مجلة جامعة بني وليد للعلوم الإنسانية والتطبيقية

تصدر عن جامعة بني وليد – ليبيا

Website: https://jhas-bwu.com/index.php/bwjhas/index

العدد التاسع والعشرون، 2023

Comparison of Ensemble learning algorithms in predicting heart disease

Bader N. Awedat [1*], Ali M. Abumrfgh[2]

[1,2] computer science, faculty of Information Technology / Azzaytuna University, Libya

*Corresponding author: bader_najep@yahoo.com

تاريخ الاستلام: 20-06-2023 تاريخ القبول: 07-07-2023 تاريخ النشر: 07-09-2023

Abstract: The main objective of this research is enhancing accuracy of predictive analysis for cardiovascular diseases (CVDs) through the implementation of ensemble learning algorithms. Ensemble learning is a strong approach that combines predictions from multiple models to amelioration overall performance. In this research, we compare the effectiveness of three ensemble learning algorithms: Random Forest, AdaBoost, and Stacking. We evaluate their performance using five criteria: Recall, Precision, F-score, Roc Auc, and Accuracy. The obtained results indicate that the AdaBoost algorithm has achieved the highest performance in the field of diagnosis using the available data. This signifies the high effectiveness of this algorithm in disease prediction and diagnosis. It is also notable that the Stacking algorithm has demonstrated strong performance, particularly in comparison to the Random Forest algorithm. Other performance standards such as Accuracy, Recall, Cohen's kappa, F-measure, Precision, and Specificity also exhibit good performance for the different algorithms. The ROC Curve metric reveals that the AdaBoost algorithm has attained the highest value (97.64), indicating its capability to effectively discriminate between true and false instances.

Keywords: Bagging, Boosting, Ensemble learning, ROC curve, Stacking.

Introduction

Cardiovascular diseases (CVDs) are a significant global cause of mortality, accounting for approximately 17.9 million deaths annually, representing 31% of all global deaths. The majority of these deaths result from heart attacks and strokes, with a significant proportion occurring prematurely in individuals under the age of 70. This presents a major public health concern that necessitates attention and effective preventive measures. CVDs frequently lead to heart failure, and the provided dataset contains 13 features that can be utilized to predict the possibility of heart disease. [1]

Heart failure is commonly observed as a result of cardiovascular diseases. This dataset comprises 13 features that aid in predicting the likelihood of heart disease. Early detection and effective management are essential for individuals with cardiovascular disease or those at a high risk of developing it, given the presence of risk factors like hypertension, diabetes, hyperlipidemia, or pre-existing conditions. Machine learning models can significantly contribute to addressing this issue.

Ensemble learning improves performance by creating and combining multiple distinct base learners using dedicated approaches. These individual models are commonly referred to as base learners, while the process of combining them is known as integration strategy. In the context of combining multiple models to improve predictive performance, ensemble learning techniques are employed. In this approach, the individual performance of each base learner doesn't need to be exceptionally strong; rather, it should surpass random guessing. Based on how the base learners are generated, ensemble learning methods can be broadly classified into two categories: 1. Parallel methods, such as Bagging, and 2. Sequential methods, such as Boosting. [2]

The concept of "wisdom of the crowd" involves merging multiple weak models or learners into a single predictive model. This methodology aims to address bias, reduce variance, and enhance accuracy by leveraging the collective knowledge and predictions of the individual models.

The aim of this research is to improve the accuracy of predictive analysis for heart disease. While individual machine learning models may have limitations in their prediction capabilities, ensemble

learning methods, such as collaborative learning, strive to address this issue by training multiple models successively to enhance the overall accuracy of the system. Additionally, ensemble learning approaches are well-suited for datasets of different sizes, making them effective in data mining tasks. These methods have demonstrated promising performance in tackling the challenges of variance and bias that are commonly encountered in data mining algorithms. In our study, we specifically concentrate on the complex task of heart disease prediction and diagnosis, which involves understanding various contributing factors.

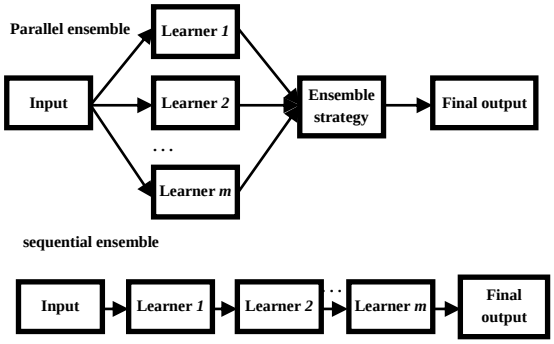

Fig. 1: Flowchart of Parallel and Sequential ensemble.[2]

Related Work

A concise overview of recent research papers comparing various machine learning algorithms and examining their outcomes.

In the study conducted by Madhumita Pal and Smita Parija [3], the random forest algorithm was employed to predict heart disease. The researchers utilized a dataset comprising 13 features for their analysis. The obtained results revealed an accuracy of 86.9%, a sensitivity value of 90.6%, and a specificity value of 82.7%. The receiver operating characteristics analysis demonstrated a diagnosis rate of 93.3% for heart disease prediction using the random forest algorithm. These findings highlight the high effectiveness of the random forest algorithm in accurately classifying heart disease [3].

In a recent study conducted by Asfandyar Khan et al. [4], a novel ensemble approach known as the 'Stacking Classifier' was proposed to enhance the performance of integrated individual classifiers and reduce the likelihood of misclassification for individual instances. The Stacking Classifier utilizes Random Forest and SVM as meta-classifiers. The experimental results indicated that the proposed stacking classifier achieved a remarkable accuracy of 0.9735 percent in diagnosing diabetes, surpassing the performance of existing models such as Naive Bayes (0.7646 percent), KNN (0.7460 percent), DT (0.7857 percent), and LDA (0.7735 percent). Similarly, in the context of cardiovascular disease, the suggested stacking classifier demonstrated superior performance compared to current models, including KNN (0.8377 percent), NB (0.8256 percent), DT (0.8426 percent), LDA (0.8523 percent), and SVM (0.8472 percent). The stacking classifier achieved a higher accuracy of 0.8871 percent [4].

T. R. Mahesh, et al [5], the researcher used the synthetic minority over-sampling technique (SMOTE). The results of the study indicate that the AdaBoost-Random Forest classifier achieves a high accuracy of 95.47% in the early detection of heart disease.

Md. Maidul Islam, et al [6], In this research, 9 algorithms were compared, including 5 Ensemble Learning algorithms, and gave the best results, the best of which was the stack algorithm.

The Stacked Ensemble Classifier showcases outstanding performance, achieving an accuracy of 0.910, sensitivity of 0.934, specificity of 0.883, best F1-score of 0.916, minimum Log Loss of 3.08, and the highest ROC value of 0.909 [6].

Among the various evaluated metrics, Random Forest demonstrates the highest sensitivity, followed by XGBoost [6]. The Stacked Classifier model attains an accuracy of 91.06%, along with an F1 score of 0.9163. In comparison, the XGBoost and Random Forest algorithms achieve accuracies of 89.78% and 89.36%, respectively, with corresponding F1 scores of 0.8972 and 0.8911. The Extra Tree Classifiers, CART, GBM, MLP, SVC, and KNN algorithms exhibit accuracies of 88.51%, 85.10%, 82.97%, 82.12%, 81.27%, and 80.00%, respectively [6].

Types of Ensemble Learning:

1.Bagging:

In the ensemble learning approach called "majority voting," multiple weak models (N in total) are trained in parallel using non-overlapping subsets of the input dataset. In the testing phase, each model is assessed individually, and the label that receives the highest number of predictions is chosen as the final prediction. This approach aims to merge the predictions from multiple models to achieve a more robust and accurate prediction.

2.Boosting:

Boosting is a machine learning technique that trains N different weak models sequentially on the entire dataset. These weak models are typically of the same type (homogeneous). In each iteration, data points that were misclassified by the previous weak model are assigned higher weights to prioritize their correct classification by the subsequent weak learner. During the testing phase, the predictions of each model are combined by assigning weights based on their test error, enabling a voting mechanism. Boosting methods are known to effectively reduce prediction bias.

3.Stacking:

In the ensemble learning approach, multiple weak models (N in total) are trained simultaneously, often of different types (heterogeneous). This training process is performed using one subset of the dataset. Once the weak models are trained, a meta learner is trained using their predictions to perform the final prediction. The meta learner utilizes the other subset of the dataset. During the testing phase, each individual model predicts its label, and these predicted labels are combined and provided to the meta learner, which generates the ultimate prediction.

Materials and Methods:

We employed group learning algorithms to predict heart failure diseases, including the utilization of the random forest algorithm and AdaBoost. Additionally, we incorporated four stacked algorithms for enhanced performance. The applied algorithms in this study are based on the analysis of a Heart Failure dataset obtained from the Kaggle repository. The dataset comprises 304 patient samples, each representing medical records. The heart failure dataset includes a wide range of features. To enhance the algorithm performance, a comprehensive analysis of these features is conducted, considering factors such as importance scores, accuracy, sensitivity, and specificity. The dataset was split into a training set comprising 70% of the data and a testing set comprising the remaining 30%.

In this study, Google Colab Notebook was utilized as the simulation tool for conducting the experiments and constructing the models. Google Colab Notebook is a convenient platform for Python programming projects, offering a wide range of features such as rich text components, code integration, and real-time data analysis capabilities. It enables the seamless integration of descriptive analysis, findings, equations, and visualizations. Similar to Jupyter Notebook, Google Colab Notebook provides a web-based interactive interface for creating and sharing interactive graphics, maps, plots, visualizations, and narrative texts. This tool proved to be invaluable in facilitating the analysis process and enhancing the efficiency of the research workflow. Moreover, Google Colab Notebook is an open-source tool, making it accessible and freely available for researchers.

Five features were used out of a total of 13 features in our study. These features are as follows: Age, resting blood pressure (trestbps), Cholesterol level (chol), Maximum heart rate achieved (thalach, ST depression induced by exercise relative to rest (oldpeak).

In addition, the target feature was utilized to determine the presence (1) or absence (0) of heart disease."

Ensemble techniques:

Ensemble techniques refer to methods that employ multiple learning algorithms or models to create an optimal predictive model. The resulting model exhibits superior performance compared to the individual base learners used independently. Ensemble learning has various other applications, such as feature selection and data fusion, among others. Additionally, ensemble techniques can be categorized into three main types: Bagging, Boosting, and Stacking.

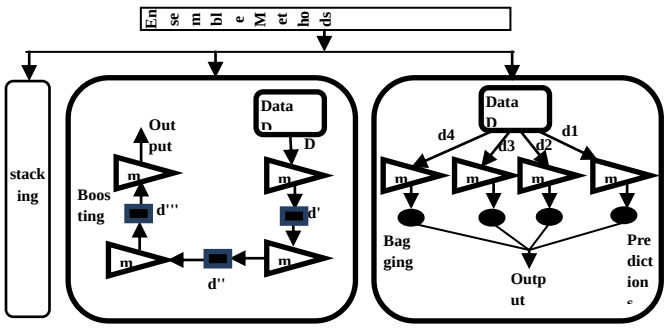

Fig. 2: Representing the types of ensemble methods.[7]

1.Bagging (Bootstrap Aggregating):

Bagging is a short form of bootstrap aggregating. It is an ensemble technique that divides a dataset into n samples with replacement. The dataset is divided into n samples, and each sample is trained individually using separate machine learning models. Then the output of all the separate models is combined into one single output by using voting (Figure 3). [8]

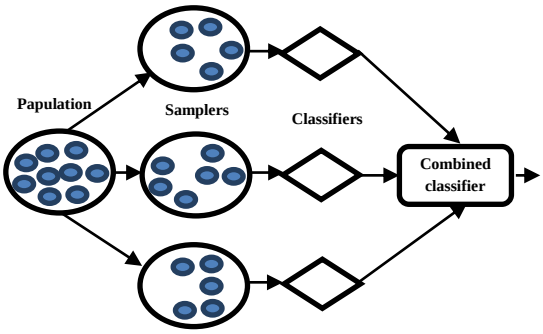

Fig. 3: Bagging with sampling [8]

Random Forest (RF) is a powerful machine learning technique that employs an ensemble of decision trees to tackle classification and regression problems. It combines multiple decision trees using bootstrap sampling, and the final classification or regression outcome is determined through majority voting or averaging [9][10]. RF is widely recognized for its robustness in handling imbalanced, missing, and multicollinear data [2][9]. The analysis process consists of two stages:

Stage 1: The random forest is constructed by randomly selecting samples with replacement from the initial dataset (training data). Subsets are created, and regression trees are built based on these smaller datasets. During the training stage, various parameters can be adjusted, including the number of variables (mtry) and the number of trees (ntree).

Stage 2: Once the random forest model is trained, predictions can be generated. The input variables for each regression tree are combined, and the final prediction is obtained by averaging the predictions from all the trees [10].

Majority Voting, also known as Hard Voting, is a classification technique in which the predicted class label ŷ is determined by taking the majority vote of multiple classifiers C_j. In other words, the class label that is predicted by the majority of the classifiers is selected as the final prediction:

$\hat{y} = mode\ \{C_1(x), C_2(x), \ldots., C_m(x)\}$ [11]

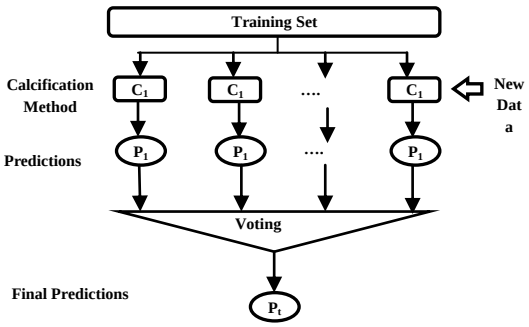

Fig. 4: Random Force Algorithm. [12]

2.Boosting:

There are primarily three types of boosting algorithms commonly used in Machine Learning:

- AdaBoost algorithm,
- Gradient descent algorithm,
- Xtreme gradient descent algorithm [7].

In 1995, Freund and Schapire introduced the Adaboost (adaptive boosting) algorithm. This algorithm operates by adjusting weights without requiring any prior knowledge of the learner's training process [13].

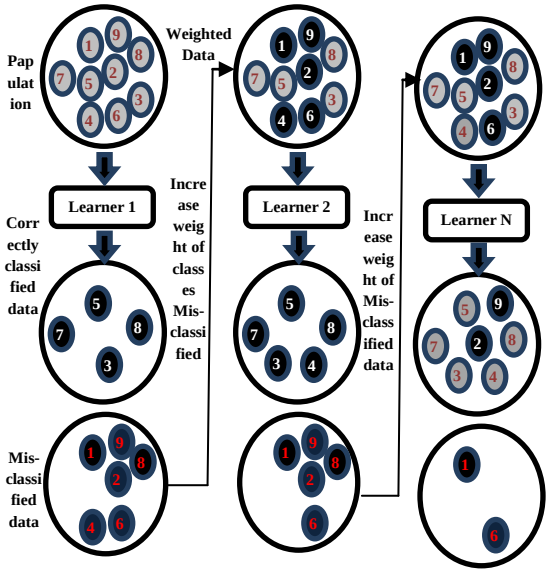

Fig. 5: AdaBoost boosting. [8]

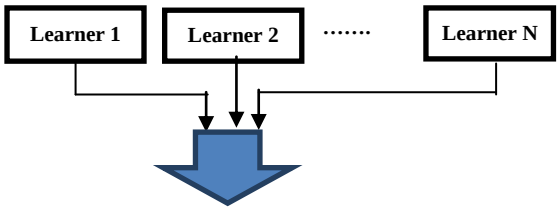

Fig. 6: Voting of n learners in AdaBoost boosting.[8]

Basic Algorithm for Boosting:

1. **Initialize:** set all examples to have equal weights
2. **For each** t = 1, …, T,
3. Learn a hypothesis h_t from weighted examples
4. **Decrease weights** of examples h_t classifies **correctly**
5. Calculateα_t, the weight of the current weak learner, h_t
6. $\boldsymbol{Return}\ h(x) = \sum_{t=1}^{T} \alpha_t\ h_t(x)$

3.Stacking:

Stacking is a powerful technique for combining predictions in ensemble learning. It involves training multiple models, referred to as base learners, to generate individual predictions. These predictions are then used as input for another model, known as the meta learner or aggregator, which learns to combine the base learners' predictions (Figure 7). Think of stacking as building a stacked architecture of machine learning models, where each layer learns to aggregate the predictions of the previous layer. Unlike traditional ensemble methods that use simple functions like majority voting to aggregate predictions, stacking leverages a model to perform this aggregation, resulting in improved performance and flexibility.

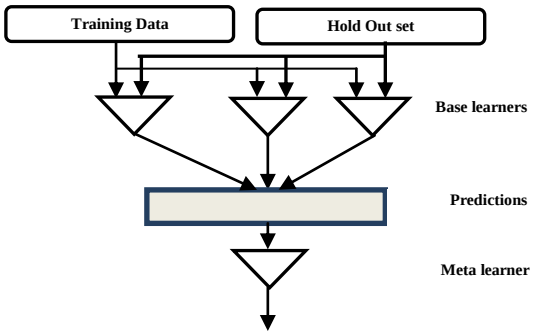

Fig. 7: Stacking. [8]

Stacking Algorithms: Two algorithms, xgboost and random forest, were used as stack algorithms and Logistic Regression was used as mate algorithm.

Dataset Description

The Heart Failure Dataset utilized in this study is sourced from the Kaggle platform [14]. The dataset used in this study is a combination of five distinct datasets, resulting in a comprehensive and diverse collection of attributes. Specifically, we focused on employing attributes that are highly relevant in predicting a patient's heart condition for this specific experiment. Furthermore, the dataset file comprises 13 medical variables for a total of 304 patients. A comprehensive description of each attribute, along with the respective value count, is presented in Table 1. The exploratory data analysis is further depicted in Table 2, accompanied by a figure illustrating the features with less than 5 columns, and another figure displaying the features with more than 5 columns. The heatmap representation can be observed in Figure 2.

Table 1. Description of Features

N.	Features	Description
1	Age	"The age of the individual, expressed in years". (Source: Ramadan A.M. Elghalid et al., 2022)
2	Sex	"The gender of the person, represented as a binary variable where 1 indicates male and 0 indicates female". (Source: Ramadan A.M. "Elghalid et al., 2022)
3	chest_pain_type	"0: asymptomatic 1: atypical angina 2: non-angina pain 3: typical angina" (Source: Ramadan A.M. Elghalid et al., 2022)
4	RestingBP	"The person's resting blood pressure upon admission to the hospital,

		measured in millimeters of mercury (mm Hg)". (Source: Ramadan A.M. Elghalid et al., 2022)
5	Cholesterol	"The measurement of the person's cholesterol level, expressed in milligrams per deciliter (mg/dL)". (Source: Ramadan A.M. Elghalid et al., 2022)
6	FastingBS	"Indicates whether the person has a fasting blood sugar level higher than 120 mg/dL. It is represented as a binary variable, where 1 indicates true (blood sugar > 120 mg/dL) and 0 indicates false (blood sugar ≤ 120 mg/dL)". (Source: Ramadan A.M. "Elghalid et al., 2022)
7	RestingECG	Resting Electrocardiographic Results: 0: Indicates probable or definite left ventricular hypertrophy based on Estes' criteria. 1: Represents a normal resting electrocardiogram. 2: Indicates the presence of ST-T wave abnormalities, such as T wave inversions and/or ST elevation or depression exceeding 0.05 mV". (Source: Ramadan A.M. "Elghalid et al., 2022)
8	MaxHR	The maximum heart rate achieved by an individual". (Source: Ramadan A.M. Elghalid et al., 2022)
9	ExerciseAngina	"The presence of exercise-induced angina, represented by a value of 1 for 'yes' and 0 for 'no". (Source: Ramadan A.M. Elghalid et al., 2022)
10	Oldpeak	"The magnitude of ST depression induced by exercise relative to rest. Note that 'ST' refers to specific positions on the ECG plot. For more information, please refer to the provided resource". (Source: Ramadan A.M. "Elghalid et al., 2022)
11	ST_Slope	The slope of the peak exercise ST segment: 0: Represents a downsloping ST segment. 1: Indicates a flat ST segment. 2: Indicates an upsloping ST segment". (Source: Ramadan A.M. "Elghalid et al., 2022)
12	Heart Disease	The Target (1 = no, 0= yes)". (Source: Ramadan A.M. Elghalid et al., 2022)

Table 2. Exploratory data analysis

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

303 rows × 14 columns

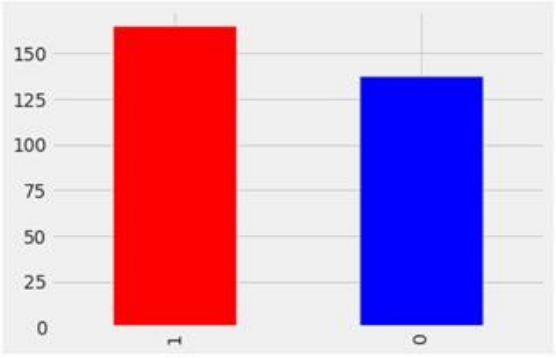

Fig. 8: Percentage of Heart Disease

We have 165 persons with heart disease and 138 persons without heart disease, so our problem is balanced. [Kaggle Inc]

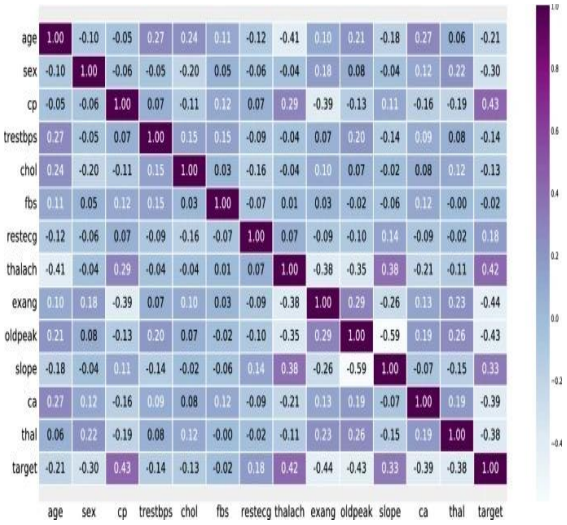

Fig. 9: Heatmap depiction of the dataset.

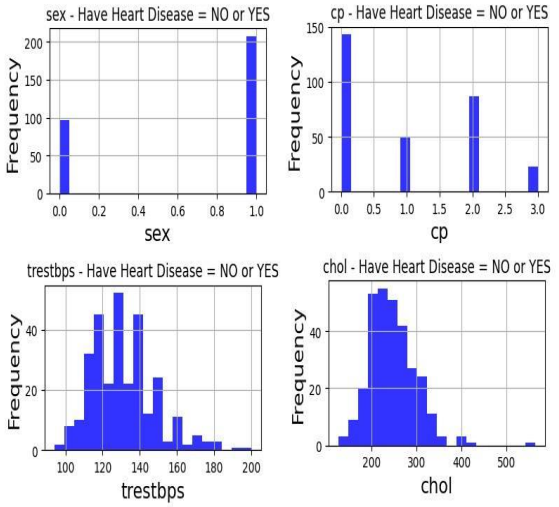

Fig. 10: Properties with less than 5 variables.

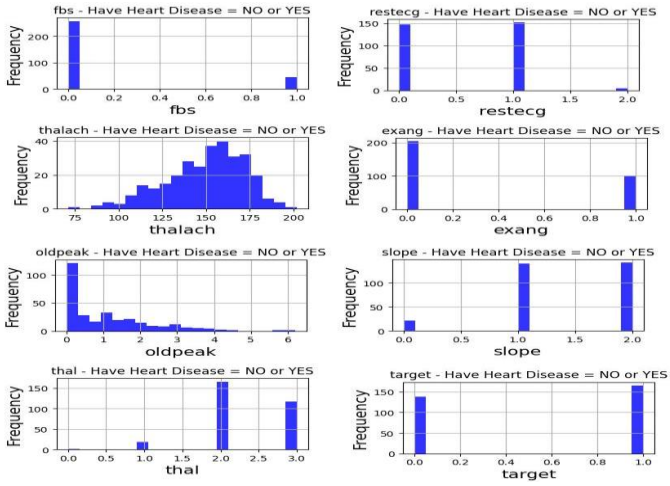

Fig. 11: Properties that have more than 5 variables.

Evaluation Metrics:

The dataset is divided into two subsets for evaluation purposes: the training dataset and the test dataset. The training dataset is used to construct the classifier, while the test dataset is employed for prediction using the trained classifier. Typically, this split allocates 80% of the data for the training dataset and 20% for the test dataset.

Table 3. Exploratory data analysis

		Predicted Class	
Actual Class	Total population = P + N	Positive (PP)	Negative (PN)
	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

True Positive (TP): Represents the number of correctly classified positive instances.

False Negative (FN): Indicates the number of positive instances incorrectly classified as negative.

False Positive (FP): Refers to the number of negative instances incorrectly classified as positive.

True Negative (TN): Represents the number of correctly classified negative instances.

Accuracy: It represents the percentage of test tuples that are correctly classified.[15]

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total Number of predictions}} \text{ [15]}$$

Precision: Measures the exactness of a classifier by calculating the percentage of positive predictions that are correct in relation to the total positive predictions [16].

$$\text{Precision} = \frac{TP}{TP+FP} \text{ [16]}$$

Recall: also referred to as sensitivity or true positive rate, is a metric that quantifies the completeness of a classifier by calculating the percentage of actual positive tuples in the test dataset that the classifier correctly identifies as positive.[16].

$$\text{Recall} = \frac{TP}{TP+FN} \text{ [16]}$$

F1-Score: The F1-Score is a measure that combines precision and recall using their harmonic mean.[16]

$$F1 = \frac{2*\text{Precision}*\text{recall}}{\text{Precision}*\text{recall}} \text{ [16]}$$

Performance Analysis of Classifiers using Area under the ROC Curve: The AUC-ROC is a crucial metric used to evaluate the accuracy of classifiers.

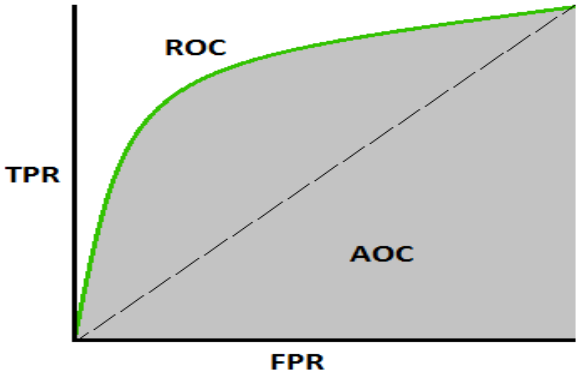

Fig. 12: ROC Curve

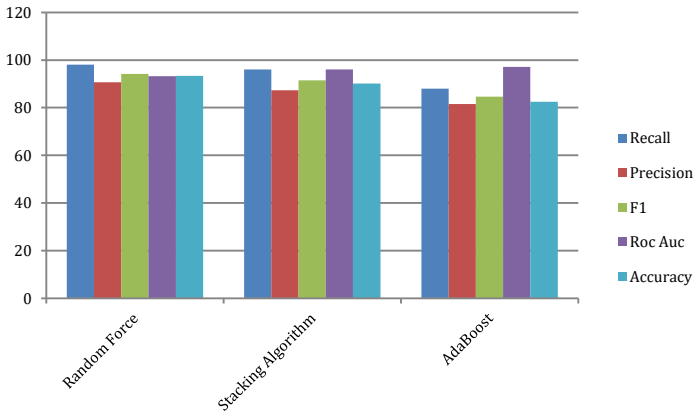

Fig. 13: Weighted average recall, precision, F-score, ROC AUC, and accuracy of the cardiovascular diseases dataset.

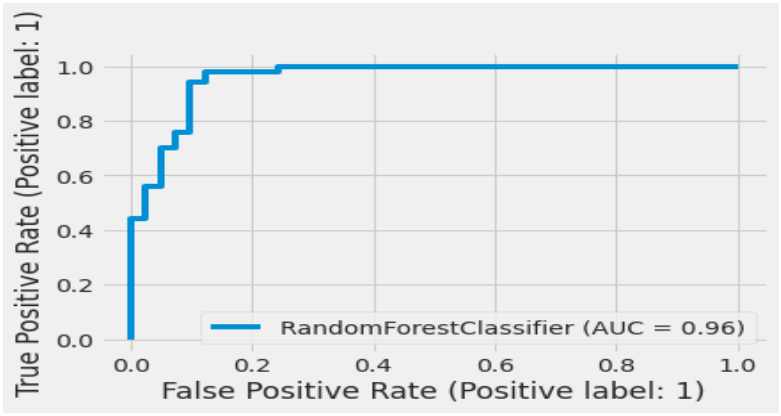

Fig. 14: ROC Curve for Random Forest Model.

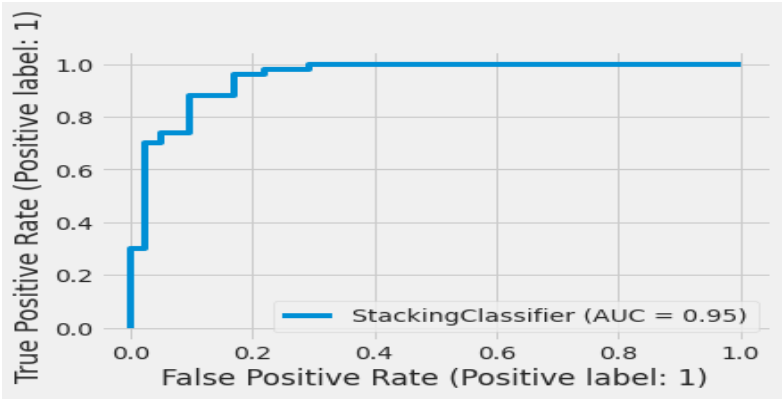

Fig. 15: ROC Curve for Stacking Model..

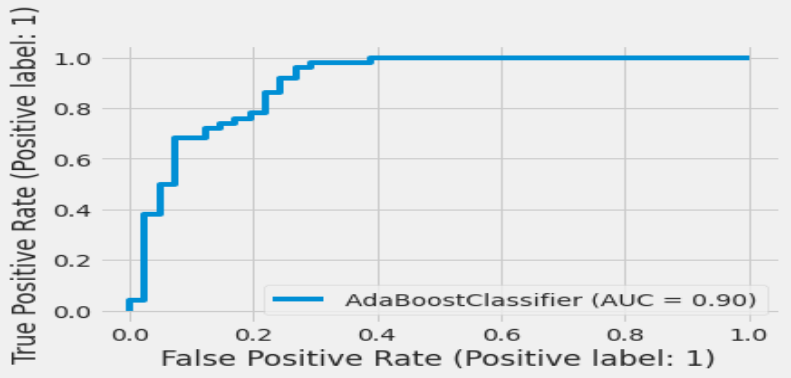

Fig. 16: ROC Curve for AdaBoost Model.

Results:

By working on the Heart Failure dataset using Python and machine learning libraries, the findings were as follows

In Table4, Table 5, the algorithms were compared in terms of Accuracy, Precision, Recall, F1-Score, ROC curve and it was found that all algorithms are equal in the total of True Positive and False Positive and the highest Accuracy of the AdaBoost Algorithm was the same as that given by ROC curve, which means that the model is able to predict the correct positive states TPR and predict the correct negative states FPR significantly.

While the Stacking Algorithm gave the lowest result and was Accuracy=84.61 while ROC curve=98.0 which means that the model is unable to predict the correct positive states TPR and predict the correct negative states FPR correctly.

Table 4. Compare Train Result of algorithms.

	Random Force Algorithm	Stacking Algorithm	AdaBoost Algorithm
TP	76	97	95
FP	21	0	2
FN	11	0	3
TN	104	115	112
Accuracy	84.91	100	97.64
Recall	90.43	100	97.39
Cohen's kappa(k)	69.34	100	95.25
F-measure	86.67	100	97.82
Precision	83.20	100	97.25
Specificity	78.35	100	97.94
Sensitivity	90.34	100	97.39
ROC Curve	91.91	100	99.56

Table 5. Compare Test Result of algorithms.

	Random Force Algorithm	Stacking Algorithm	AdaBoost Algorithm
TP	36	30	31
FP	5	11	10
FN	1	3	6
TN	49	47	44
Accuracy	93.41	84.61	97.64
Recall	98.0	94.0	88.0
Cohen's kappa(k)	86.56	68.37	68.37
F-measure	94.23	87.03	84.61
Precision	90.74	81.03	81.48
Specificity	90.0	81.03	80.0
Sensitivity	98.0	94.0	88.0
ROC Curve	93.19	0.98	97.64

Discussions

Based on the presented results, we can highlight some potential discussions and conclusions:

1. AdaBoost Algorithm Performance: The AdaBoost algorithm demonstrates excellent performance in the evaluation of heart failure-related diseases. With high accuracy, recall, and precision rates observed in both the training and testing datasets, the algorithm showcases its strong ability to accurately identify such conditions.

2. Stacking Algorithm Performance: Although the Stacking algorithm demonstrated ideal performance in the training dataset, its performance was slightly lower in the testing dataset. Two algorithms, Random Forest and XGB, were used as the stack, with Logistic Regression as the output. The small number of stacked algorithms may be the reason behind this performance difference.

3. Random Forest Algorithm Performance: The results show that the Random Forest algorithm has achieved good performance but slightly lower than AdaBoost in the testing dataset. There may be a need to review and improve the factors influencing the performance of this algorithm.

4. Importance of Feature Analysis and Selection: Based on the use of only five features out of a total of 13 in the study, it can be said that accurate feature analysis and selection play a significant role in improving the performance of algorithms. Further studies should be considered for more analysis and verification of the importance of different features.

5. Statistical Metrics Performance: The values of statistical metrics such as Cohen's kappa and F-measure indicate that the statistical models used have a good ability to predict heart failure-related cases. There may be additional improvements to be considered to enhance the values of these metrics.

Conclusion

In this paper, 3 Ensemble learning algorithms were compared and the AdaBoost algorithm gave the highest prediction in the dataset cardiovascular diseases (CVDs). It can be said that in diagnosing diseases using data mining algorithms, it is not possible to rely on Accuracy scales as a measure of the model's accuracy in prediction, the important thing is to know the degree to which the model fully understands the positive correct and negative correct states to give a correct and reliable accuracy ratio. The AdaBoost algorithm has proven to work in medical datasets more than the Random Force Algorithm and the Stacking Algorithm.

Abbreviations and Acronyms

CVDs: cardiovascular diseases. SMOTE Technique: SMOTE stands for Synthetic Minority Over-sampling Technique. ROC Curve: ROC refers to Receiver Operating Characteristic. XGB Algorithm: XGB stands for Extreme Gradient Boosting, CART Algorithm: CART stands for Classification and Regression Trees. GBM: Gradient Boosting Machine. MLP: Multi-Layer Perception. SVC: Support Vector Machine. KNN: K-Nearest Neighbors. TP: True Positive, FN: False Negative. FP: False Positive. TN: True Negative, AUC-ROC (Area Under the Receiver Operating Characteristic Curve)

References

[1] Kaggle David Lapp, 2019, Heart Disease Dataset, Kaggle, Date Access 29-9-2022, direct Link: https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset,

[2] Yiheng Li, Weidong Chen. A Comparative Performance Assessment of Ensemble Learning for Credit Scoring. Mathematics 2020, 8, 1756.

[3] Madhumita Pal, Smita Parija, 2021, Prediction of Heart Diseases using Random Forest, J.Phys.:Conf.Ser.1817 012009.

[4] Asfandyar Khan,. et al, Cardiovascular and Diabetes Diseases Classification Using Ensemble Stacking Classifiers with SVM as a Meta Classifier, Diagnostics, 2022, https://doi.org/10.3390/diagnostics12112595

[5] T. R. Mahesh, et al, AdaBoost Ensemble Methods Using K-Fold Cross Validation for Survivability with the Early Detection of Heart Disease, Computational Intelligence and Neuroscience ،Volume 2022.

[6] Md. Maidul Islam, Tanzina Nasrin Tania, Sharmin Akter, and Kazi Hassan Shakib, 2022, An Improved Heart Disease Prediction Using Stacked Ensemble Method, CC BY-NC-ND 4.0.

[7] Yash Khandelwal, 2021, Ensemble Stacking for Machine Learning and Deep Learning, Analytics Vidhya, Date Access 2-9-2022, direct Link: https://www.analyticsvidhya.com/blog/2021/08/ensemble-stacking-for-machine-learning-and-deep-

[8] Alok Kumar, Mayank Jain, (2020), Ensemble Learning for AI Developers Learn Bagging, Stacking, and Boosting Methods with Use Cases, Apress, ISBN-13 (electronic): 978-1-4842-5940-5.

[9] Byeon, H. Exploring Factors for Predicting Anxiety Disorders of the Elderly Living Alone in South Korea Using Interpretable Machine Learning: A Population-Based Study. Int. J. Environ. Res. Public Health 2021, 18, 7625.

[10] Ahmad, M., Kamiński, P., Olczak, P., Alam, M., Iqbal, M., Ahmad, F., Sasui, S., Khan, B. Development of Prediction Models for Shear Strength of Rockfill Material Using Machine Learning Techniques. Appl. Sci. 2021, 11, 6167.

[11] S. Raschka. Python Machine Learning. Packt Publishing Ltd, Third Edition, 2019.

[12] Dharmaraj Patil, Jayantrao Patil, 2018, Malicious URLs Detection Using Decision Tree Classifiers and Majority Voting Technique, Cybernetics and Information Technologies 18(1):11-29.

[13] Saini, A. (2021, September 15). Master the AdaBoost Algorithm: Guide to Implementing & Understanding AdaBoost. Analytics Vidhya. Retrieved January 15, 2023, from https://www.analyticsvidhya.com/blog/2021/09/adaboost-algorithm-a-complete-guide-for-beginners/

[14] Ramadan A.M. Elghalid, Ahmed Alwirshiffani, Abdelhafid Ali I. Mohamed, Fatimah Husayn, Amir Aldeeb, Aisha Andiasha. "Comparison of Some Machine Learning Algorithms for Some Machine Learning Algorithms for Predicting Heart Failure", 2022 International Conference on Engineering & MIS (ICEMIS.

[15] Muhammad Sakib Khan Inan, Istiakur Rahman, (2022), Integration of explainable artificial intelligence to identify significant landslide causal factors for extreme gradient boosting based landslide susceptibility mapping with improved feature selection., Machine Learning Applied to Geo-technical Engineering, arXiv, v1.

[16] Nabeela Ashraf1, Waqar Ahmad2, Rehan Ashraf3, A Comparative Study of Data Mining Algorithms for High Detection Rate in Intrusion Detection System, Annals of Emerging Technologies in Computing (AETiC) Vol.2, No.1, 2018.