Online-Journals.org

iJOE International Journal of
Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 9 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i09.55425

PAPER

Improving the Efficiency of Predicting the Heart Diseases Using Optimized Feature Selection and Ensemble Machine Learning Techniques

Archana P.(✉),
Shashikala S. V.

Department of Computer
Science and Engineering,
BGS Institute of Technology,
Adichunchanagiri University,
BG Nagara, Mandya,
Karnataka, India

archanahavishsurya@gmail.com

ABSTRACT

Millions of people worldwide suffer from heart failure, a chronic illness that makes an effective machine learning (ML)-based approach for early detection and treatment necessary. Although medication is still the mainstay of care, exercise is becoming recognized as a useful adjunctive therapy for the management of heart failure. In this work, we used patient health parameter data to design a ML-based method to enhance heart failure detection. Improving the early detection of heart failure is our goal in an effort to save lives. To find the most important features for enhancing performance, we conducted a comparative analysis of ten distinct ML algorithms and applied feature engineering methodologies. By developing a novel new feature set, we improved our strategy and obtained the best accuracy ratings. The proposed system works on the statistical dataset and CT scan images. Numerous experiments were carried out to assess the efficacy of different algorithms, and our suggested approach outperformed other cutting-edge models, attaining impressive accuracy. Cross-validation approaches were employed to validate all applied procedures. On the CT scan dataset, AdaBoost (AB) achieved 100% accuracy, while gradient boosting (GB) led with 96% on the statistical dataset. Accuracy improved with random or synthetic data. Notably, applying a soft voting ensemble of all models further boosted accuracies to 98% and 95% on the respective datasets. Our study advances heart failure early detection techniques, which make important scientific contributions to the medical world.

KEYWORDS

heart disease, machine learning (ML), heart failure rate, classification, cross validation

1 INTRODUCTION

Since cardiovascular diseases (CVDs) are the primary cause of death worldwide, efficient management and preventative techniques are required. Different strategies

Archana, P., Shashikala, S.V. (2025). Improving the Efficiency of Predicting the Heart Diseases Using Optimized Feature Selection and Ensemble Machine Learning Techniques. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(9), pp. 27–42. https://doi.org/10.3991/ijoe.v21i09.55425

Article submitted 2025-03-09. Revision uploaded 2025-05-23. Final acceptance 2025-05-23.

for comprehending and forecasting cardiovascular outcomes are highlighted by recent research [1]. In their discussion of the connection between glycaemic index and cardiovascular outcomes, [2] stresses the significance of glycaemic control in lowering cardiovascular risks. Using association rule discovery, [3] presented an early technique for heart disease prediction that aided in the detection of important patterns in medical data. A cluster-based decision tree learning technique was presented in [4] to pick features optimally and increase the precision of heart disease forecasts. The usefulness of ensemble learning approaches in the early diagnosis of cardiac disease was illustrated in [5], highlighting the advantages of merging several machine learning (ML) models for improved predictive performance. A stacking model fusion-based predictive classifier for cardiovascular disease was developed in [6] and concluded by emphasizing the role of various ML algorithms in the prediction of cardiovascular disease, underscoring advancements in predictive healthcare [7]. This classifier combines multiple models for improved accuracy.

Different ML classification algorithms have been used in various applications including healthcare, trust prediction, vehicular networks, and many more [8]–[10]. Utilizing ML and deep learning approaches to improve diagnostic accuracy and early detection has been the main focus of recent research on heart disease prediction. The potential of integrated models in physiological data analysis was demonstrated in [11], who devised a method to diagnose chronic heart failure from heart sounds using a stack of ML classifiers. The authors in [12] emphasized the major healthcare difficulties associated with heart failure and its impact on worldwide public health.

Comprehensive statistics on heart disease and stroke were presented in [13], emphasizing risk factors, mortality rates, and prevalence. In their assessment of safe and reliable ML methods for the healthcare industry, [14] addressed important concerns such model dependability and data privacy. Predictive analytics has advanced, as demonstrated in [15], which demonstrated how ML classifiers could accurately predict coronary heart disease.

In order to forecast cardiac illness, [16] analysed several ML algorithms and provided insights into the most efficient approaches. In order to enhance clinical decision-making, [17] created a decision support system for cardiac disease prediction utilizing various ML models. In their evaluation of early heart failure prediction methods, [18] emphasized the significance of prompt intervention.

Existing techniques for predicting cardiac disease have a number of severe drawbacks, despite tremendous progress. There are few attempts to combine statistical data and CT scan pictures for thorough analysis, and many methods are not generalizable across various data sources. Furthermore, approaches that use principal component analysis (PCA) or other black-box models frequently have poor interpretability, which makes clinical validation and acceptance difficult. The limited use of robust assessment methods, which are essential for attaining dependable performance and include ensemble optimization and systematic hyper parameter tuning, is another drawback. Furthermore, overfitting usually results from relying on short or unbalanced datasets, which lowers the models' efficacy in actual clinical settings.

1.1 Motivation and problem statement

Heart disease is one of the major causes of death and serious illness worldwide. Because of the disease's complexity and range of symptoms, early discovery is crucial for improving outcomes and lessening the burden on the healthcare system. However, correct diagnosis is challenging to achieve. ML, with its proven success in various disciplines, has considerable promise for healthcare. These algorithms improve the identification of cardiac illness by analysing vast datasets to find complex patterns and generate precise predictions. By applying these strategies, automated decision-support systems may efficiently handle a wide range of patient data, giving physician's insightful information for efficient diagnosis and treatment planning.

1.2 Research contributions

The key contributions of this paper are as follows:

- Using performance classification criteria as its primary focus, this study assesses and applies ten well-known ML algorithms to the statistical and CT scan heart datasets.
- To attain the best accuracy, the grid search CV hyper parameter tuning method with ten-fold cross-validation was applied. Performance was assessed by accuracy and negative log loss.
- On randomly created data, every ML model was put to the test and achieved better accuracy.
- In the end, a soft voting ensemble approach was used to merge all classifiers to improve the accuracy of the model. The effectiveness of the results was demonstrated by comparing them with those of prior relevant works.

1.3 Paper organization

The paper is organized to methodically investigate the use of ML for heart disease identification. The literature on heart disease classification and prediction models is reviewed in Section 2. The suggested methodology is explained in Section 3, along with specifics about the dataset and ML model that were employed. The implementation is the main topic of Section 4, which also presents the outcomes and offers a discussion of the conclusions. The paper is finally concluded in Section 5, which summarizes the main findings and contributions of the study.

2 RELATED WORKS

The application of various ML and data mining approaches has led to substantial breakthroughs in the field of disease prediction, especially for heart disease. In order to assess various supervised ML algorithms' efficacy in illness prediction, [19] compared them, highlighting the crucial significance of algorithm selection. The potential of cutting-edge optimization approaches was demonstrated in [20] when they presented a novel optimization algorithm to improve heart disease prediction accuracy. Using several ML techniques, [21] investigated the impact of food contamination on gastrointestinal morbidity. Their findings can be applied to more general disease prediction scenarios, such as cardiovascular health. By combining ML and data mining algorithms, [22] created a novel feature reduction model that increased prediction efficiency by decreasing data dimensionality. The MIFH framework was proposed in [23], which combines several ML approaches to improve the diagnosis of heart disease.

In order to anticipate heart illness, [28] developed a decision support system that makes use of several ML models. This system has potential uses in clinical settings. [24] underlined the significance of data completeness in accurate illness prediction by concentrating on enhancing classifier accuracy by resolving missing data through imputation techniques. By identifying important risk factors, [25] showed how well logistic regression (LR) techniques work in predicting cardiovascular disease. In order to forecast cardiac illness, [26] analysed several data mining and ML algorithms, offering insights into the relative effectiveness of various techniques. Last but not least, [27] investigated hybrid ML approaches, combining many algorithms to improve prediction performance and demonstrating the advantages of combining various ML techniques for increased accuracy.

Although a lot has been learned about applying ML and data mining approaches to the prediction of cardiac disease, there are still a few unanswered questions that need to be investigated further:

- Algorithm selection and optimization: Research has shown the value of choosing and refining the appropriate algorithms, as evidenced by the works of [22]. Nevertheless, a thorough framework that methodically assesses and chooses the best algorithms and optimization strategies for particular datasets and prediction tasks is still required.
- Feature reduction and selection: By decreasing data dimensionality, feature reduction models—such as the one by [26]—and sophisticated feature selection methods—discussed by [25]—have the potential to improve prediction efficiency. To preserve prediction accuracy without sacrificing important information, further study is required to create more complex, adaptive feature reduction and selection techniques that can dynamically adapt to a variety of changing datasets.
- Handling data imbalances and missing data: The work of [28] on using imputation approaches to resolve missing data emphasizes how crucial complete data is to precise prediction. In order to guarantee that models continue to function well even when working with faulty datasets, future research should concentrate on creating reliable techniques to handle data imbalances and missing values.
- Integration of diverse ML techniques: Multiple ML algorithms combined into hybrid and ensemble techniques have been demonstrated to increase prediction performance in a number of studies, including [27] and [28]. Nonetheless, further research is required to determine the best arrangements and blends of these methods. Future research should examine how various ML models perform in concert with one another and determine the most effective ways to integrate them in order to optimize heart disease prediction accuracy and robustness.
- Real-time and long-term monitoring systems: A large number of the models in use today, including the ones put forth by [25] and [26], concentrate on forecast accuracy utilizing past data. Systems that can detect cardiac illness in real-time and anticipate its long-term danger are becoming more and more necessary.

By considering all these gaps, it may be possible to create heart disease prediction systems that provide more accurate, dependable, and efficient outcomes and progress the field of medical diagnostics.

3 PROPOSED WORK

The proposed approach is shown in Figure 1 which describes a system that uses the best feature selection to assist people in determining their risk of heart disease. The following are the six main steps: (i) Data collection with varying parameters from several sources; (ii) Pre-processing the data; (iii) Splitting data; (iv) Classification using K-nearest neighbour (KNN), support vector classification (SVC), decision trees

(DT), and other ML classifiers; (v) Performance metrics; and (vi) K-fold technique for model validation.

3.1 Dataset description

We have considered two different datasets for study and experimentation. The statistical dataset [29] has 12 features spread across 1900 instances and CT scan heart images [30]. The statistical dataset contains both normal and diseased data in categorical and continuous. The list of features considered are: 'Age', 'Sex', 'Chest Pain', 'Resting_BP', 'Cholestrol', 'Fasting_Blood_Sugar', 'Resting_Electrocardiographic', 'Max_Heart_Rate', 'Exercise_Induced_Angina', 'Old_Peak', 'Slope', 'Target'. We have extracted the seven categories of features from the CT scan images such as: 1) Texture Features, 2) Entropy, 3) Local Binary Pattern, 4) Gabor Filter, 5) Intensity-Based Features, 6) Shape Features, and 7) Edge and Boundary Features. The features are listed as: 'Contrast', 'Energy', 'Homogeneity', 'Entropy', 'LBP_Mean', 'Gabor_Mean', 'Mean_Intensity', 'Std_Dev', 'Skewness', 'Kurtosis', 'Area', 'Perimeter', 'Eccentricity', 'Canny_Edge_Count', 'Sobel_Gradient_Magnitude'.

3.2 Data pre-processing

Real-world data is frequently incomplete and unformatted due to the large number of nulls and errors it contains.

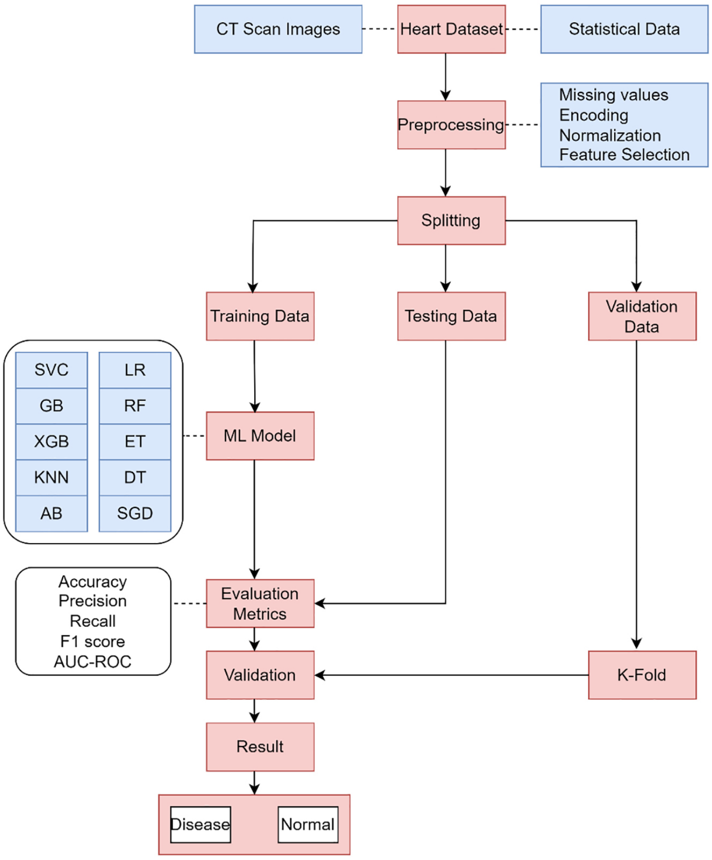

Fig. 1. Working of proposed system

Missing values: To find and handle outliers, this stage entails an exploratory data analysis technique. To deal with missing values, we employ an iterative imputer. More accurate imputations are ensured by the iterative imputation procedure, which modifies one feature in light of the others.

Normalization of data: The dataset's data fields could have different types. Every data value needs to be standardized to the same type for categorization to be effective. For accurate results, we scale parameter values between zero and one. The values in the numerical columns are converted to a standard scale without being changed or removed.

Feature selection: Choosing the right features is essential to enhancing model performance. Sequential forward selection (SFS) is what we use to weed out less important traits. Because of its interpretability and performance advantages, SFS was chosen over PCA for the prediction of heart disease using statistics and CT scan image datasets. Because SFS chooses real features instead of altering them, physicians will find the results more significant. It assesses each feature according to how well it contributes to increasing model accuracy, making sure that predictive and pertinent variables are kept. SFS is model-specific and performs well with both structured and extracted image features, in contrast to PCA, which is unsupervised and may miss crucial information. Better accuracy, less overfitting, and the preservation of crucial clinical insights were the results of this strategy.

3.3 Dataset splitting

There are two partitions for the patient dataset: one for testing (20%) and one for training (80%). The ML model is constructed using the training set, and its performance is assessed using the testing set.

3.4 Classification model

We have utilized several well-known ML algorithms for the screening and classification of heart disease. These algorithms includes, KNN, SVC, DT, random forest (RF), extra trees (ET), AdaBoost (AB), LR, gradient boosting (GB), stochastic gradient descent (SGD), extreme gradient boosting (XGBoost). All the ML algorithms have been implemented on both datasets and ensemble them for better accuracy.

The application of ensemble approaches, which integrate several ML models to increase accuracy and resilience, is emphasized in recent developments in the prediction of cardiac disease. In order to decrease overfitting and boost reliability, methods such as RF and GB construct numerous DT and aggregate their predictions. The goal of AB and XGBoost is to improve misclassified predictions by successively modifying weights. These ensemble approaches work very well with a variety of data sources, including CT scan pictures and clinical information. These methods can improve real-world applicability in cardiac disease prediction, decrease model bias, and increase accuracy by utilizing the advantages of several algorithms.

3.5 Model evaluation metrics

Evaluation metrics play a crucial role in evaluating how well predictive models perform in categorization tasks. They offer insights into a model's effectiveness and assist in quantifying how well it accomplishes its objectives. We can assess the performance of the model by contrasting its predictions with the actual results on the training and testing datasets. The performance evaluation metrics include accuracy, recall, precision, F1-score, confusion matrix, and AUC-ROC.

3.6 Model validation

The dataset is split into K equal parts using this procedure, with the remaining data being used for training and each component being tested once. To guarantee thorough validation, this procedure is carried out again for every component of the dataset. The model's performance is then carefully assessed and validated by comparing the outcomes from each iteration with the real results utilizing the entire dataset. The model can also be evaluated based on randomly generated data values depending on the actual samples.

4 EXPERIMENTAL RESULTS AND DISCUSSION

To efficiently run and evaluate the suggested model, a Windows 10 PC equipped with a 2.9 GHz Core i7 processor, 8 GB RAM, Intel HD Graphics 620 GPU, and 5 GB of disk space was used.

Prior to implementation, it is imperative to optimize the performance of a ML model in order to guarantee optimal precision. The model's learning behaviour is controlled by hyper parameters, which must be properly adjusted for this optimization. To fine-tune the model for improved performance, it is usually necessary to fit the model. grid search CV is an effective method for hyper-parameter optimization. Using this method, a thorough grid of possible hyper-parameter values is created, and each combination is assessed.

4.1 Performance analysis of machine learning models

Every ML classification algorithm is executed based on the parameters. The best model is identified by using the best estimator applied to the mentioned parameters. Further the best model of algorithms are tested based on the 'new synthetic (random)' data generated using random over sampler () method.

The summary shown in Table 1 uses metrics like accuracy, recall, and F1-score for two-class (0 and 1) prediction to compare the classification performance of different ML models on two datasets: Dataset 1 (CT scan images) and Dataset 2 (statistical). Models for Dataset 1 typically perform well and with high accuracy. KNN and SVC show good accuracy, recall, and F1-scores in both classes, with accuracies of 95% and 97%, respectively. With accuracy ranging from 94% to 99%, DT, RF, and ET also demonstrate their competence in managing the dataset's features. Both AB and GB receive excellent overall grades and flawless accuracy for class 0. On the other hand, Dataset 2 exhibits marginally different model performances. DT and LR perform worse, with accuracies of 88% and 86%, respectively, while KNN and SVC maintain good accuracy of 95% and 90%, respectively. On the other hand, ensemble techniques such as GB, XGB, ET, RF, and GB continue to perform well, with accuracy levels between 93% and 96%. Because Dataset 1 presumably has better pre-processing or features that are more appropriate, overall, it provides more favourable conditions for model performance. Across both datasets, ensemble approaches consistently beat simpler models, highlighting their applicability to difficult classification problems. The performance variations underscore the significance of dataset attributes and the possible requirement for customized model selection and optimization tactics grounded in particular data subtleties.

Table 1. Classification reports of all machine learning models

	Class	Dataset 1 (CT Scan)			Dataset 2 (Statistical)		
		Precision	Recall	F1-score	Precision	Recall	F1-score
KNN	0	0.92	0.96	0.94	0.96	0.93	0.95
	1	0.97	0.94	0.96	0.95	0.97	0.96
SVC	0	0.96	0.96	0.96	0.88	0.91	0.89
	1	0.97	0.97	0.97	0.92	0.90	0.91
DT	0	0.92	0.94	0.93	0.84	0.91	0.86
	1	0.96	0.94	0.95	0.90	0.86	0.88
RF	0	0.98	0.99	0.98	0.96	0.93	0.95
	1	0.99	0.98	0.99	0.95	0.97	0.96
ET	0	0.99	0.99	0.99	0.92	0.93	0.93
	1	0.99	0.99	0.99	0.95	0.93	0.94
AB	0	1.0	1.0	1.0	0.88	0.88	0.88
	1	1.0	1.0	1.0	0.90	0.90	0.90
LR	0	0.95	0.94	0.95	0.85	0.84	0.85
	1	0.96	0.97	0.96	0.87	0.88	0.87
SGD	0	0.94	0.98	0.96	0.85	0.82	0.83
	1	0.98	0.96	0.97	0.86	0.88	0.87
GB	0	0.98	0.95	0.96	0.96	0.95	0.96
	1	0.97	0.98	0.97	0.96	0.97	0.97
XGB	0	0.98	0.98	0.98	0.94	0.92	0.93
	1	0.98	0.98	0.98	0.93	0.95	0.94

Results with synthetic data using the random over sampler method are included in Table 2, which also compares the accuracy of several ML algorithms on two datasets. On Dataset 1, models such as KNN, RF, ET, AB, and GB get 100% training accuracy and high testing accuracy (95%-99%). Moreover, SVC, DT, LR, SGD, and XGB exhibit strong performance. Dataset 2 shows a modest reduction in performance. While SVC declines to 90% testing accuracy, RF and KNN maintain 100% training accuracy and 95% testing accuracy. ET drops to 94% and DT drops to 88% testing accuracy. LR, SGD, and AB exhibit significant drops, with testing accuracies ranging from 85% to 89%. For all classifiers, synthetic data consistently increases testing accuracy.

Table 2. Accuracy of machine learning algorithms on different datasets

	Dataset 1 (CT Scan)			Dataset 2 (Statistical)		
	Training Accuracy	Testing Accuracy	Synth. Accuracy	Training Accuracy	Testing Accuracy	Synth. Accuracy
KNN	100%	95%	99.82%	100%	95%	99.12%
SVC	100%	97%	99.39%	96%	90%	94.91%
DT	96%	94%	97.24%	98%	88%	98.01%
RF	100%	98%	99.74%	100%	95%	99.20%
ET	100%	99%	99.74%	100%	94%	98.96%
AB	100%	100%	99.65%	95%	89%	94.19%
LR	97%	95%	95.86%	89%	86%	89.62%
SGD	97%	96%	98.56%	92%	85%	93.22%
GB	100%	97%	99.39%	99%	96%	99.20%
XGB	99%	98%	99.65%	99%	93%	98.48%

4.2 Performance assessment of machine learning models at each fold

By testing a model on different subsets of the data, cross-validation assesses its performance and offers information on its generalization and accuracy. Overfitting may be indicated by accuracy loss, which is the difference between training and validation accuracies. We can determine which models generalize well or badly by examining the accuracy and accuracy loss mean and variation across folds. The accuracy and loss of several ML models across 10 times for two datasets. Models such as KNN, SVC, DT, RF, ET, AB, LR, SGD, GB, and XGB perform well for Dataset 1, with accuracies ranging from 96% to 99% and negligible losses (0.01–0.05). More variance is seen in Dataset 2, as KNN and RF maintain high accuracy at 93% while SVC and DT decline to 88% and 86%, respectively. While LR and SGD exhibit more losses, suggesting inferior performance, ET, GB, and XGB continuously exhibit strong performance. The summary of performance results are shown in Table 3. Figure 2 displays the ROC-AUC scores along with the standard deviation.

Table 3. Accuracy and loss of machine learning algorithms at each fold

	Dataset		F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean Value
KNN	**1**	Acc	0.97	0.93	0.99	0.99	0.93	0.94	0.96	0.98	0.98	0.98	0.97
		Loss	0.03	0.07	0.01	0.01	0.07	0.06	0.04	0.02	0.02	0.02	0.04
	2	Acc	0.90	0.89	0.89	0.88	0.89	0.99	1.00	0.92	1.00	0.96	0.93
		Loss	0.10	0.11	0.11	0.12	0.11	0.01	0.00	0.08	0.00	0.04	0.07
SVC	**1**	Acc	0.98	0.98	0.98	1.00	0.98	0.99	0.97	0.99	0.99	0.96	0.98
		Loss	0.02	0.02	0.02	0.00	0.02	0.01	0.03	0.01	0.01	0.04	0.02
	2	Acc	0.88	0.89	0.80	0.88	0.86	0.91	0.96	0.84	0.94	0.87	0.88
		Loss	0.12	0.11	0.20	0.12	0.14	0.09	0.04	0.16	0.06	0.13	0.12

(Continued)

Table 3. Accuracy and loss of machine learning algorithms at each fold *(Continued)*

	Dataset		F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Mean Value
DT	1	Acc	0.96	0.95	0.96	0.99	0.94	0.97	0.95	0.98	0.99	0.95	0.96
		Loss	0.04	0.05	0.04	0.01	0.06	0.03	0.05	0.02	0.01	0.05	0.04
	2	Acc	0.72	0.83	0.81	0.76	0.86	0.93	0.99	0.86	0.97	0.90	0.86
		Loss	0.28	0.17	0.19	0.24	0.14	0.07	0.01	0.14	0.03	0.10	0.14
RF	1	Acc	0.98	0.96	1.00	0.99	0.97	0.98	0.99	0.99	0.98	0.96	0.98
		Loss	0.02	0.04	0.00	0.01	0.03	0.02	0.01	0.01	0.02	0.04	0.02
	2	Acc	0.84	0.87	0.83	0.87	0.88	0.76	0.79	0.77	0.85	0.76	0.82
		Loss	0.16	0.13	0.17	0.13	0.12	0.24	0.21	0.23	0.15	0.24	0.18
ET	1	Acc	0.98	0.97	0.99	0.99	0.98	0.97	0.96	0.99	0.97	0.95	0.98
		Loss	0.02	0.03	0.01	0.01	0.02	0.03	0.04	0.01	0.03	0.05	0.03
	2	Acc	0.92	0.88	0.88	0.90	0.90	0.99	1.00	0.93	1.00	0.95	0.94
		Loss	0.08	0.12	0.12	0.10	0.10	0.01	0.00	0.07	0.00	0.05	0.06
AB	1	Acc	0.98	0.98	0.99	1.00	0.98	0.99	0.98	0.99	0.98	1.00	0.99
		Loss	0.02	0.02	0.01	0.00	0.02	0.01	0.02	0.01	0.02	0.00	0.01
	2	Acc	0.89	0.86	0.85	0.84	0.84	0.89	0.93	0.81	0.96	0.86	0.87
		Loss	0.11	0.14	0.15	0.16	0.16	0.11	0.07	0.19	0.04	0.14	0.13
LR	1	Acc	0.97	0.95	0.98	0.99	0.95	0.94	0.97	0.97	0.95	0.96	0.96
		Loss	0.03	0.05	0.02	0.01	0.05	0.06	0.03	0.03	0.05	0.04	0.04
	2	Acc	0.84	0.87	0.83	0.87	0.88	0.76	0.79	0.77	0.85	0.76	0.82
		Loss	0.16	0.13	0.17	0.13	0.12	0.24	0.21	0.23	0.15	0.24	0.18
SGD	1	Acc	0.97	0.94	0.97	0.99	0.94	0.92	0.96	0.97	0.95	0.93	0.95
		Loss	0.03	0.06	0.03	0.01	0.06	0.08	0.04	0.03	0.05	0.07	0.05
	2	Acc	0.87	0.87	0.90	0.86	0.82	0.76	0.78	0.76	0.67	0.72	0.80
		Loss	0.13	0.13	0.10	0.14	0.18	0.24	0.22	0.24	0.33	0.28	0.20
GB	1	Acc	0.98	0.96	0.98	1.00	0.97	0.97	0.99	0.98	0.98	0.99	0.98
		Loss	0.02	0.04	0.02	0.00	0.03	0.03	0.01	0.02	0.02	0.01	0.02
	2	Acc	0.92	0.90	0.89	0.89	0.87	0.97	1.00	0.88	1.00	0.93	0.93
		Loss	0.08	0.10	0.11	0.11	0.13	0.03	0.00	0.12	0.00	0.07	0.07
XGB	1	Acc	0.98	0.96	0.99	1.00	0.97	0.98	0.98	0.99	0.98	0.97	0.98
		Loss	0.02	0.04	0.01	0.00	0.03	0.02	0.02	0.01	0.02	0.03	0.02
	2	Acc	0.92	0.92	0.85	0.87	0.89	0.97	1.00	0.93	1.00	0.94	0.93
		Loss	0.08	0.08	0.15	0.13	0.11	0.03	0.00	0.07	0.00	0.06	0.07

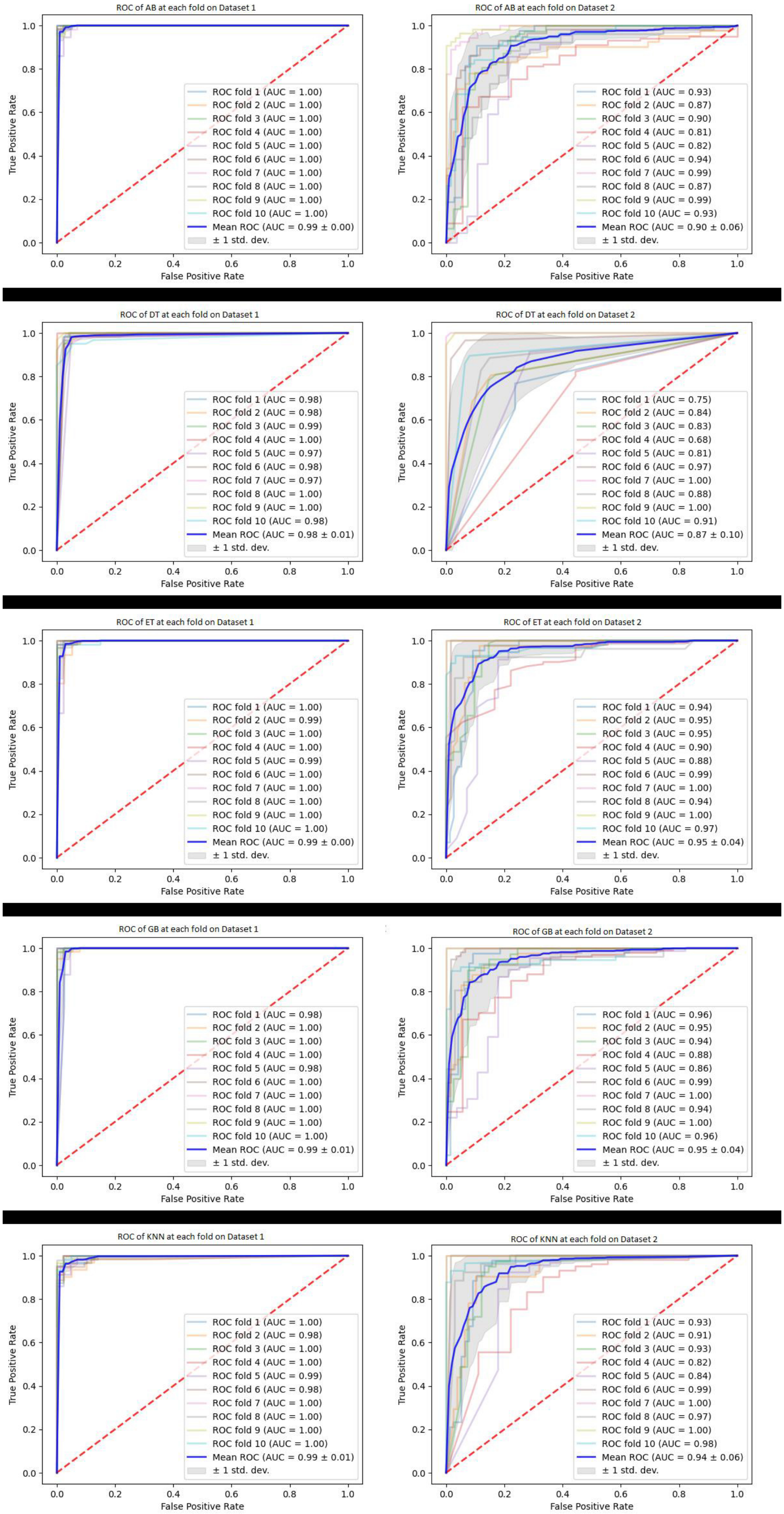

Fig. 2. ROC-AUC curves at each fold

4.3 Performance assessment of ensemble classifier

The methodology that is being suggested utilizes an ensemble approach to merge several ML models, producing an output that is more accurate as a whole than it would be with just one model. Soft voting is the ensemble method employed here, averaging the estimated probabilities from six distinct models to arrive at a final forecast. When averaging probabilities instead of solely taking into account the majority vote, soft voting "gives more weight" to confident forecasts, which is why it frequently performs better than hard voting. Mean majority voting and weighted majority voting are investigated in soft voting ensembles. To improve forecast accuracy, the soft voting ensemble (SVE) makes use of each model's unique capabilities. In addition, the SVE makes data mistakes and outliers more tolerant and lowers the chance of overfitting.

Let $P_{i,j}$ be the probability of class j for a given test instance, as predicted by model i. The formula used to determine the average projected probability for each class across all models.

$$P_j = \frac{1}{n}\sum_{i=1}^{n} P_{i,j} \tag{1}$$

Where n (in this case, ten) is the number of models. The class prediction with the highest average probability is the final one to be made:

$$\hat{y} = \arg max_j P_j \tag{2}$$

Table 4 gives the performance of ensemble classifiers on both considered datasets. The ensemble classifier's remarkable 0.98 accuracy for Dataset 1 was attained. With a precision of 0.99, which was likewise high, the classifier had an extremely low false positive rate. With a recall of 0.98, the model's ability to recognize true positives is demonstrated. The F1 score, which takes memory and precision into account, was 0.99. The AUC achieved a flawless 1.00, indicating exceptional ability to discriminate between classes. Although marginally lower, the performance for Dataset 2 was still noteworthy. With an accuracy of 0.95 and a precision of 0.94, the true positive and false positive rates were well-balanced. With a recall of 0.97, the model demonstrated a high degree of ability to identify true positives. AUC was 0.97 and the F1 score matched the accuracy of 0.95, indicating strong classification performance. The overall ROC-AUC is shown in Figure 3.

Table 4. Performance of ensemble classifiers

Metric	Dataset 1 (CT scan)	Dataset 2 (Statistical)
Accuracy	0.98	0.95
Precision	0.99	0.94
Recall	0.98	0.97
F1 Score	0.99	0.95
AUC	1.00	0.97

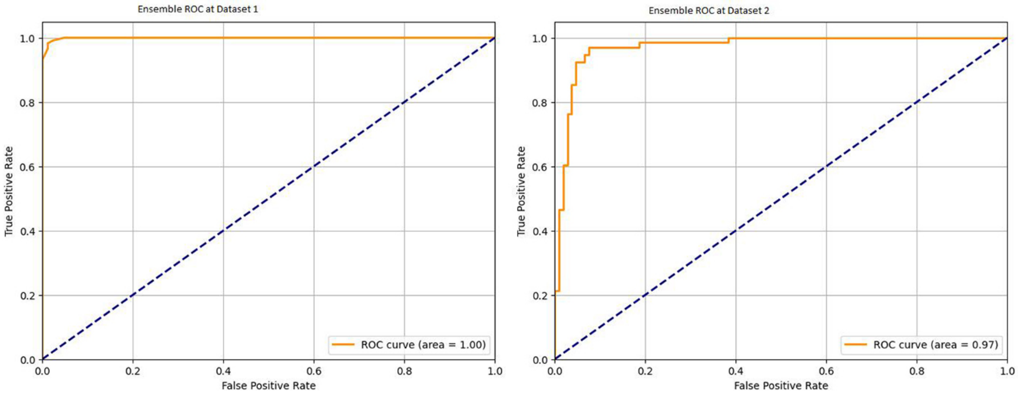

Fig. 3. ROC-AUC curves at of ensemble model

4.4 Performance comparison

Several studies on the statistical dataset demonstrate how ML classifiers significantly impact the accuracy of heart disease detection. One study, for example, used SGD, LR, SVM, NB, and ensemble approaches to obtain 93% accuracy, while another used RF, LR, and SVM to get 92%. While the CART approach obtained 87%, a combination of NN, MLPNN, AB, SVM, LR, ANN, and RF produced 93.9%. The combination of RF, KNN, LR, NB, GB, AB, and a Soft Voting Ensemble yielded the greatest accuracy, 95%. Using classifiers such as KNN, SVC, DT, RF, and a soft ensemble, the suggested model demonstrated 98% accuracy in tests conducted on the Indian and Cleveland datasets. It is shown in Figure 4.

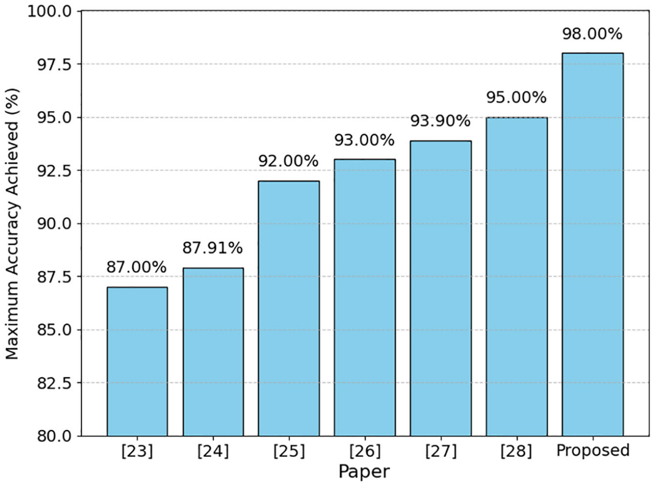

Fig. 4. Performance comparison with previous works

5 CONCLUSION

This work presents a successful ML method for heart disease diagnosis. The models were optimized with the grid search CV hyperparameter optimization technique and cross-validated 10 times on TWO distinct datasets (statistical and CT scan image dataset) in order to achieve optimal accuracy. Based on measures including accuracy, precision, recall, and F1-score, ten ML classifiers were evaluated and contrasted. The results showed that on the CT scan image dataset, the AB classifier had the highest accuracy rate at 100%, while on the statistical dataset, GB had the best performance at 96%. Using random or generated data to evaluate the model has increased its accuracy. Notably, accuracies were further increased to 98% and 95% on the corresponding datasets by applying the soft voting ensemble classifier to all 10 models. The ML model must be included into useful apps in order to enable real-time cardiac disease prediction. Web applications, mobile apps, or other software systems can be used to achieve this. The model could be used to determine a patient's risk of heart disease by being used in real-world settings like clinics or hospitals.

6 REFERENCES

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

[2] R. P. Choudhury and N. Akbar, "Beyond diabetes: A relationship between cardiovascular outcomes and glycaemic index," *Cardiovascular Research,* vol. 117, no. 8, pp. e97–e98, 2021. https://doi.org/10.1093/cvr/cvab162

[3] C. Ordonez, "Association rule discovery with the train and test approach for heart disease prediction," *IEEE Trans. Inform. Technol. Biomed*, vol. 10, no. 2, pp. 334–343, 2006. https://doi.org/10.1109/TITB.2006.864475

[4] G. Magesh and P. Swarnalatha, "RETRACTED ARTICLE: Optimal feature selection through a cluster-based DT learning (CDTL) in heart disease prediction," *Evol. Intel.*, vol. 14, pp. 583–593, 2020. https://doi.org/10.1007/s12065-019-00336-0

[5] K. Rohit Chowdary, P. Bhargav, N. Nikhil, K. Varun, and D. Jayanthi, "Early heart disease prediction using ensemble learning techniques," *Journal of Physics: Conference Series*, vol. 2325, p. 012051, 2022. https://doi.org/10.1088/1742-6596/2325/1/012051

[6] J. Liu, X. Dong, H. Zhao, and Y. Tian, "Predictive classifier for cardiovascular disease based on stacking model fusion," *Processes*, vol. 10, no. 4, p. 749, 2022. https://doi.org/10.3390/pr10040749

[7] B. N. Ravi Kumar, B. J. Ambika, Naveen Chandra Gowda, H. N. Veena, B. Ben Sujitha, and D. Roja Ramani, "An efficient breast cancer detection using machine learning classification models," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 20, no. 13, pp. 24–40, 2024. https://doi.org/10.3991/ijoe.v20i13.50289

[8] H. N. Veena, K. K. Patil, P. Vanajakshi, A. Ambore, and N. C. Gowda, "An enhanced RNN-LSTM model for fundus image classification to diagnose glaucoma," *SN Computer Science*, vol. 5, 2024. https://doi.org/10.1007/s42979-024-02867-5

[9] L. Shalini, S. S. Manvi, B. Gardiner, and N. C. Gowda, "Image based classification of COVID-19 infection using ensemble of machine learning classifiers and deep learning techniques," in *2022 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI),* 2022, pp. 1–6. https://doi.org/10.1109/ICDSAAI55433.2022.10028859

[10] A. R. Sohara Banu, V. Nagaveni, N. C. Gowda, T. Yerriswamy, M. Gururaj, and N. Varshitha, "Well Bot—A reinforcement learning and sentiment analysis-based chatbot implementation for mental well-being," in *Soft Computing and Signal Processing (ICSCSP 2023)*, in Lecture Notes in Networks and Systems, V. S. Reddy, J. Wang, and K. Reddy, Eds., Springer, Singapore, vol. 864, 2024, pp. 471–478. https://doi.org/10.1007/978-981-99-8628-6_40

[11] O. Iparraguirre-Villanueva, J. Andia-Alcarraz, F. Saba-Estela, and A. Epifanía-Huerta, "Mobile application with augmented reality as a support tool for learning human anatomy," *International Journal of Engineering Pedagogy (iJEP)*, vol. 14, no. 1, pp. 82–95, 2024. https://doi.org/10.3991/ijep.v14i1.46845

[12] V. Pavithra and V. Jayalakshmi, “Hybrid feature selection technique for prediction of cardiovascular diseases,” *Materials Today: Proceedings*, vol. 81, pp. 336–340, 2023. https://doi.org/10.1016/j.matpr.2021.03.225

[13] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, “Secure and robust machine learning for healthcare: A survey,” *IEEE Rev. Biomed. Eng.,* vol. 14, pp. 156–180, 2021. https://doi.org/10.1109/RBME.2020.3013489

[14] Ch. Anwar ul Hassan *et al.*, “Effectively predicting the presence of coronary heart disease using machine learning classifiers,” *Sensors*, vol. 22, no. 19, p. 7227, 2022. https://doi.org/10.3390/s22197227

[15] R. Katarya and S. K. Meena, “Machine learning techniques for heart disease prediction: A comparative study and analysis,” *Health Technol.*, vol. 11, pp. 87–97, 2020. https://doi.org/10.1007/s12553-020-00505-7

[16] M. Gjoreski, M. Simjanoska, A. Gradisek, A. Peterlin, M. Gams, and G. Poglajen, “Chronic heart failure detection from heart sounds using a stack of machine-learning classifiers,” in *2017 International Conference on Intelligent Environments (IE)*, IEEE, 2017, pp. 14–19. https://doi.org/10.1109/ie.2017.19

[17] N. S. Mansur Huang, Z. Ibrahim, and N. Mat Diah, “Machine learning techniques for heart failure prediction,” *Malaysian Journal of Computing (MJoC)*, vol. 6, no. 2, pp. 872–884, 2021. https://doi.org/10.24191/mjoc.v6i2.13708

[18] S. P. Patro, G. S. Nayak, and N. Padhy, “Heart disease prediction by using novel optimization algorithm: A supervised learning prospective,” *Informatics in Medicine Unlocked*, vol. 26, p. 100696, 2021. https://doi.org/10.1016/j.imu.2021.100696

[19] M. Trigka and E. Dritsas, “Long-term coronary artery disease risk prediction with machine learning models,” *Sensors*, vol. 23, no. 3, p. 1193, 2023. https://doi.org/10.3390/s23031193

[20] S. J. Pasha and E. S. Mohamed, “Novel Feature Reduction (NFR) model with machine learning and data mining algorithms for effective disease risk prediction,” *IEEE Access*, vol. 8, pp. 184087–184108, 2020. https://doi.org/10.1109/ACCESS.2020.3028714

[21] A. Gupta, R. Kumar, H. Singh Arora, and B. Raman, “MIFH: A machine intelligence framework for heart disease diagnosis,” *IEEE Access*, vol. 8, pp. 14659–14674, 2020. https://doi.org/10.1109/ACCESS.2019.2962755

[22] P. Rani, R. Kumar, N. M. O. S. Ahmed, and A. Jain, “A decision support system for heart disease prediction based upon machine learning,” *J. Reliable. Intell. Environ.,* vol. 7, pp. 263–275, 2021. https://doi.org/10.1007/s40860-021-00133-6

[23] M. Ozcan and S. Peker, “A classification and regression tree algorithm for heart disease modeling and prediction,” *Healthcare Analytics*, vol. 3, p. 100130, 2023. https://doi.org/10.1016/j.health.2022.100130

[24] G. N. Ahamad *et al.*, “Influence of optimal hyperparameters on the performance of machine learning algorithms for predicting heart disease,” *Processes*, vol. 11, no. 3, p. 734, 2023. https://doi.org/10.3390/pr11030734

[25] R. Yilmaz and F. H. Yağin, “Early detection of coronary heart disease based on machine learning methods,” *Medical Records,* vol. 4, no. 1, pp. 1–6, 2022. https://doi.org/10.37990/medr.1011924

[26] F. Rustam, A. Ishaq, K. Munir, M. Almutairi, N. Aslam, and I. Ashraf, “Incorporating CNN features for optimizing performance of ensemble classifier for cardiovascular disease prediction,” *Diagnostics*, vol. 12, no. 6, p. 1474, 2022. https://doi.org/10.3390/diagnostics12061474

[27] B. P. Doppala, D. Bhattacharyya, M. Janarthanan, and N. Baik, “A reliable machine intelligence model for accurate identification of cardiovascular diseases using ensemble techniques,” *Journal of Healthcare Engineering*, vol. 2022, no. 1, p. 2585235, 2022. https://doi.org/10.1155/2022/2585235

[28] N. Chandrasekhar and S. Peddakrishna, “Enhancing heart disease prediction accuracy through machine learning techniques and optimization,” *Processes*, vol. 11, no. 4, p. 1210, 2023. https://doi.org/10.3390/pr11041210

[29] M. Siddhartha, “Heart Disease Dataset (Comprehensive),” IEEE Dataport, 2020, https://dx.doi.org/10.21227/dz4t-cm36

[30] N. Tomar, “CT Heart Dataset,” kaggle, 2021. https://www.kaggle.com/datasets/nikhilroxtomar/ct-heart-segmentation

7 AUTHORS

Archana P. (Research Scholar) earned a Bachelor of Engineering degree in Information Science and Engineering (ISE) from Visvesvaraya Technological University, Belagavi, in 2009. She obtained her Master’s in Computer Science and Engineering (CSE) from Visvesvaraya Technological University, Belagavi, in 2012. She is currently pursuing a Ph.D degree in Computer Science and Engineering and has been working as an Assistant Professor at Adichunchanagiri Institute of Technology for the past 13 years. She has published more than 20 research papers in reputed International and National Journals and Conferences. Her interest areas are Image Processing, Artificial Intelligence, Data Mining, Machine Learning and Deep Learning (E-mail: archanahavishsurya@gmail.com).

Dr. Shashikala S. V. (Research Supervisor) is currently serving as the Professor and Head of the Department of Computer Science and Engineering at BGS Institute of Technology, Adichunchanagiri University, B.G. Nagara, Nagamangala Taluk, Karnataka. She has 30 years of teaching experience in the field of computer science education. She holds a Master’s degree in Computer Science and Engineering and Ph.D. in the same field from Visvesvaraya Technological University (VTU), Belagavi. Her scholarly contributions include 30 publications in reputed International and National Journals and Conferences. She has also authored two book chapters, filed three patents—one of which has granted. Actively involved in academic governance, she serves as a member of Board of Studies (BOS), Board of Examiners (BOE), and several other professional bodies and academic councils. She has successfully organized multiple International and National Conferences, Hackathons, Workshops, Faculty Development Programs (FDPs), Seminars, and Industrial visits. Dr. Shashikala has secured funding, including ₹60 lakhs from VGST and ₹20,000 from KSCST for student projects. Her research interests include Artificial Intelligence, Machine Learning, Data Mining, Networking, and Image Processing (E-mail: shashisv7@gmail.com).