Journal of Carcinogenesis
Covering Genesis to Therapeutics

Journal of Carcinogenesis, ISSN: 1477-3163
2025; Vol 24: Issue 3s, https://carcinogenesis.com/

Hyperparameter Tuning in Machine Learning Classifiers: Performance Insights for CVD Prediction

Maitri Bhavsar[1]*, Dr. Manish Patel[2]

[1]*Assistant professor, Sankalchand Patel College of Engineering, Sankalchand Patel University, Visnagar, India mhbai_spce@spu.ac.in

[2]Professor, Sankalchand Patel College of Engineering, Sankalchand Patel University, Visnagar, India, mmpatelit_spce@spu.ac.in

***Corresponding Author:** Maitri Bhavsar

*Assistant professor, Sankalchand Patel College of Engineering, Sankalchand Patel University, Visnagar, India

ABSTRACT

Globally, Cardiovascular Diseases (CVDs) continue to be a leading cause of morbidity and mortality, placing a significant burden on healthcare systems. Managing the risk factors linked to CVDs requires early identification and resolution. In this article, machine learning techniques (MLT) for predicting heart disease were presented. This approach presents the effects of dimension reduction and hyper-parameter adjustment on the prediction of heart disease. Furthermore, to aid in the prediction of cardiovascular diseases (SVCs), six classifiers are used: Random Forest (RF), Gradient Boosting (GB), Support Vector Classifier (SVC), Decision Tree (DT), Adaboost (AB), ensemble classifier (EC), and one type of dimensional reduction model based on principal component analysis (PCA). Grid search cross validation (GridSearchCV) is used as an automated hyper-parameter tuning algorithm. Thirty percent of the 920 patients with 14 attributes used in the Jupyter Notebook are test data. Evaluation measures that include precision, accuracy, recall, F1-score, confusion matrix, and receiver operating characteristic (ROC) curve are used to assess the performance of six classifiers.

Keywords: *Hyper-parameter tuning, classification algorithm, prediction, heart disease, and dimension reduction.*

How to Cite: Maitri Bhavsar, Dr. Manish Patel, (2025) Hyperparameter Tuning in Machine Learning Classifiers: Performance Insights for CVD Prediction, *Journal of Carcinogenesis, Vol.24, No.3s*, 583-592.

1. INTRODUCTION

The World Health Organization (WHO) cardiovascular diseases (CVDs) are a leading cause of death, accounting for 32% of all deaths globally [1]. In order to lessen the catastrophic consequences of CVDs, early identification and prediction of cardiovascular risks are now crucial. In order to forecast cardiovascular events, traditional risk prediction techniques like the Framingham Heart Study and the SCORE model frequently rely on fundamental demographic characteristics (such as age, gender, blood pressure, and cholesterol levels). Although helpful, these conventional models are constrained in their capacity to take into consideration intricate, non-linear interactions in the data and disregard a number of personal health factors that may influence the risk of CVD.

In the last few years, Machine Learning (ML) approaches have drawn a lot of attention as a potential remedy for enhancing cardiovascular risk prediction due to their ability to manage large and intricate datasets. More precise and individualized predictions are made possible by ML algorithms' ability to automatically identify patterns in data that human analysts might miss. For the prediction of cardiovascular disorders, researchers have been investigating a variety of machine learning techniques, from ensemble methods to deep learning, with encouraging findings that frequently surpass those of traditional models.

Ensemble learning techniques, which combine the forecasts from multiple models to increase overall precision and resilience, are given special attention. For the CHD dataset, RF using the TPOT classifier achieved the greatest accuracy of 97.52%. To forecast the CAD using the Z-Alizadeh Sani dataset, Moloud Abdar *et al.* [18] employed three different forms of SVM. Additionally, they presented the N2Genetic optimizer, a novel optimization technique. The accuracy of N2Genetic-nuSVM was 93%, according to the results. Using the Cleveland dataset, James Meng *et al.* [19] presented a clinical knowledge-increased machine learning algorithm to detect Ischemic Heart Disease (IHD). The top SVM-based model achieved 94% accuracy. The performances of several people were examined by Khaled Mohamad Almustafa [20].

© 2025 Journal of Carcinogenesis | Published for Carcinogenesis Press by Wolters Kluwer-Medknow

To predict the likelihood of CVDs, three datasets were pooled, and the DT outperformed other machine learning methods such as the adaBoost (AB) classifier, Extra Trees Classifier (ET), SVM, gradient boost, multi-layer perceptron (MLP), XGB, RF, KNN, and LR, with the highest accuracy of 99.16% in [23].

A range of data preparation methods have been employed by researchers for sample analysis and classification algorithms in all of their previous investigations on the estimation of heart disease. Furthermore, the majority of the time, the effects of dimension reduction were not examined using various datasets and classification techniques. Hyper-parameter tweaking based on GridSearchCV was used in certain research to attain good classification algorithm performance.

Several classification techniques, including RF, GB, support vector classifier (SVC), DT, AB, and ensemble classifier EC, were used in this work to maximize the prediction of heart disease. A dataset comprising 920 patients and 14 attributes, of which 30% are test data, was used to run six algorithms. This manuscript's state-of-the-art is the use of principal component analysis (PCA) with hyper-parameter tuning and untuning methods for dimensional reduction in the prediction of cardiac disease.

The remainder of the manuscript's research is organized as follows: Section 2 cites relevant literature on metrics, hyper-parameter tuning, dimension reduction techniques, and classification techniques. Section 3 displays the study's findings, which include a confusion matrix, GridSearchCV results, and feature reduction results. In the end, parts 4 and 5 presented the discussion and conclusion parts, respectively.

Table 1 Attribute of heart disease database [20].

Attributes	Definition and values
age	The person's age in years
sex	The person's sex (1 = male, 0 = female)
cp	The chest pain experienced (Value 1: typical angina, Value 2: atypical angina, Value 3: non-anginal pain, Value 4: asymptomatic)
trestbps	The person's resting blood pressure (mm Hg on admission to the hospital)
chol	The person's cholesterol measurement in mg/dl
fbs	The person's fasting blood sugar (> 120 mg/dl, 1 = true; 0 = false)
restecg	Resting electrocardiographic measurement (0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria)
thalach	The person's maximum heart rate achieved
exang	Exercise induced angina (1 = yes; 0 = no)
oldpeak	ST depression induced by exercise relative to rest ('ST' relates to positions on the ECG plot. See more here)
slope	the slope of the peak exercise ST segment (Value 1: upsloping, Value 2: flat, Value 3: downsloping)
ca	The number of major vessels (0–3)
thal	A blood disorder called thalassemia (3 = normal; 6 = fixed defect; 7 = reversable defect)
target	Heart disease (0 = no, 1 = yes)

2. RELATED WORK

As a computer tool, MLT are important in the identification of cardiac disease. A few steps can be taken to estimate the risk of heart disease. Data preprocessing, the initial stage, entails looking at the features, null values, duplicate data, and out-of-range values [30]. The popular data analysis library is called Pandas [31]. Second, the performance of the classifier can be enhanced by dimensional reduction, hyper-parameter adjustment, and data standardization [27]. Using the data as input, the proposed model for predicting the likelihood of heart disease is constructed.

This graphic shows that the pipeline method, which included PCA and classification algorithms, was fitted to the X_train and Y_train using two sequences. Additionally, X_train and Y_train were fitted with hyper-parameter tuning. Accuracy, precision, recall, and F1-score are other outcomes of prediction.

2.1 Classification Techniques

A supervised learning method for predicting the appropriate class for a given input is data classification [32]. To make

data easier to find, retrieve, and save for later use, a classifier groups it into classes. The classification model is trained using the initial training data and evaluated using the test data.

2.1.1. RF

To create decision trees, RF chooses random samples. It makes predictions based on each tree and averages them together. Principal features of a dataset are also chosen by this classification technique [34].

2.1.2. GB

A potent MLT that is a member of the ensemble methods family is the Gradient Boosting Classifier. It combines a collection of weak learners, usually decision trees, in a step-by-step manner to create a powerful classifier.

2.1.3. SVC

SVC divides the data into two groups by determining the optimal hyperplane by mapping the data points to a high-dimensional space. To import SVC, an integrated model called SVM is executed. Recent studies have shown that using SVM offers an appropriate option for control purposes [36].

2.1.4. DT

This popular classifier uses a tree-like model of Boolean judgments as a decision assistance tool. Every path starts at the root and decides on a data splitting order until the leaf node yields a Boolean result [33].

2.1.5. Adaboost

In order to create a strong classifier, the ensemble learning algorithm AdaBoost (Adaptive Boosting) combines several weak learners, typically decision stumps or single-split decision trees.

2.1.6. Ensemble Classifier

A machine learning model known as an ensemble classifier makes a final conclusion by combining several basic classifiers, which can be either strong or weak learners.

The premise is that, particularly when the base models have a range of errors, On average, a group of models is more reliable and accurate than a single model.

2.2. Dimension reduction methods

An unsupervised learning technique called the "dimension reduction method" converts high-dimensional data into low-dimensional data while preserving some of the original data's key characteristics. Applying the dimension reduction method has several benefits, such as increasing performance, decreasing the requirement for computer resources, and expediting computational processes [39]. The following is a description of three popular dimension reduction techniques:

Table 2 The best Values performance of models with PCA (hyper parameter).

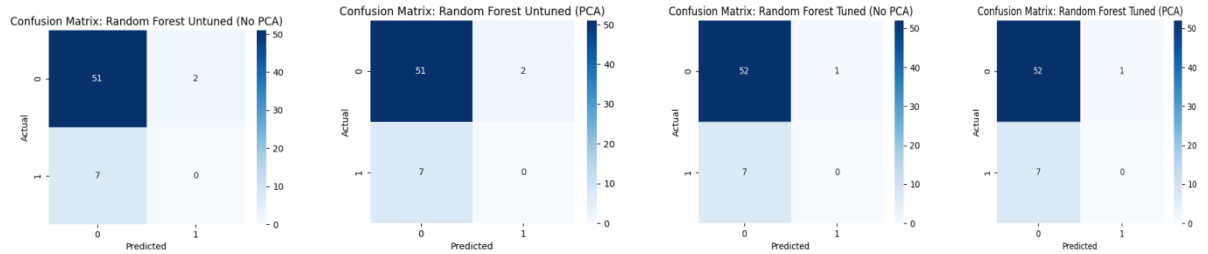
Model	Accuracy	Precision	Recall	F1
Random Forest Untuned (No PCA)	0.766667	0.500000	0.50	0.495192
Random Forest Untuned (PCA)	0.833333	0.685185	0.62	0.639423
Random Forest Tuned (No PCA)	0.766667	0.500000	0.50	0.495192
Random Forest Tuned (PCA)	0.816667	0.627273	0.57	0.580952
Gradient Boosting Untuned (No PCA)	0.800000	0.620192	0.60	0.607843
Gradient Boosting Untuned (PCA)	0.816667	0.648248	0.61	0.623073
Gradient Boosting Tuned (No PCA)	0.816667	0.676252	0.69	0.682540
Gradient Boosting Tuned (PCA)	0.816667	0.648248	0.61	0.623073
SVC Untuned (No PCA)	0.800000	0.640000	0.64	0.640000
SVC Untuned (PCA)	0.816667	0.663399	0.65	0.656071
SVC Tuned (No PCA)	0.800000	0.640000	0.64	0.640000
SVC Tuned (PCA)	0.816667	0.663399	0.65	0.656071
Decision Tree Untuned (No PCA)	0.766667	0.624224	0.66	0.635417
Decision Tree Untuned (PCA)	0.783333	0.598039	0.59	0.593538
Decision Tree Tuned (No PCA)	0.783333	0.639116	0.67	0.650381
Decision Tree Tuned (PCA)	0.783333	0.598039	0.59	0.593538
AdaBoost Untuned (No PCA)	0.800000	0.640000	0.64	0.640000
AdaBoost Untuned (PCA)	0.766667	0.580000	0.58	0.580000
AdaBoost Tuned (No PCA)	0.800000	0.656250	0.68	0.666048

© 2025 Journal of Carcinogenesis | Published for Carcinogenesis Press by Wolters Kluwer-Medknow

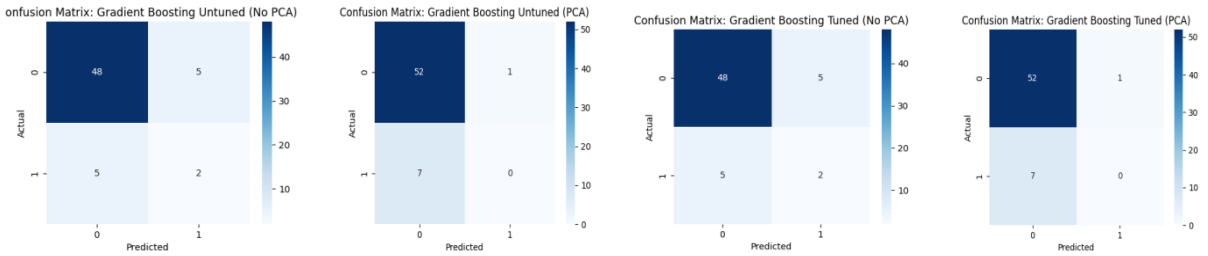
AdaBoost Tuned (PCA)	0.800000	0.656250	0.68	0.666048
Ensemble Untuned (No PCA)	0.783333	0.598039	0.59	0.593538
Ensemble Untuned (PCA)	0.816667	0.648248	0.61	0.623073
Ensemble Tuned (No PCA)	0.816667	0.676252	0.69	0.682540
Ensemble Tuned (PCA)	0.783333	0.567385	0.55	0.554540

2.2.1. PCA

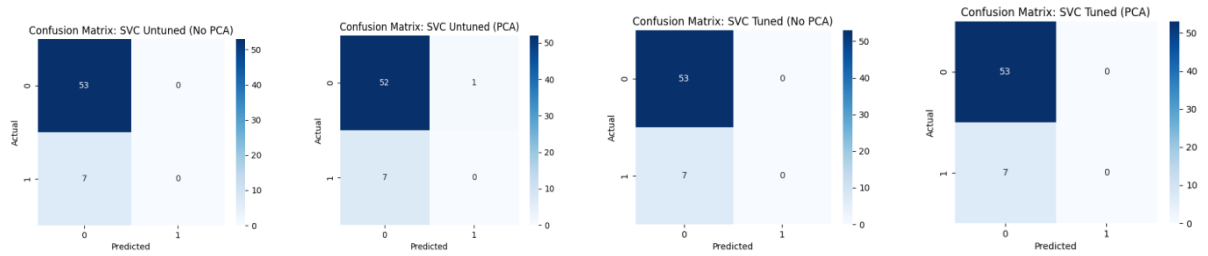
In four phases, PCA is used to minimize information loss and reduce the dimensionality of a dataset. Initially, the dataset is standardized. Secondly, the features' covariance matrix is computed. The covariance matrix's eigenvalues and eigenvectors are then ascertained. The eigenvalues and their matching eigenvectors are finally sorted. This algorithm searches for directions with the maximum variance regardless of labels. [40,41].



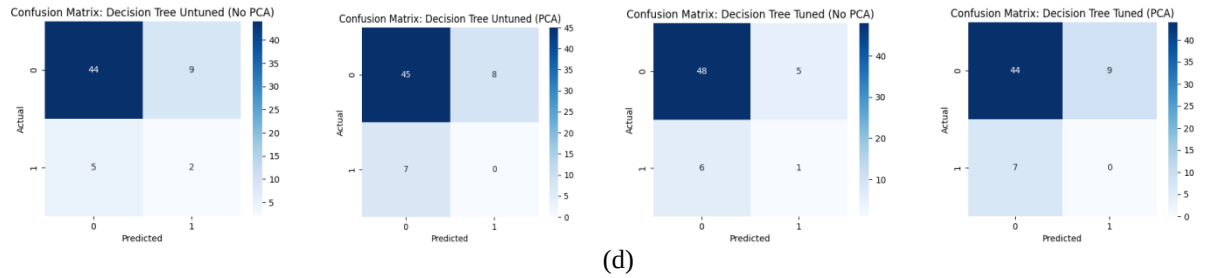
(a)



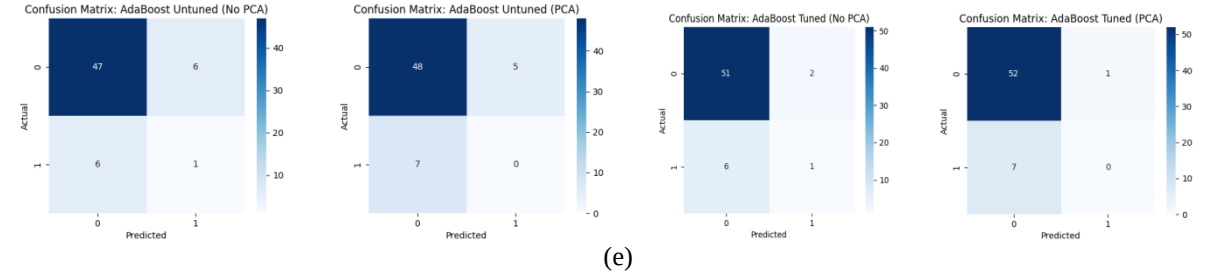
(b)



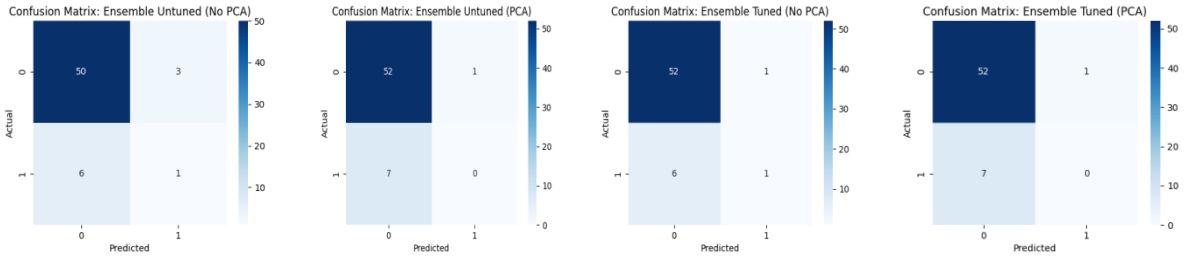
(c)



(d)



(e)



(f)

Fig. 1. Confusion matrix of classification techniques using PCA with Hyper Parameter Tunning on dimensionally reduced datasets. (a) RF. (b) GB. (c) SVC. (d) DT. (e) AB, (f) Ensemble Classifier.

2.2.2. SMOTE

A data preparation technique called SMOTE (Synthetic Minority Over-sampling Technique) is utilized in machine learning datasets to rectify class imbalance. The model is frequently skewed in favor of the majority class when trained directly on such data. SMOTE produces artificial samples rather than merely replicating the minority class's samples.

Our evaluation's goal is to draw attention to the SMOTE technique's contribution to the creation of highly precise and trustworthy ML models. We tested popular ML models, including NB, LR, RotF, MLP, KNN, J48, bagging, RF, voting, and stacking, in this direction. We assessed each models's accuracy, recall, precision, and AUC following ten-fold cross-validation, both with and without SMOTE.

Table 3 The best Values performance of models with SMOTE (NO PCA).

Model	Accuracy	Precision	Recall	F1
Random Forest Untuned	0.850000	0.375000	0.428571	0.400000
Random Forest Tuned	0.800000	0.272727	0.428571	0.333333
Gradient Boosting Untuned	0.816667	0.250000	0.285714	0.266667
Gradient Boosting Tuned	0.833333	0.333333	0.428571	0.375000
SVC Untuned	0.783333	0.200000	0.285714	0.235294
SVC Tuned	0.783333	0.200000	0.285714	0.235294
Decision Tree Untuned	0.800000	0.272727	0.428571	0.333333
Decision Tree Tuned	0.783333	0.250000	0.428571	0.315789
AdaBoost Untuned	0.766667	0.181818	0.285714	0.222222
AdaBoost Tuned	0.816667	0.333333	0.571429	0.421053
Ensemble Untuned	0.800000	0.222222	0.285714	0.250000
Ensemble Tuned	0.816667	0.300000	0.428571	0.352941

Additionally, Table 3 shows that the ML models attained very high-performance metrics following the application of SMOTE. When considering the obtained experimental results of the recall metric, the decrease in false-negative predictions makes it significantly better than the No-SMOTE case. This is crucial and plays a major part in creating effective machine learning models and methods.

2.3. Hyper-parameter tuning

GridSearchCV is a ML approach that uses hyper-parameter tweaking to find the optimal values for a particular model. To generate hyper-parameters, this method makes use of two variables: The quantity of layers and the pace of learning. After defining a selection of values, the hyper-parameter combination is estimated. The optimal hyper-parameters are ultimately chosen and processed [35].

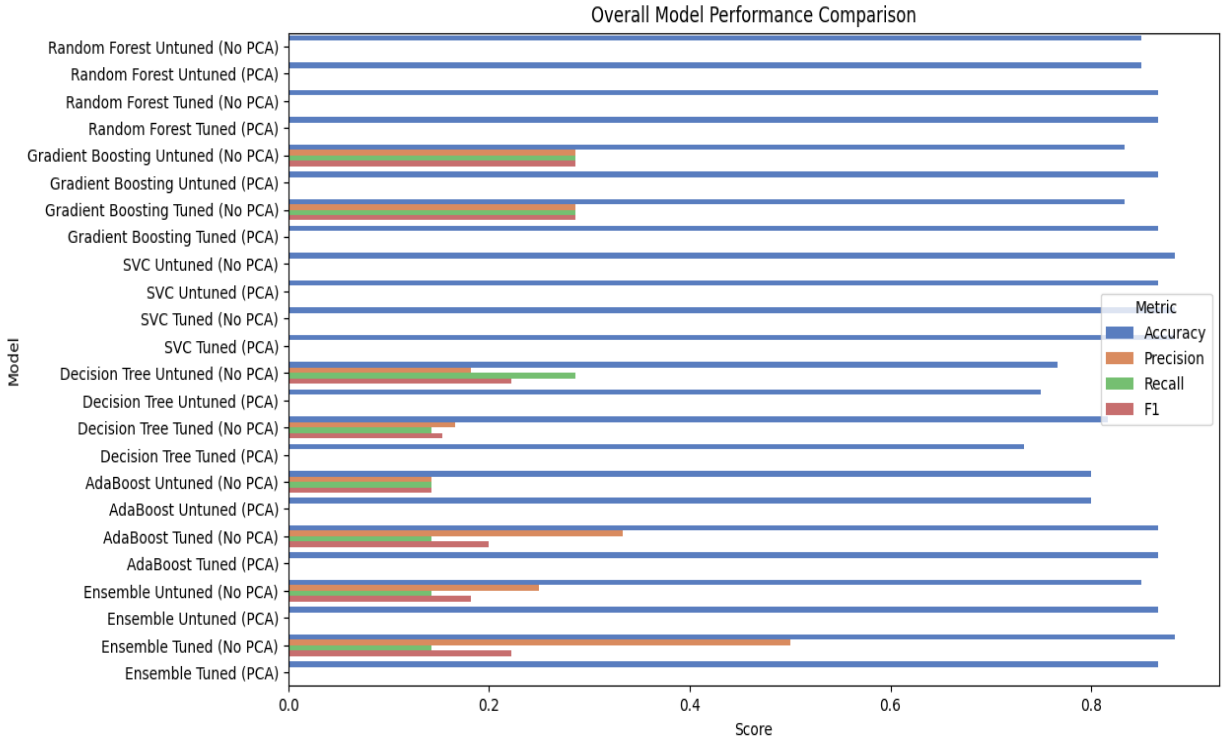


Fig 2 : Overall Model Performance Comparison

2.3.1. Confusion Matrix

The confusion matrix, which is entirely dependent on the True, False, Positive, and Negative outputs, shows the quantity of accurate and inaccurate predictions as well as the kinds of mistakes that classifiers make [42].

2.3.2. Accuracy

The percentage of accuracy is a measure of how well the linked classifier predicts. TP + TN / TP + TN + FP + FN = Accuracy.

2.3.3. Recall

Recall is the quantity of positive cases that the classifier correctly predicted. It also makes reference to the sensitivity. Memory is equal to TP / TP + FN.

2.3.4. Precision

Precision is also known as positive predictive value. It displays the percentage of favorable outcomes that are truly favorable. Accuracy = TP / TP + FP.

2.3.5. F1-score

The F1-score combines recall and precision to determine the model's accuracy. Recall ∗ precision / precision + recall = F1.

2.5.3. ROC

The ROC curve is a graphical tool for evaluating binary classifier performance. It illustrates how a model's true positive rate and false positive rate are traded off at different classification criteria. [22].

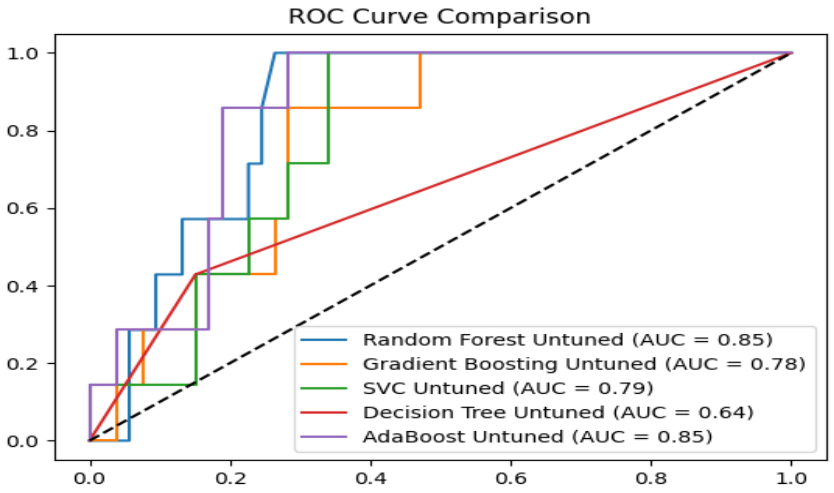


Fig 3 : ROC Curve Comparison

3. RESULTS

920 samples and 14 features from four databases—Cleveland (304), Hungarian (293), Switzerland (123), and Long Beach, Virginia (200)—make up the equivalent dataset [20] used in this investigation. Furthermore, 30% of all samples were used as test samples to predict cardiac disease. Table 1 compiles the relevant attributes together with their definition. Fig. 2 shows how each attribute is distributed in respect to other attributes. Additionally, two characteristics need to be changed. Initially, there were five values for ca: 0, 1, 2, 3, and 4. They are taken out and replaced with value = 3 since value = 4 is out of range. Second, the incorrect value of 0 is regarded as the slope column's missing value. As a result, the SimpleImputer algorithm is used to fill it utilizing the median calculation strategy. The median of the slope attribute then reaches value = 1. The matching x is standardized using StandardScaler once the target column for creating x data, y data, is dropped. To present a wonderful forecast, this study applies six various classification algorithms, including RF, GB, SVC, DT, AB, and Ensemble Classifier, together with three different dimension reductions, including PCA and GridSearchCV.

3.1. Feature reduction results

The pipeline method is used in this work with two transformers, namely classifiers and PCA. For PCA, eight numbers are taken into account. Table 2 shows the values of four assessment measures for six classifiers. Each assessment metric shows the best values that can be achieved with the specific number of PCA and parameter for each classifier. Out of all the classifiers, DT and RF perform the best based on the available data.

3.2. GridSearchCV results

As can be seen in Table 3, DT and KNN obtained the same and best results among classifiers using GridSearchCV. Those two have somewhat different RF values. DT has no parameter in this method.

3.3. Confusion matrix

Figures two and three show the confusion matrices of six classification strategies for each dimension reduction method and GridSearchCV. There are 308 samples in the related test data. [41] The numbers of disease (N) and no disease (P) are 149 and 159, respectively.

4. DISCUSSION

To examine the performance of six classifiers, two distinct methods were used at the preprocessing stage. PCA is one kind of dimension reduction technique, based on the information provided in Table 2. Additionally, the accuracy of classifiers is entirely increased by the GridSearchCV hyper-parameter tuner technique. Furthermore, the best performance is shown by DT and RF.

It is evident that a ROC of greater than 0.8 indicates that the classification algorithms are doing exceptionally well. Fig. 3 displays the ROC RF, GB, SVC, DT, and AB for this investigation. The ROC of RF, AB increased greatly from 0.85 to 0.85 via GridSearchCV, as is evident. Six classification algorithms under the dimension reduction and hyper-parameter tuning method are used to categorize the heart disease dataset. PCA is regarded as eight dimensions, one dimension, and

© 2025 Journal of Carcinogenesis | Published for Carcinogenesis Press by Wolters Kluwer-Medknow

one to three dimensions, respectively, when it comes to dimension reduction. Additionally, n_estimators = 7 for RF, kernel = "rbf" for SVC, and solver = "liblinear" for LR are the parameters of the various classifiers with PCA.

The "linear" parameter is substituted for the "rbf" parameter in the LDA method to enhance SVC performance. "C" = 100 and "gamma" = 0.1 for SVC, "C" = 0.1 and "penalty" = 12 for LR, "var_smoothing" = 0.351 for Gaussian, and n_estimators = 30 for RF and n_neighbors = 1 for KNN are the parameters of six classifiers with GridSearchCV when hyper-parameter tweaking is used. NB. SimpleImputer, StandardScaler, the train_test_split approach, dimension reduction techniques, GridSearchCV, classifiers, and metrics are all imported using Scikit-learn, an open-source Python package [48].

As is evident, each study may use a different type of dataset related to heart disease. In the meantime, five investigations that did not use the GridSearchCV technique examined 60 samples and 14 attributes [16,19,23,28]. Furthermore, neither GridSearchCV nor the dimension reduction technique were used in the preparation of the data. GridSearchCV was used in [24], however the outcomes were not as strong as those of the approach that was described, which used three classifiers. Several classification algorithms, including five classifiers shared by the suggested strategy, were used to classify a dataset consisting of 4728 samples and 14 characteristics.

In order to classify using six distinct classification methods, the method ultimately applied all of the specified preprocessing techniques, such as GridSearchCV and three different forms of dimension reductions, to a merged dataset. By identifying the optimal parameter and reducing the dataset size, classification greatly enhanced the outcomes. In the meantime, RF, GB, SVC, DT, AB, and Ensemble Classifier achieved, respectively, 85%, 81%, 78%, 80%, 76%, and 80% accuracy.

This manuscript makes use of publically available datasets, including four different datasets from Kaggle with 920 samples and 14 characteristics. To improve data preparation for categorization, a variety of data pretreatment techniques are also used for sample analysis. As a result, noteworthy outcomes are anticipated. To accomplish a notably high level of accuracy with DT and RF classifiers, the related dataset is short and carefully managed. The results obtained were really 85% rather than 100%. In fact, none of the classifiers achieved 100% accuracy. To find the best classification performance in this study, A range of categorization algorithms are also employed.

5. CONCLUSION

Six popular classifiers—RF, GB, SVC, DT, AB, and Ensemble Classifier—as well as data pretreatment approaches, such as three different kinds of dimension reduction techniques and hyper-parameter tuning, are employed in this study to effectively predict the risk of CVDs. With this approach, a variety of metrics, including accuracy, recall, precision, F1-score, confusion matrix, and ROC curve, were used to assess the classifiers' performance. Preprocessing techniques had a minor impact on heart disease prediction, according to a number of recent research [16–29], but PCA's hyper-parameter tuning strategy enhanced heart disease prediction.

In subsequent research, we will use deep learning algorithms in conjunction with hyper-parameter tuning and dimension reduction techniques to confirm their impact on the prediction of heart disease. To increase the classifier's performance, a variety of preprocessing strategies were used, such as various dimension reduction techniques and hyper-parameter tuning methods, and their effectiveness was simultaneously examined. Every effort was made in this investigation to obtain the most likely classification results. It is evident that six traditional classifiers achieved excellent results by employing a variety of suitable preprocessing techniques. How various preprocessing techniques and hyperparameter tuning might impact the end findings is crucial to achieving the best outcome in this study.

REFERENCES

[1] Barbara Martins, Diana Ferreira, Cristiana Neto, Antonio Abelha, Jose Machado,“Data Mining for Cardiovascular Disease Prediction,” Journal of Medical Systems (2021) 45: 6, https://doi.org/10.1007/s10916-020-01682-8© Springer, 5 January 2021.

[2] Najmul Hasan and Yukun Bao, “Comparing different feature selection algorithms for cardiovascular disease prediction,” Health and Technology https://doi.org/10.1007/s12553-020-00499-2 © Springer, October 2020.

[3] Amin Ul Haq ,Jian Ping Li , Muhammad Hammad Memon , Shah Nazir, and Ruinan Sun, “A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms,” Hindawi Mobile Information Systems Volume 2018, Article ID 3860146, 21 pages https://doi.org/10.1155/2018/3860146.

[4] Ahmed I. Taloba , Rayan Alanazi, Osama R. Shahin , Ahmed Elhadad , Amr Abozeid , and Rasha M. Abd El-Aziz, “Machine Algorithm for Heartbeat Monitoring and Arrhythmia Detection Based on ECG Systems,” Hindawi Computational Intelligence and Neuroscience Volume 2021, Article ID 7677568, 9 pages https://doi.org/10.1155/2021/7677568.

[5] Karl-Patrik Kresoja, Matthias Unterhuber, Rolf Wachter,Holger Thiele,Philipp Lurz, "A cardiologist's guide to machine learning in cardiovascular disease prognosis prediction" Basic Research in Cardiology (2023) 118:10 https://doi.org/10.1007/s00395-023-00982-7 © Springer, 20 March 2023.

[6] Sibo Prasad Patro, Neelamadhab Padhy, Dukuru Chiranjevi, "Ambient assisted living predictive model for cardiovascular disease prediction using supervised learning" Evolutionary Intelligence https://doi.org/10.1007/s12065-020-00484-8, © Springer-Verlag GmbH Germany, part of Springer Nature 2020.

[7] Saravanan Srinivasan , Subathra Gunasekaran , Sandeep Kumar Mathivanan , Benjula Anbu Malar M. B , Prabhu Jayagopal & Gemmachis Teshite Dalu, "An active learning machine technique based prediction of cardiovascular heart disease from UCI-repository database" Scientific Reports | (2023) 13:13588 |https://doi.org/10.1038/s41598-023-40717-1.

[8] Ahmad Shaker Abdalrada, Jemal Abawajy, Tahsien Al-Quraishi, Sheikh Mohammed Shariful Islam, "Machine learning models for prediction of co- occurrence of diabetes and cardiovascular diseases: a retrospective cohort study" Journal of Diabetes & Metabolic Disorders (2022) 21:251–261,https://doi.org/10.1007/s40200-021-00968-z, © Springer, 12 January 2022.

[9] Kazi Mostaq hridoy, Saiful Islam Akash, Fahmida Afrose Dipti, Mahamudul Hasan, Jesan Ahammed ovi, Md Al-Imran, Dr. Sarwar Jahan " Heart Disease Algorithm using Machine Learning Algorithms" 2023 4th International Conference on Big Data Analytics and Practices (IBDAP) | 979-8-3503-0019-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/IBDAP58581.2023.10271997.

[10] Niloy Biswas , Md Mamun Ali ,Md Abdur Rahaman,Minhajul Islam ,Md. Rajib Mia ,Sami Azam ,Kawsar Ahmed ,Francis M. Bui , Fahad Ahmed Al-Zahrani , and Mohammad Ali Moni , "Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques" Hindawi BioMed Research International Volume 2023, Article ID 6864343, https://doi.org/10.1155/2023/6864343.

[11] Madhukar Chavan, Shashank Kumar Singh, Saurabh Bansod and Prashant Pal, "Design and Implementation of Heart Disease Prediction Using Artificial Neural Network" 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM) | 979-8-3503-4779-1/23/$31.00 ©2023 IEEE| DOI:10.1109/ICONSTEM56934.2023.10142267

[12] M. Durairaj and Nandhakumar Ramasamy, "Intelligent Prediction Methods and Techniques Using Disease Diagnosis in Medical Database: A Review," 2015, pp. 2153-2160 © International Science Press.

[13] Sana S. Zadawale1 , Savita Bakare, KLE Dr MSSCET, "ECG Signal Based Heart Disease Prediction System using DWT and SVM", IJARCCE Vol. 6, Issue 7, July 2017.

[14] Steven N. Blair1, "An Cardio respiratory Fitness: Overview of Non-exercise Estimated Estimation Equations, Cross-Validation and Application," Journal of Science in Sport and Exercise (2019) 7 June2019 © Beijing Sport University 2019.

[15] A.Saravana Kumar and Ponnaiyah Ramajayam, "Heart Disease Detection Using Radial Basis Function Classifier" ICTACT, September 2020, Vol: 01, Issue: 04.

[16] Maryam I. Al-Janabi1 , Mahmoud H. Qutqut1, Mohammad , "Machine Learning Classification Techniques for Heart Disease Prediction: A Review," IJET, Queen's University, Kingston,Canada.

[17] Tabreer T. Hasan, Manal H. Jasim and Ivan A. Hashim, "Heart Disease Diagnosis System based on Multi-Layer Perceptron neural network and Support Vector Machine ,"IJCET, Vol.7, No.5 (Sept/Oct2017).

[18] Moloud Abdar, Wojciech Ksiazek, U.Rajendra Acharya, Ru-San Tan, Vladimir Makarenkov, Paweł. Pławiak, A new machine learning technique for an accurate diagnosis of coronary artery disease, Comput. Methods Prog. Biomed. 179 (2019), https://doi.org/10.1016/j.cmpb.2019.104992.

[19] James Meng MA, MB, BChir, Ruiming Xing MSc, Inside the "black box": Embedding clinical knowledge in data-driven machine learning for heart disease diagnosis, (Cardiovasc. Digit. Health J. (2022), https://doi.org/10.1016/j. cvdhj.2022.10.005.

[20] Khaled Mohamad Almustafa, Prediction of heart disease and classifiers' sensitivity analysis, "Almus BMC Bioinforma. (2020), https://doi.org/10.1186/s12859-020- 03626-y.

[21] K Rohit Chowdary , P Bhargav , N Nikhil , K Varun and Dr D Jayanthi, "Early heart disease prediction using ensemble learning techniques," Journal of Physics: Conference Series 2022.

[22] M. Sheetal Singh, Khelchandra Thongam , Prakash Choudhary , and P. K. Bhagat , "An Integrated Machine Learning Approach for Congestive Heart Failure Prediction", https://doi.org/ 10.3390/diagnostics14070736, March 2024.

[23] Khandaker Mohammad Mohi Uddin, Rokaiya Ripa, Nilufar Yeasmin, Nitish Biswas, Samrat Kumar Dey, Machine learning-based approach to the diagnosis of cardiovascular vascular disease using a combined dataset, Intell. -Based Med. 7(2023), https://doi.org/10.1016/j.ibmed.2023.100100.

[24] Ghulab Nabi Ahamad, Shafiullah, Hira Fatima, Imdadullah, S. M. Zakariya , Mohamed Abbas , Mohammed S. Alqahtani and Mohammed Usman, "Influence of Optimal Hyperparameters on the Performance of Machine Learning Algorithms for Predicting Heart Disease" Processes 2023, 11, 734https://doi.org/10.3390/pr11030734.

[25] Prabhakaran D, Jeemon P, Sharma M, et al. “The changing patterns of cardiovascular diseases and their risk factors in the states of India: the Global Burden of Disease Study” 1990e2016. Lancet Glob Health. 2018. https://doi.org/10.1016/s2214-109x(18)30407-8.

[26] Amirhossein Najafi, Alireza Nemati, Mahdi Ashrafzadeh, Sarfaraz Hashemkhani Zolfani, Multiple-criteria decision making, feature selection, and deep learning: a golden triangle for heart disease identification, Eng. Appl. Artif. Intell. 125 (2023), https://doi.org/10.1016/j.engappai.2023.106662.

[27] Van der Heijden AA, Abramoff MD, Verbraak F, van Hecke MV, Liem A, Nijpels G. “Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System.”ActaOphthalmol.2017;96(1):63e68.https://doi.org/10.1111 /aos.13613.

[28] Maini E, Venkateswarlu B, Gupta A. Applying machine learning algorithms to develop a universal cardiovascular disease prediction system. In: International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018. 2018:627e632.

[29] Ahmad, G.N.; Ullah, S.; Algethami, A.; Fatima, H.; Akhter, S.M.H. “Comparative Study of Optimum Medical Diagnosis of Human Heart Disease Using ML Technique with and without Sequential Feature Selection.” IEEE Access 2022, 10, 23808–23828.

[30] Sushant K. Singh, Robert W. Taylor, Biswajeet Pradhan, Ataollah Shirzadi, Predicting sustainable arsenic mitigation using machine learning tecgniques, Ecotoxicol. Environ. Saf. 232 (2022), https://doi.org/10.1016/j.ecoenv.2022.113271.

[31] Mc.Kinney Wes, Python for data analysis: Data wrangling with Pandas. NumPy, and IPython", O’Reilly Media, Inc, 2017.

[32] Ayat Naji Hussain, Sahar Adil Abboud, Basim Abdul baki Jumaa, Mohammed Najm Abdullah, Impact of feature reduction techniques on classification accuracy of machine learning techniques in leg rehabilitation, Meas.: Sens. 25 (2023), https://doi.org/10.1016/j.measen.2022.100544.

[33] Bahzad Taha Jijo, Adnan Mohsin Abdulazeez, Classification based on decision tree algorithm for machine learning, J. Appl. Sci. Technol. Trends 02 (01) (2021) 20–28.

[34] Sunwoo Han, Hyunjoong Kim, Yung-Seop Lee, Double random forest (no page), Mach. Learn. 109 (2020) 1569–1586, https://doi.org/10.1007/s10994-020- 05889-1.

[35] Mienye, I.D.; Sun, Y.;Wang, Z. “An improved ensemble learning approach for the prediction of heart disease risk.” Inform. Med. Unlocked 2020, 20, 100402.

[36] Eric Noboa, Melinda Racz, Laszlo Szucs, Peter Galambos, Gergely Marton, Gyorgy Eigner, Development of an EMG based SVM supported control solution for the PlatypOUs education mobile robot using MindRove headset, 54-15, IFAC Pap. (2021), 〈https://creativecommons.org/licenses/by-nc-nd/4.0/〉, 54-15.

[37] Magesh, G.; Swarnalatha, P. “Optimal feature selection through a cluster-based DT learning (CDTL) in heart disease prediction”.Evol. Intell. 2021, 14, 583–593.

[38] Bharti, R.; Khamparia, A.; Shabaz, M.; Dhiman, G.; Pande, S.; Singh, P. “Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning.” Comput. Intell. Neurosci.2021, 2021, 8387680.

[39] Md. Faisal Kabir, Tianjie Chen, Simone A. Ludwig, A performance analysis of dimensionality reduction algorithms in machine learning models for cancer prediction, Healthc. Anal. 3 (2023), https://doi.org/10.1016/j.health.2022.100125.

[40] Lan T. Jolliffe, Jorge Cadima, Principal component analysis: a review and recent developments, R. Soc. – Philos. TRANACTIONS A (2015). 〈http://dx.org/10.109 8/rsta.2015.0202〉.

[41] Ayat Naji Hussain, Sahar Adil Abboud, Basim Abdul baki Jumaa, Mohammed Najm Abdullah, Impact of feature reduction techniques on classification accuracy of machine learning techniques in leg rehabilitation, Meas.: Sens. 25 (2023), https://doi.org/10.1016/j.measen.2022.100544.

[42] Rui Silva, Pedro Melo-Pinto, t-SNE: a study on reducing the dimensionality of hyperspectral data for the regression problem of estimating oenological parameters, Artif. Intell. Agric. 7 (2023), https://doi.org/10.1016/j.aiia.2023.02.003.

© 2025 Journal of Carcinogenesis | Published for Carcinogenesis Press by Wolters Kluwer-Medknow