feature set selection and Stacking approaches to use Boosted Random Subspace SVMs for the diagnosis of cardiovascular diseases. The framework exists in three layers.

1. Multiple diverse SVM classifiers are generated by training each one on a randomly sampled subset of features using the random subspace method.
2. It uses boosting to iteratively reweight samples by directing later classifiers to the misclassified instances in order to minimize bias and variance.
3. Stacking is employed in which a meta-learner is trained to best combine the outputs of the ensemble members, which incorporate higher-order dependencies among base learners.

Additionally, Bayesian Optimization automatically picks hyperparameters at the base learner and meta learner levels to avoid tedious, laborious manual tuning and perform best every time. The primary contributions of this study are summarized as follows.

1. This research develops an ensemble system that combines Random Subspace SVMs, boosting, and stacking methods to handle difficult distributions of CVD dataset clinical information.
2. The model uses Bayesian Optimization to find perfect settings and strengthens the performance of the entire ensemble system.
3. The framework proves better at medical diagnosis by showing enhanced results across all evaluation metrics on cardiovascular datasets through statistical tests.

The next parts of this study follow this structure. Section 2 examines existing research about using ensemble techniques for CVD diagnosis. Section 3 explains our new system design and explains how to optimize the model using mathematical methods. Section 4 presents information about data sources, model results, and expert evaluations. Section 5, the end of this research work, presents the conclusion and directions for upcoming scientific studies.

2. Related Work

The research into cardiovascular disease diagnosis benefits greatly from a literature review because it creates an understanding of existing conditions and unveils medical and technical limitations. The review establishes knowledge about risk factors, diagnostic tools, and treatment procedures

yet points out shortcomings with present methods because of limited sensitivity and restricted use cases. The previously conducted research helps determine appropriate datasets and features and machine learning methods for methodological validity. Knowledge gained from the literature review enables researchers to develop research hypotheses that rely on established evidence and enable meaningful model benchmarking. The combination of exhaustive background research strengthens both the quality and academic integrity of research, making it stronger for publication.

Jayaraman M and Pichai S [11] research combines several learning methods to increase accuracy in predicting cardiovascular disease onset. Using 70,000 records, the authors first filtered out damaged data records via a box plot algorithm for outlier detection. The dataset needed processing, so we split it into training and testing data sets. The authors tested a group of basic recognition tools such as Support Vector Machines (SVM), Decision Trees (DT), and Random Forests (RF). The combined model showed outstanding results at 88.39% by correctly recognizing cardiac heart events while avoiding incorrect assessments for both types of data. This study proves that combining different training methods can precisely forecast when patients will develop cardiovascular diseases in medical settings. The successful combination of effective dataset preparation and multiple learning methods improves our ability to make accurate diagnoses at the beginning of treatment planning.

Ganie and colleagues [12] developed an ensemble learning approach to boost diagnosis accuracy of heart diseases. The study team applied Gradient Boosting XGBoost and AdaBoost boosting methods to analyze heart disease features from the UCI Machine Learning repository. The core actions of my process are to format the data and train the model to check results. The dataset underwent exploratory analysis to discover and deal with missing data by imputation, and then the interquartile range was employed to detect and replace outliers.

The filtered data is divided into 70 percent training and 30 percent testing data. We evaluated all boosting models by splitting samples into 10 sections for testing and training while repeating the process 10 times. The Gradient Boosting method delivered 92.20% success, which exceeded the results of XGBoost and AdaBoost. The Gradient Boosting model outperformed other methods in recognizing heart disease cases accurately, while its detection measures performed well for positive and negative disease samples. The authors believe this hybrid system can recognize various illnesses effectively and suggest future work on transferred learning technologies.

Tiwari A., together with Chugh and colleagues [13], present a reliable method for improving our ability to predict cardiovascular diseases. The authors used all available datasets from IEEE Data Port that combined data from Hungary, Cleveland, Long Beach, VA, Switzerland, and Statlog. The research team used four separate feature selection methods, LASSO and Relief, in particular, to find the best disease predictors. Afterwards, we determined the most helpful elements using Chi-Square statistical values and p-values.

The model designer combined ExtraTrees Classifier, Random Forest, and XGBoost in a stacked ensemble classifier. This approach combines various models with their strengths to get better forecasting results as a whole. The proposed model's performance results reveal its accuracy, precision, recall, F1-score, specificity, sensitivity, MCC, and AUC-ROC scores. The combined system delivered an excellent prediction result of 92.34% while demonstrating better performance than any previous studies.

Sharma et al. [14] developed a study using classification and ensemble methods for heart disease diagnosis while improving medical choice-making at the health centre. The model used several basic machine learning algorithms for weak learners to build a dependable predictive system. The process groups different basic low-performing classifiers to build an accurate and precise enhanced model. The model uses initial medical signs to detect heart disease at the beginning before problems become severe. Grouping several forecasting models together increased heart disease diagnosis accuracy better than single-classification models. Using several classification methods strengthened our predictive scheme.

In a research project, Wenhao Chi and colleagues [15] examined how MTBO speeds up hyperparameter selection in SVM classifiers when detecting pulmonary nodules. The author used MTBO to make hyperparameter tuning in SVM more efficient to speed up pulmonary nodule screenings. This research project followed a specific set of steps to accomplish its tasks. CT scans of lungs required preprocessing to detect their nodules before applying nine different image classification methods across different numbers of bins and quantization values. Radiomic features of nodule image data were generated using different shape measurements plus statistical values and texture patterns. The team developed an SVM classifier with an RBF kernel for each resolution work strategy. The MTBO method optimized the hyperparameters C and γ for every SVM classifier by sharing information from all tasks using a combined version of the Gaussian process. MTBO sped up the hyperparameter tuning process for all classifiers more than traditional STBO methods. MTBO reduced classification loss and RMSE scores for all tasks when optimizing various linked models simultaneously. This research uses MTBO technology to boost medical diagnosis work by decreasing processing time and advancing image-scanning accuracy.

Francesco Girlanda, Olga Demler [16] and their team present innovative techniques to better estimate cardiovascular disease using self-supervised learning to merge MRI, ECG and clinical readout information. The framework uses multiple steps in self-supervised learning to develop its performance. The Masked Autoencoder technique lets us train the ECG encoder, which extracts useful patterns from ECG measurements. The system uses an Image Encoder to recognize important information within cardiac MRI inputs. The connection between MRI data and simpler resources such as ECG and clinical records directs the system to recognize patterns better. The network receives special training to determine if a patient has experienced a heart attack. Using multiple types of self-supervision achieved 7.6% better performance than standard training methods in accuracy results. The model recognizes both heart disease cases and non-disease cases more effectively in this situation. The research uses self-supervised learning techniques to analyze multiple data types and produces efficient and reliable CVD warnings despite small labeled dataset availability.

Mohapatra et al. [17] brought forward an automatic system that helps doctors find heart problems faster by looking at medical records on patients' Electronic Health (EHRs). Our system creates automatic heart irregularity and disease detection tools to help medical staff identify heart problems faster and more precisely. The research sets up a two-stage ensemble approach made up of separate ML predictors to analyze various patterns in the dataset. The final prediction system uses base-level outputs to elevate the model's total effectiveness. Our model shows performance numbers of 92% accuracy along with 92.6% precision, sensitivity and specificity of 92.6% and 91%, respectively. The numbers prove the model can reliably detect when patients have or do not have heart problems. The stacking model showed better accuracy than single traditional ML methods when it comes to medical diagnosis. This research shows that stacking ensembles with advanced ML methods helps doctors make better heart disease decisions and handle patient cases faster.

The researchers Qusay Shihab Hamad, Hussein Samma, and Shahrel Azmin Suandi [18] provide a complete review of updated work that uses metaheuristic algorithms to improve CNN hyperparameters for medical imaging tasks. The author wants to simplify CNN hyperparameter tuning for experts and minimize tuning time by looking at metaheuristic optimization methods. This review studies medical image diagnosis research from 2019 to 2022, which optimizes CNN hyperparameters using Genetic Algorithms (GA), Particle Swarm Optimization (PSO), Harris Hawks Optimization (HHO), and Arithmetic Optimization Algorithm (AOA). These optimization algorithms help doctors make better clinical decisions between various medical settings, especially when diagnosing brain growths, coronavirus infections, and breast cancer. The study obtained 98.8% effective results in detecting COVID-19 through optimising CNN networks. Our research documents the most adjusted hyperparameters: learning rate, batch size, number of CNN layers, filters, dropout ratio, and fully connected layer design. Metaheuristic search methods help find appropriate model settings without extensive manual trial and error to boost model performance at a faster development speed. Medical image research becomes more effective with these algorithms that improve tool performance and speed up diagnostic model creation.

Mert Özcan and Serhat Peker [19] studied how to predict heart disease and understand the variable connections using the CART algorithm. The research uses CART as a supervised learning technique for heart disease prediction and determines how input variables connect to this condition. Scientists worked with one dataset from five sources that included 1,190 results with eleven total data points. Our preprocessing methods helped improve how well the model distinguishes accurate results. The CART system trains and forecasts heart disease events. The CART model showed 87% prediction accuracy, which proves it can effectively detect heart diseases. Healthcare providers find it easy to work with the model because its extracted decision rules do not need advanced knowledge inputs to use the tool. Research demonstrates that CART medical models work well because they show important insights and are easy to use. The model's useful decision guidelines help medical staff and patients with time or budget limitations. This investigation boosts understanding of medical machine learning by showing that CART models aid doctors with heart disease evaluation and care setup.

V. K. Sudha and D. Kumar [20] designed an advanced heart disease prediction system using CNN and LSTM networks. Our project creates a deep learning system that enhances heart disease predictions through a combination of CNN and LSTM architecture benefits. The proposed method uses CNN to find medical data patterns while LSTM tracks relevant time-related information. The system-validated results rely on k-fold cross-validation and perform better than SVM, Naïve Bayes, and Decision Trees as traditional algorithms. The amalgamated CNN-LSTM system achieves 89% accuracy through machine learning, outperforming all typical models. This research shows how joining CNN and LSTM networks increases heart disease detection accuracy, which leads to better early diagnosis results.

Ezekiel Adebayo Ogundepo and Waheed Babatunde Yahya [21] study which supervised machine learning classifiers works best for heart disease prediction. Our project examines how ten basic classification algorithms identify heart disease in Cleveland data and show their output on the Statlog dataset. We initially analysed the Cleveland dataset through the Chi-square test to find which bio-clinical factors

were statistically connected to heart disease. We split the Cleveland data into 70% training and 30% testing parts before conducting 200 independent partitions.

The Statlog dataset confirmed our findings from ten separate training and testing experiments on the data. Several specific bio-clinical factors strongly relate to heart disease status according to our analysis (statistical significance < 0.001). The Support Vector Machine (SVM) model delivered outstanding results with 85% accuracy and scores of 82% sensitivity, 88% specificity, 87% precision, an area under the ROC curve of 91%, and a Log Loss of 38%. According to this study, model selection plays a vital role in medical diagnosis, and SVM demonstrates great potential as a heart disease forecasting method.

The research from Manikandan et al. [22] examines how joining the Boruta feature selection model with multiple machine learning tools helps find heart disease better. Our goal is to test if Boruta feature selection leads to better performance when machine learning classifiers predict heart disease. Boruta feature selection helped us test its impact on the Heart Disease Dataset at Cleveland Clinic on LR DT SVM RF and XGBoost.

Boruta found the most useful attributes of the data before training occurred. The utilization of Boruta produced a selection set of six features from thirteen. Boruta feature selection helped Logistic Regression reach 88.52% accuracy, outperforming other tested classifiers. Boruta feature selection demonstrates in research that it boosts machine learning classifiers for heart disease prediction by selecting the best features.

This experiment merges CNN and LSTM deep networks to create a better system for CVD prediction, as explained by Hossain et al. [23]. Their study seeks to make a CVD prediction model by mixing CNN and LSTM neural networks with explainable AI methods to enhance detection outcomes. The system combines CNN layers to read medical information patterns, while LSTM layers detect trends in time-based data. Our team used feature engineering procedures to boost the model's output results.

The model results required SHAP values to show which medical indicators affected CVD risk most. The hybrid CNN-LSTM system reached 74.15% accuracy in not using feature engineering, then 73.52% with feature engineering, which outperformed the existing best results.

Our model gained better interpretability because SHAP values showed us exactly which input factors drive its prediction results. This research proves how linking deep learning systems with AI techniques provides a better way to find CVD early in patients. From the above investigation, it is identified that several important limitations exist that decrease its effectiveness across different models. The models needed to enhance their approach since they did not use advanced methods to select features and extract deep features.

Unbalanced medical data was improperly handled in the developed models, especially when working with rare cardiovascular disease researchers. This research mostly used Cleveland and Statlog datasets that can only represent brief population details from specific areas.

The presented models proved harder to run because their ensemble structure used boosting mechanisms with meta-learning that raised scalability problems on big or complex datasets. Hyperparameter tuning is needed to improve the model's performance. The research evaluated basic machine learning methods yet excluded deep learning options, which provide better results but create high interpretation difficulties.

Only basic model explainability was included, and some dark box systems like SVMs made transparency harder to see in the study, with no evaluation methods used throughout the experiment. There is a need to test many deep learning systems with different datasets to show how well deep learning works in medical diagnosis.

3. Methodology

The methodology develops an ensemble framework based on Boosted Random Subspace [24] Support Vector Machines (SVMs) and stacked generalization implementation for optimum CVD diagnosis results [25]. A normalization phase, along with imputation methods, is used to process the dataset before analysis to overcome missing data problems and unify the format.

Feature selection helps simplify dimensions while maintaining only significant features. After pre-processing the dataset, the Random Subspace method creates diverse feature subsets that multiple SVM classifiers use for training.

A boosting algorithm tackles SVM systems by multiple iterations that specifically address wrongly classified data points for performance optimization.

A stacked generalization layer uses outputs from the boosted SVMs to learn effective prediction combinations. The Bayesian optimization method allows the researchers to modify base and meta-level model hyperparameters to achieve maximum diagnostic accuracy [25].

The Ensemble undergoes quality checks through cross-validation and performance measurements, including accuracy, sensitivity, specificity and AUC-ROC to confirm clinical validity. The sequence of operations is illustrated in Figure 1.

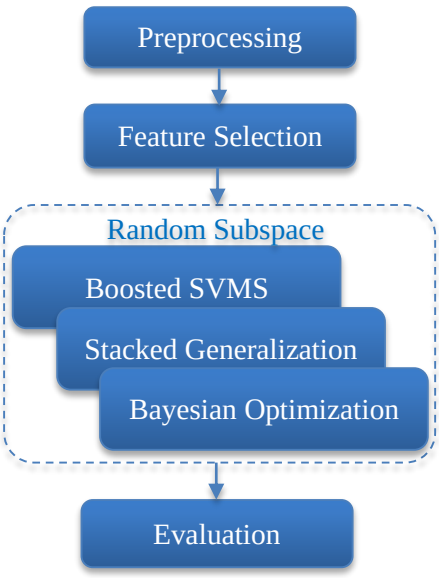

Fig. 1 An optimized framework of Random subspace, boosted SVM and stacked generalization with Bayesian Optimization

3.1. Bayesian Optimization

Bayesian Optimization [26] is a model-based probabilistic technique for identifying optimal function minima or maxima within cost-intensive evaluation non-analytic functions or noisy data scenarios. The technique finds its prime application during hyperparameter model tuning for machine learning systems because every evaluation cycle demands expensive computational resources.

The basic idea of Bayesian Optimization is:

1. The objective function receives a surrogate representation by applying a Gaussian Process function.
2. Choose the next point for evaluation using an acquisition function that estimates uncertainty and improves predicted values.
3. The surrogate model gets updated through previous observations to approach the global optimum quickly.

3.1.1. Bayesian Optimization to Find the Best SVM Parameters

The Bayesian optimization system finds the best Support Vector Machine (SVM) hyperparameters by testing various values of C, kernel selection, and specific gamma in the Radial Basis Function (RBF) kernel [27]. The method differs from grid or random search by using probabilistic learning to pick the best next hyperparameters based on what it has discovered so far.

This repeated method uses exploration and exploitation practices to inspect the function model, which typically measures classification results. Bayesian Optimization decreases testing time by selecting the best hyperparameter sets and helps improve the accuracy of the cardiovascular disease diagnosis model.

Bayesian Optimization to minimize the value of f(x) by adjusting the vector of hyperparameters x:

$$X^* = arg\min_{x \in X} f(X) \tag{1}$$

In the case of SVM, the X could be any of the following.

- C: regularization parameter
- γ: kernel coefficient for 'rbf'
- Type of kernel (ex: -linear','rbf', 'polynomial', etc.)

Bayesian Optimization builds an objective function model using a substitute model typically based on Gaussian Processes (GP):

$$f(X) \sim GP(m(X), k(X, X^{'})) \tag{2}$$

It uses an acquisition function (e.g., Expected Improvement, Upper Confidence Bound) to guide the search:

Adopting an acquisition method (such as Expected Improvement or Upper Confidence Bound) efficiently performs the search space.

$$X_{next} = arg\max_{X} \alpha(X|\hat{f}) \tag{3}$$

Table 1. Advantages of bayesian optimization over other methods

Method	Limitation	Bayesian Optimization Advantage
Grid Search	Exhaustive and inefficient in high dimensions	Requires fewer evaluations by being sample-efficient
Random Search	Ignores past information when selecting next points	Uses prior evaluations to guide search (probabilistic learning)
Gradient-Based	Needs differentiable functions	Works on black-box, non-differentiable, noisy functions
Manual Tuning	Time-consuming and subjective	Automated, principled optimization process

3.1.2. Key Advantages of Bayesian Optimization

Bayesian Optimization is simple and efficient in finding optimal parameters with fewer function evaluations. The algorithm works best with complex models because it requires less training time. Because of its random search technique, the method helps escape local minimum traps. The technique can optimize problems without needing information about gradients or function structure. The algorithm promotes exploration and exploitation automatically by using expected improvement or upper confidence bound methods. Table 1 shows the advantages of Bayesian Optimization over other methods.

3.2. Training a Boosted Ensemble SVM with Random Subspace

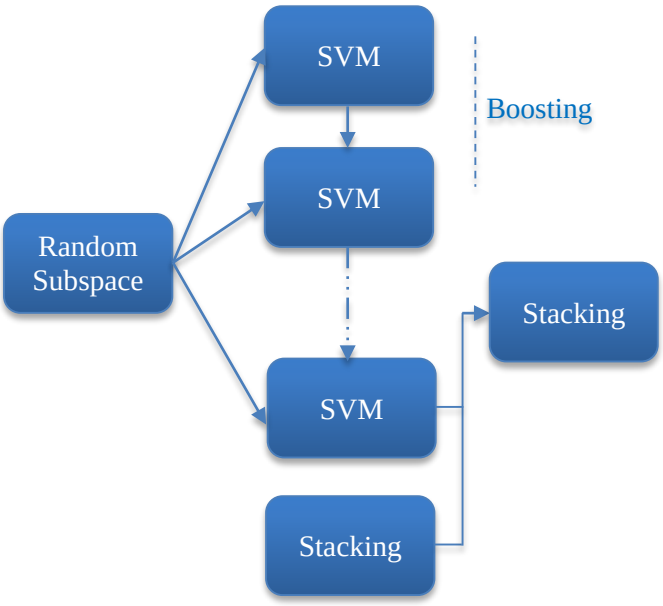

Fig. 2 Pipe-line diagram of Boosted Random Subspace SVM with stacking

To improve diagnostic accuracy and generalization of the model, a boosted ensemble learning strategy is used with the Random Subspace method [28], which is shown in Figure 2. First, it generates multiple training subsets by randomly picking features from the original dataset. A series of base classifiers – SVMs in this case – are trained over these feature subsets, ensuring diversity in the learners.

SVMs are applied with boosting, where the SVM classifiers are trained sequentially and focus more on instances that were misclassified by the previous one so that the bias is reduced, and there is an improvement in the overall performance. To form a strong ensemble model, the weighted scheme is applied to the predictions from all boosted SVMs trained on different subspaces. The basis for this hybrid method was merged from the capabilities of feature diversity (from Random Subspace) and iterative error correction (from boosting), which results in a robust and accurate framework for making cardiovascular disease diagnosis.

A Mathematical formulation for training a Boosted Ensemble using Random Subspace, which includes both techniques into a formal process:

Let the training dataset be:

$$D=\{(x_i,y_i)\}_{i=1}^{N},\ x_i \in R^d, y_i \in \{+1,-1\} \quad (4)$$

Step 1: Random Subspace Sampling.

For each base learner $t = 1,2,\dots,T$:

- Randomly select a subset of features $F_t \subset \{1,2,\dots,d\}$
- Construct a training set using only the features in F_t:

$$D_t = \{(x_i^{(F_t)}, y_i)\}_{i=1}^{N}$$

Where $x_i^{(F_t)}$ is the projection x_i onto the feature subspace F_t.

Step 2: Boosting (AdaBoost-style)

Initialize weights for each sample:

$$w_i^{(1)} = \frac{1}{N}, \text{i=1, …,N}$$

For each iteration $t = 1,2,\dots,T$:

1. Train a base classifier h_t (e.g., SVM) on D_t wit sample weights $w_i^{(t)}$
2. Compute weighted classification error:

$$\varepsilon_t = \sum_{i=1}^{N} w_i^{(t)} \cdot \prod(h_t(x_i^{(F_t)}) \neq y_i)$$

3. Compute classifier weight: $\alpha_t = \frac{1}{2} ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$
4. Update sample weights:

$$w_i^{(t+1)} = w_i^{(t)} \cdot exp\left(-\alpha_t y_i h_t(x_i^{(F_t)})\right)$$

Normalize $w_i^{(t+1)}$ so that $\sum_i w_i^{(t+1)} = 1$

Final Prediction:

The ensemble classifier $H(x)$ aggregates the predictions of base learners:

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t \cdot h_t(x^{(F_t)})\right) \quad (5)$$

This formulation helps understand how Random Subspace makes features diverse while Boosting decreases the error and increases the robustness.

3.3. Stacked Generalization with Logistic Regression

Several base classifiers work together using Stacked Generalization to produce better results than the separate models would achieve alone. Various boosted SVMs trained from random subspaces are taught individual approaches before using them in Stacked Generalization [29]. After producing outcome predictions on a validation set, the models feed their results into a meta-classifier for training that learns which combination of outputs creates the best predictions.

Generally, Logistic Regression is used as a meta-classifier because it offers easy interpretation and successful results. This method uses weighted addition to produce the final prediction by analyzing how base learners generate results and the resulting class membership likelihood. Multiple individual models improve accuracy when they work together in stacking and can handle diagnostic tasks better than any one model by itself.

3.4. Mathematical Formulation to implement Stacked Generalization with Logistic Regression as the Meta-Classifier

Let:

- $h_1(x), h_2(x), \dots, h_T(x)$: outputs of T base classifiers (e.g., boosted SVMs)
- $z_i = [h_1(x_i), h_2(x_i), \dots, h_T(x_i)]$: the feature vector for the sample $x_i$used by the meta-classifier

Meta-level logistic regression model:
Train logistic regression $z_i$to predict y_i:

$$P(y = 1|z) = \frac{1}{1+exp(-(w^T z+b))} \quad (6)$$

Here, w denotes the weight vector for the base classifier outputs, and b denotes the bias term.

The parameters w b are learned by minimizing the log-loss function over the validation data:

$$L = -\sum_{i=1}^{N}[y_i\, log(\,P(y_i)) + (1 - y_i)\, log(\,1 - P(y_i))] \quad (7)$$

In short, this process effectively learns how to weight the base classifiers' predictions using logistic regression in a way that the result of the final prediction is more accurate and robust.

3.5. Cross-Validation of the Final Ensemble Model

Cross-validation [30] enables strict testing to show whether the combined model of boosted SVM voting classifiers and stacked generalization delivers good predictions. Here, the dataset is split up into several folds (k-fold cross-validation is often used), and the model is trained on k−1 folds and tested on the remaining one. This is repeated k times, where each of the k folds will be taken as a test set once. Within each fold, base learners are trained within the ensemble learning framework and boosting on their respective random subspace features and the meta classifier (e.g., logistic regression) is trained on out-of-fold predictions. In order to guarantee that results are not biased and are representative of the model's actual predictive strength, the final performance metrics, i.e., accuracy, sensitivity, specificity, and AUC, are averaged across all folds. This approach to validation enables the detection of overfitting while ensuring the robustness of the Ensemble used for diagnosing cardiovascular disease.

4. Results and Analysis

4.1. Dataset Description

This study uses the Heart Disease Dataset (Comprehensive) that Manu Siddhartha created at Liverpool John Moores University and made available on IEEE DataPort [31]. The dataset includes five prominent heart disease records from Cleveland, Hungarian, Switzerland, Long Beach, VA, and Statlog (Heart), which were combined to create a single set containing 1,190 examples with 11 core variables. It has become the largest free database to research coronary artery disease.

This database exists to build heart disease detection systems by helping machine learning methods predict diseases at their early stages. The data contains important clinical measurements: age, gender, chest pain type, blood pressure, cholesterol levels, blood sugar readings, initial ECG results, maximum heart rate, exercise-related chest pain, ST segment depression during exercise compared to rest, and ST segment response. The dataset, however, is a rich resource for CAD-related machine learning and data mining algorithms though, but it is worth noting limitations in the dataset. In particular, there are 272 duplicates, and some imputed missing values with zeros, which can influence the performance and accuracy of predictive models. This indicates that the data analysis researchers should apply suitable preprocessing on the data before proceeding.

4.2. Experimental Setup

In Windows 10 Professional (64-bit) operating system, the proposed cardiovascular disease (CVD) detection model was performed based on MATLAB 2021b software. It was implemented on a laptop brand, Lenovo, at the ThinkPad T480 with an Intel Core i5 8th Generation processor, having 4 cores and 8 threads with a base clock speed of 1.6 GHz and a maximum turbo clock speed of 3.4 GHz. Such system architecture offered enough computing power in data preprocessing, training, and evaluation. MATLAB environment installed some of the most important toolboxes, like the Statistics and Machine Learning Toolbox, with the aim of implementing the classifiers like the logistic regression and support vector machines (SVM), the Image Processing Toolbox that will process the clinical data (in case it was required), and the Deep Learning Toolbox introducing components on the basis of neural networks or the Optimization Toolbox that will tune hyperparameters and improve performance. All of these toolboxes helped to work out the framework of hybrid classification, including sigmoid-hyperbolic activation function and logistic regression-based analysis, refining them. All in all, the chosen system and the software environment were effective and viable to implement the suggested methodology.

4.3. Performance Metrics

To measure how well the stacked ensemble model accurately diagnoses, the accuracy, precision, sensitivity, F1-score and AUC scores are computed [32, 33]. These measurement tools help us completely test the classifier's results, especially in healthcare, since wrong diagnoses can lead to serious effects.

1. Accuracy

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (8)$$

2. Precision

$$\text{Precision}=\frac{\text{TP}}{\text{TP+FP}} \quad (9)$$

3. Recall

$$\text{Recall}=\frac{\text{TP}}{\text{TP+FN}} \quad (10)$$

4. F1 Score

$$\text{F1Score}=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (11)$$

5. Area Under the Curve (AUC): It evaluates the model's capability to differentiate between classes for various threshold settings. The greater AUC suggests better separability.

The Table 2 Showing the evaluation of performance metrics between the different models. Figure 3 shows the graphical analysis between the proposed method and other models. Accuracy measures the overall correctness of the model. The proposed method achieves the highest accuracy at 96.39%, significantly outperforming all others.

SVM comes second at 88.24%, indicating solid performance but still about 8.15% lower than the proposed method. The proposed method likely benefits from more sophisticated learning or ensemble mechanisms, resulting in fewer classification errors.

Precision shows how well a model identifies actual positive results. The proposed method demonstrates 96.51% precision with only a small number of predictions that actually proved wrong, while Decision Tree claimed 88.11% and Naïve Bayes achieved 85.18%. The Proposed Method provides dependable results because it generates few incorrect positive predictions, especially when these mistakes lead to substantial costs.

The recall shows how often the model finds real positive results. The SVM model reaches 97.14% positive case detection, which is just 0.5% better than what the Proposed Method delivers. The K-NN algorithm achieves 87.12% accuracy, while Random Forest follows with 86.65%. Despite finding more positive cases than other models, the Support Vector Machine shows a decrease in accuracy. The proposed method reaches close to the best results in both accuracy and performance.

F1-Score represents the average performance between precision and recall since it uses their harmonic mean. The proposed method leads the pack at 96.58% because it exhibits reliable performance throughout all trials. SVM provides results that are 7% lower than the proposed method. The proposed method shows excellent results both in identifying actual positives (recall) and reducing false detections (precision) based on its F1-Score measurement.

Table 2. Performance metrics analysis between proposed model vs State of the art models

Model	Accuracy	Precision	Recall	F1-Score	AUC
SVM	0.8824	0.8336	0.9714	0.8972	0.9483
Logistic Regression	0.8269	0.8331	0.8410	0.8370	0.9007
Decision Tree	0.8647	0.8811	0.8601	0.8705	0.8975
K-NN	0.8563	0.8589	0.8712	0.8650	0.9129
Naïve Bayes	0.8345	0.8518	0.8315	0.8415	0.9021
Random Forest	0.8445	0.8437	0.8665	0.8549	0.9195
Proposed Method	0.9639	0.9651	0.9666	0.9658	0.9637

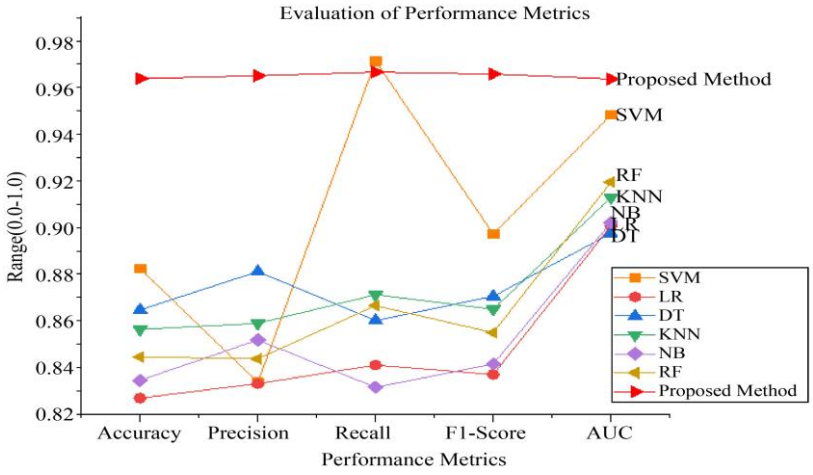

Fig. 3 Evaluation of performance metrics between different models

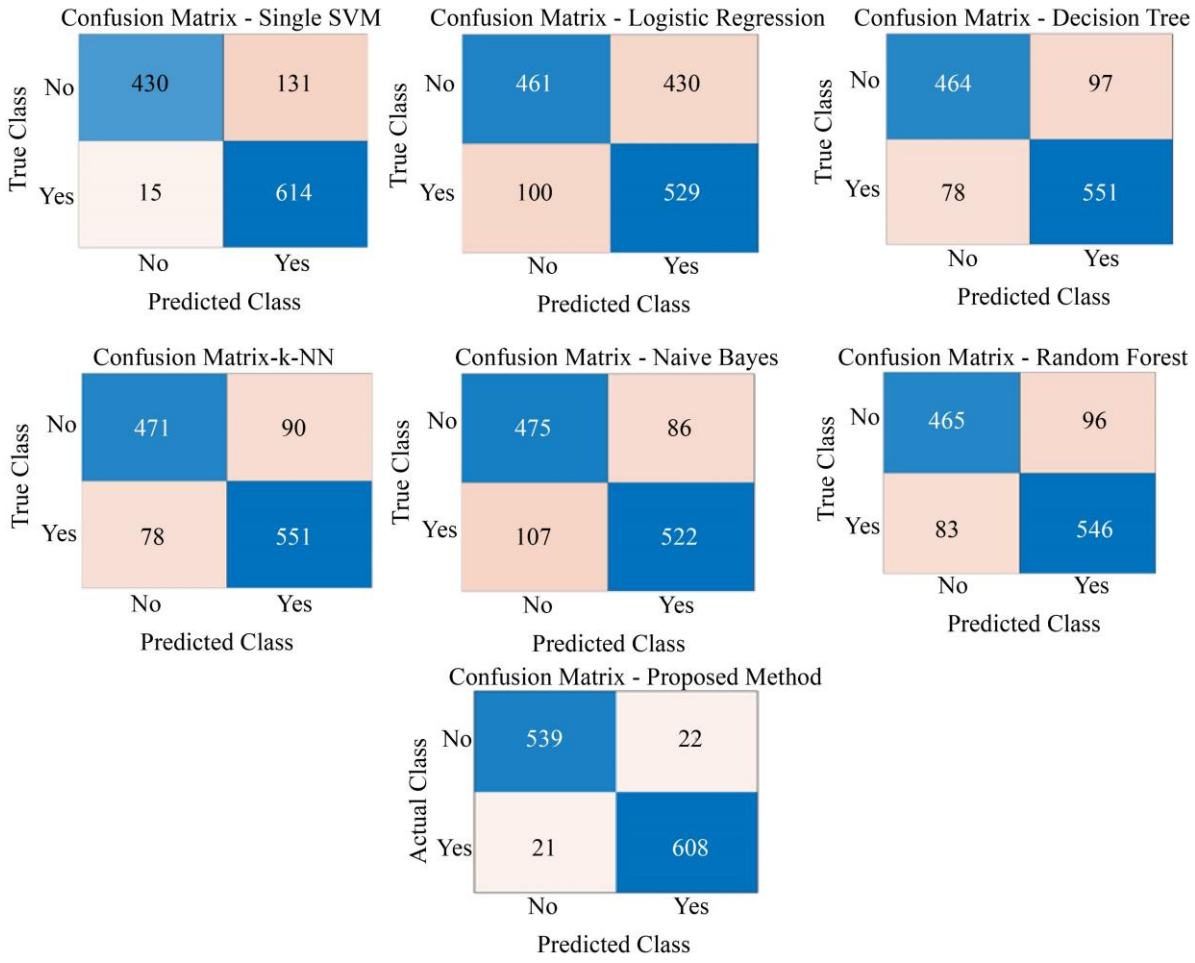

Fig. 4 Confusion matrices

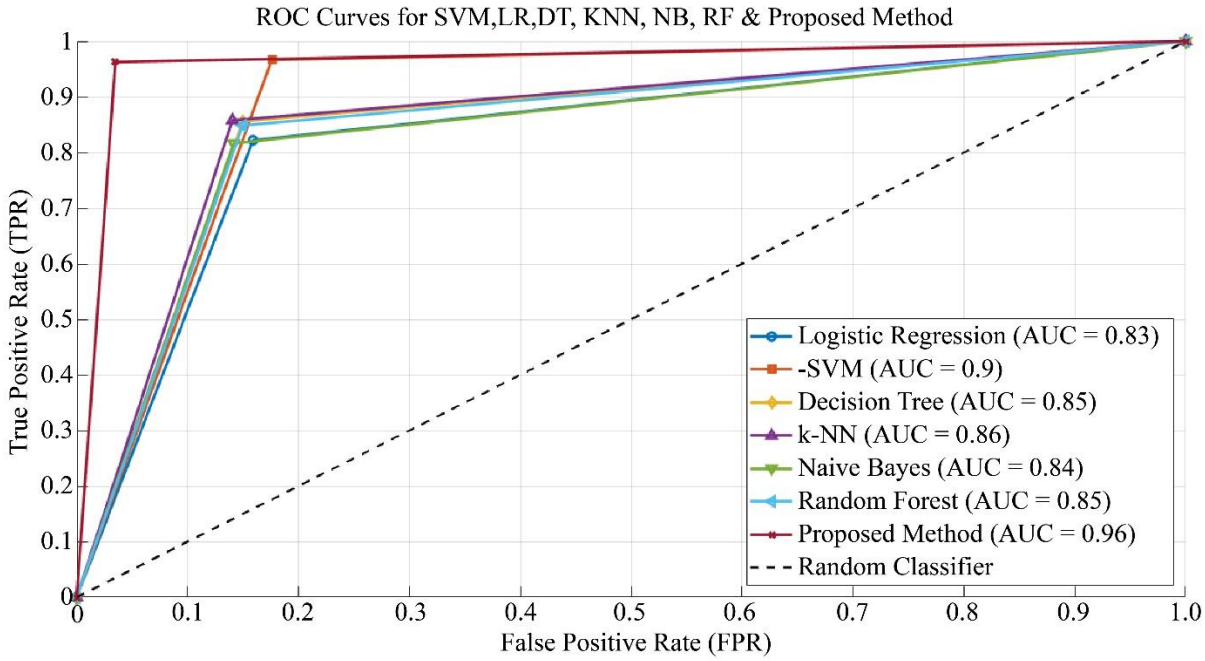

Fig. 5 Analysis of ROC curve and AUC scores

Table 3. Summary of comparative performance

Reference No.	Year	Methodology	Best Accuracy
[19]	2022	Basic CART model	87.00%
[35]	2021	Individual models: RF, LR, SVM with 10-fold CV	92.00%
[36]	2022	Ensemble with AdaBoost combining SVM, LR, RF, ANN, MLPNN	93.39%
[37]	2022	Stacked Ensemble using LR, RF, SGD, GDC, ADA	91.84%
[38]	2023	Soft Voting Ensemble of RF, LR, SVE, KNN, NB, with GB and AB	95.00%
[39]	2024	Two-stage stacking using RF, DT, XGB	96.00%
Proposed	**—**	**Optimized Ensemble of Boosted Random Subspace with stacked SVMs (Bayesian-tuned)**	**96.39%**

AUC shows how well a model recognizes differences between different types of items. The proposed method showed excellent classification skills, achieving a 96.37% result. Random Forest achieved a strong performance of 91.95%, while SVM surpassed it at 94.83%. According to AUC results, the proposed method demonstrates its ability to place positive instances at the top of its rankings accurately based on imbalanced data.

Based on the results from this analysis, our method shows clear improvement in every measurement. SVM maintains its position as a reliable standard model that performs well in recall and AUC. The proposed method performs better across multiple metrics as other baseline models demonstrate lower-than-average results. Figure 3. Showing the graphical analysis of performance metrics between proposed and other models.

4.4. Confusion Matrix Analysis

A confusion matrix is a table that the analyst uses to determine how well the classification algorithm is performing [34]. This gives the overall bar chart, which shows how many predictions were correct and incorrect, and splits it up for each class. Visualizing a model's accuracy using the matrix, most common in multiclass or imbalanced datasets, is helpful.

The true positive rate (correct identification of patients with heart disease) and true negative rate (correct identification of healthy individuals) were high, and a diagonal largely dominated the confusion matrix. This implies that the Ensemble was able to learn discriminative patterns well from the normalized feature space. Figure 4 shows the confusion matrices generated from the proposed method and other models.

4.5. Comparison of ROC Curves and AUC Scores

A Receiver Operating Characteristic diagram shows how well a model detects problems during the evaluation process [33]. It shows how the True Positive Rate relates to the False Positive Rate when adjusting detection thresholds. AUC analyzes model performance, while class separation improves when AUC values rise. The analysis of AUC scores and ROC curves between proposed and state-of-the-art models is shown in Figure 5.

$$AUC(approximate) = \frac{1}{2} \times (1 + TPR - FPR) \quad (12)$$

- The Proposed Method delivers superior performance by achieving an AUC of 0.9641, touching the best result of all classification models thus demonstrating outstanding classification capabilities.
- SVM provide effective results through an AUC of 0.90, although they display high sensitivity and show some tendency toward false positives.
- Random Forest and k-NN achieve comparable results to each other as they produce evaluation scores of 0.85 to 0.86 based on AUC.
- The AUC measurements demonstrate that Logistic Regression and Naïve Bayes provide fewer accurate predictions in this application.
- The Proposed Method stands as the leading model in both sensitivity and specificity tests because its AUC exceeds 0.5 in all predictions.

4.6. Baseline Model Comparison

Many researchers use machine learning approaches to predict heart disease on the IEEE Dataport dataset through their research. Research into heart disease prediction uses different approaches that include model-building methods and optimization procedures. Table 3 shows how this research developed and performed in relation to others.

The proposed method presents a performance that outperforms the state-of-the-art, and this is evidence of stacked ensemble Learning, boosting methods, and automated hyperparameter optimization's efficacy. Thus, it is considered the state of art approach for heart disease prediction on this dataset.

4.7. Practical Implications

The suggested ensemble learning system brings real benefits to CVD diagnosis in medical practice. The model helps medical teams perform better decisions at the right time with higher test precision because it gives more precise results for patient care. Its ability to work with medical structures makes it ready for use in Clinical Decision Support Systems (CDSS). This model system lets machine learning perform automatic tuning work that hospitals and diagnostic centres can easily use. The framework needs more testing and system integration into electronic record systems to help healthcare organizations identify heart disease risks sooner and create individual care plans.

5. Conclusion

An optimized framework of ensemble learning was proposed in this study to improve the diagnosis of cardiovascular disease based on boosted random subspace SVMs and stacked generalization for the promotion of diagnostic performance. Random subspace selection was integrated to increase the model diversity, and boosting and stacking were used to capture some of the complex patterns in the data, improving classification accuracy and robustness. Finally, Bayesian Optimization was used to fine-tune hyperparameters efficiently and improve model performance. The research evaluates the proposed method through standard test collections and shows it delivers superior results than existing approaches in disease identification tasks. The proposed ensemble model proves useful in heart disease detection because it shows 96.39% accuracy, surpassing other single classifiers and basic ensemble methods in clinical heart disease testing. Moreover, the framework is designed for the necessary structural flexibility and predictive strength to be a promising tool for computer-aided diagnosis. However, the study is limited in that model interpretability, computational cost, and clinical validation are critically lacking. However, the proposed method clearly represents a huge progress in exploiting advanced ensemble learning in cardiovascular risk estimation. Moreover, future extensions will aim at explaining more, simplifying models and validating the approach in clinical real-world settings to support reliable and transparent decision making.

The next studies will test this model on big medical datasets from various practices to make sure the results apply in different settings. To enable interpretability to clinical users, we will prioritize the incorporation of explainable AI techniques. The complexity and resource requirements will be decreased by exploring efficient feature selection methods and lightweight model variants. Class imbalance and overfitting will be considered, and strategies to handle them, including synthetic sampling and advanced regularization. Lastly, introduction into EHR systems and prospective clinical trials will be pursued to evaluate effectiveness in the real world.

References

[1] Cardiovascular Diseases, Key Facts, World Health Statistics, 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] Robin P Choudhury, and Naveed Akbar, "Beyond Diabetes: A Relationship between Cardiovascular Outcomes and Glycaemic Index," *Cardiovascular Research*, vol. 117, no. 8, pp. E97-E98, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] World Heart Report 2023 Confronting the World's Number One Killer, World Heart Federation, 2023. [Online]. Available: https://world-heart-federation.org/resource/world-heart-report-2023/

[4] Zakaria K.D. Alkayyali, Syahril Anuar Bin Idris, and Samy S. Abu-Naser, "A Systematic Literature Review of Deep and Machine Learning Algorithms in Cardiovascular Diseases Diagnosis," *Journal of Theoretical and Applied Information Technology*, vol. 101, no. 4, pp. 1353-1365, 2023. [Google Scholar] [Publisher Link]

[5] Rahma Atallah, and Amjed Al-Mousa, "Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method," *2019 2nd International Conference on new Trends in Computing Sciences*, Amman, Jordan, pp. 1-6, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[6] Vivek Pandey, Umesh Kumar Lilhore, and Ranjan Walia, "A Systematic Review on Cardiovascular Disease Detection and Classification," *Biomedical Signal Processing and Control*, vol. 102, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[7] Xuewei Du et al., "Progress, Opportunities, and Challenges of Troponin Analysis in the Early Diagnosis of Cardiovascular Diseases," *Analytical Chemistry*, vol. 94, no. 1, pp. 442-463, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Yongqiang Weng et al., "Application of Support Vector Machines in Medical Data," *2016 4th International Conference on Cloud Computing and Intelligence Systems*, Beijing, China, pp. 200-204, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[9] Rosita Guido et al., "An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review," *Information*, vol. 15, no. 4, pp. 1-36, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Dakhaz Mustafa Abdullah et al., "Machine Learning Applications Based on SVM Classification a Review," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 81-90, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Muralidharan Jayaraman, and Shanmugavadivu Pichai, "Improving the Prediction Accuracy of Onset of Cardiovascular Diseases, using Ensemble Learning," *Proceedings of the 1st International Conference on Artificial Intelligence, Communication, IoT, Data Engineering and Security*, Lavasa, Pune, India, pp. 1-6, 2023. [CrossRef] [Publisher Link]

[12] Shahid Mohammad Ganie et al., "An Improved Ensemble Learning Approach for Heart Disease Prediction Using Boosting Algorithms," *Computer Systems Science and Engineering*, vol. 46, no. 3, pp. 3993-4006, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Achyut Tiwari, Aryan Chugh, and Aman Sharma, "Ensemble Framework for Cardiovascular Disease Prediction," *Computers in Biology and Medicine*, vol. 146, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] Narendra Kumar Sharma et al., "Enhancing Heart Disease Diagnosis: Leveraging Classification and Ensemble Machine Learning Techniques in Healthcare Decision-Making," *Journal of Integrated Science and Technology*, vol. 13, no. 1, pp. 1-8, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[15] Wenhao Chi et al., "Alleviating Hyperparameter-Tuning Burden in SVM Classifiers for Pulmonary Nodules Diagnosis with Multi-Task Bayesian Optimization," *Arxiv*, pp. 1-12, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Francesco Girlanda et al., "Enhancing Cardiovascular Disease Prediction through Multi-Modal Self-Supervised Learning," *Arxiv*, pp. 1-18, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Subasish Mohapatra et al., "A Stacking Classifiers Model for Detecting Heart Irregularities and Predicting Cardiovascular Disease," *Healthcare Analytics*, vol. 3, pp. 1-10, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Qusay Shihab Hamad, Hussein Samma, and Shahrel Azmin Suandi, "Optimization of Convolutional Neural Network Hyperparameter for Medical Image Diagnosis Using Metaheuristic Algorithms: A Short Recent Review (2019-2022)," *Arxiv*, pp. 1-8, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Mert Ozcan, and Serhat Peker, "A Classification and Regression Tree Algorithm for Heart Disease Modeling and Prediction," *Healthcare Analytics*, vol. 3, pp. 1-9, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] V.K. Sudha, and D. Kumar, "Hybrid CNN and LSTM Network for Heart Disease Prediction," *SN Computer Science*, vol. 4, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Ezekiel Adebayo Ogundepo, and Waheed Babatunde Yahya, "Performance Analysis of Supervised Classification Models on Heart Disease Prediction," *Innovative Systems and Software Engineering*, vol. 19, pp. 129-144, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[22] G. Manikandan et al., "Classification Models Combined with Boruta Feature Selection for Heart Disease Prediction," *Informatics in Medicine Unlocked*, vol. 44, pp. 1-12, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[23] Md Maruf Hossain et al., "Cardiovascular Disease Identification Using a Hybrid CNN-LSTM Model with Explainable AI," *Informatics in Medicine Unlocked*, vol. 42, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[24] Ye Tian, and Yang Feng, "RaSE: Random Subspace Ensemble Classification," *Journal of Machine Learning Research*, vol. 22, no. 45, pp. 1-93, 2021. [Google Scholar] [Publisher Link]

[25] Peiwen Tan, "Ensemble-Based Hybrid Optimization of Bayesian Neural Networks and Traditional Machine Learning Algorithms," *Arxiv*, pp. 1-23, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[26] Ahmed M. Elshewey et al., "Bayesian Optimization with Support Vector Machine Model for Parkinson Disease Classification," *Sensors*, vol. 23, no. 4, pp. 1-21, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[27] Alaa M. Elsayad, Ahmed M. Nassef, and Mujahed Al-Dhaifallah, "Bayesian Optimization of Multiclass SVM for Efficient Diagnosis of Erythemato-Squamous Diseases," *Biomedical Signal Processing and Control*, vol. 71, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[28] Alexandros Vamvakas et al., "Breast Cancer Classification on Multiparametric MRI – Increased Performance of Boosting Ensemble Methods," *Technology in Cancer Research & Treatment*, vol. 21, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[29] Sruthi Nair et al., "Combining Varied Learners for Binary Classification Using Stacked Generalization," *Arxiv*, pp. 1-9, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[30] T.R. Mahesh et al., "An Efficient Ensemble Method Using K-Fold Cross Validation for the Early Detection of Benign and Malignant Breast Cancer," *International Journal of Integrated Engineering*, vol. 14, no. 7, pp. 204-216, 2022. [Google Scholar] [Publisher Link]

[31] Heart Disease, IEEE Dataport. [Online] Available: https://ieee-dataport.org/keywords/heart-disease

[32] Gireen Naidu, Tranos Zuva, and Elias Mmbongeni Sibanda, "A Review of Evaluation Metrics in Machine Learning Algorithms," *Artificial Intelligence Application in Networks and Systems*, pp. 15-25, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[33] Oona Rainio, Jarmo Teuho, and Riku Klén, "Evaluation Metrics and Statistical Tests for Machine Learning," *Scientific Reports*, vol. 14, no. 1, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[34] S. Sathyanarayanan, and B. Roopashri Tantri, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, vol. 27, no. 4S, pp. 4023-4031, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[35] Rüstem Yılmaz, and Fatma Hilal Yağın, "Early Detection of Coronary Heart Disease Based on Machine Learning Methods," *Medical Records*, vol. 4, no. 1, pp. 1-6, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[36] Bhanu Prakash Doppala et al., "A Reliable Machine Intelligence Model for Accurate Identification of Cardiovascular Diseases Using Ensemble Techniques," *Journal of Healthcare Engineering*, vol. 2022, no. 1, pp. 1-13, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[37] Najmu Nissa, Sanjay Jamwal, and Mehdi Neshat, "A Technical Comparative Heart Disease Prediction Framework Using Boosting Ensemble Techniques," *Computation*, vol. 12, no. 1, pp. 1-22, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[38] Nadikatla Chandrasekhar, and Samineni Peddakrishna, "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization," *Processes*, vol. 11, no. 4, pp. 1-31, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[39] Subhash Mondal et al., "An Efficient Computational Risk Prediction Model of Heart Diseases Based on Dual-Stage Stacked Machine Learning Approaches," *IEEE Access*, vol. 12, pp. 7255-7270, 2024. [CrossRef] [Google Scholar] [Publisher Link]