Research Reports on Computer Science
https://ojs.wiserpub.com/index.php/RRCS/

UNIVERSAL WISER PUBLISHER

Research Article

Comparative Machine Learning Approaches to Analyzing the Illnesses of the Chronic Renal and Heart Diseases

Muhammad Arslan[1], Waqas Ahmad[2], Aman Ullah Yasin[1], Jahanzaib Ali Khan[3], Muhammad Nadeem[2*], Syeda Wajiha Zahra[2]

[1]Department of Creative Technology, Air University, Islamabad 44000, Pakistan
[2]Department of Computer Science, Alhamd Islamic University, Islamabad 44000, Pakistan
[3]Department of Computer Science and Software Engineering, International Islamic University, Islamabad
E-mail: nadeem72g@gmail.com

Received: 21 June 2023; **Revised:** 25 September 2023; **Accepted:** 27 October 2023

Abstract: The considerable increase in the risk of clinical events associated with chronic renal disease makes it a severe global public health issue. Chronic kidney disease (CKD) is a severe global public health issue, increasing the risk of clinical events and being associated with renal failure, cardiovascular disease, and early mortality. An accurate and timely diagnosis is essential. This research paper focuses on the global public health issue of chronic kidney disease (CKD) and its association with cardiovascular disease. It emphasizes the importance of accurate diagnosis and timely intervention for CKD, which poses significant risks to patients' health. The study proposes a machine learning (ML) approach using deep neural networks and feature selection methods to diagnose CKD and heart attack disease. The ensemble learning algorithms used in this study are decision tree (DT), logistic regression (LR), Naive Bayes (NB), random forest (RF), support vector machine (SVM), and gradient boosted trees (GBT) classifier, as well as one deep learning technique called recurrent neural network (RNN). Feature selection techniques like correlation coefficient methods are used to identify critical characteristics. The evaluation of the proposed approach was conducted using accuracy, precision, recall, and F1 measure metrics. The study employed all features for grid search and testing in each approach.

***Keywords*:** heart disease, CKD, LR, DT, RF, SVM, NB, RNN, GBT, feature selection, correlation coefficient

1. Introduction

The considerable increase in the risk of clinical events associated with chronic renal disease makes it a severe global public health issue. Chronic kidney disease (CKD) is viewed as a severe risk to society's health in the current day [1]. Chronic renal failure and heart disease both carry significant health hazards. Various types of healthcare data are currently collected in both clinical and non-clinical contexts, with the digital record of a patient's medical history acting as the most important source of information for healthcare analytics [2]. CKD can be stopped with early detection and the right medical attention. However, due to a shortage of nephrologists, not all patients with chronic renal illnesses receive a precise diagnosis. Despite having more than 10 years of expertise in the medical field, many healthcare professionals still have a low degree of awareness regarding CKD. Therefore, an automated and accurate method of

DOI: https://doi.org/10.37256/rrcs.2220233255

CKD diagnosis is needed to aid medical professionals [3]. The implementation of computer-assisted diagnostics is necessary due to the rising prevalence of patients with chronic renal illness and the dearth of specialists in this field [4]. It has received a lot of attention that CKD has a high mortality rate. The World Health Organization (WHO) claims that chronic diseases now pose a serious threat to developing countries. Renal failure develops ultimately if CKD is not treated early. In 2016, 336 million men and 417 million women died as a result of chronic renal illness, which claimed the lives of 753 million people globally. In comparison to people between the ages of 45 and 64 (12%) and 18 to 44 (6%), adults over the age of 65 are more likely to develop CKD (38%). Women experience CKD at a somewhat higher rate (14%) than men do [5]. Heart illness is a blanket term encompassing a number of conditions that have a negative impact on your heart. According to the WHO, cardiovascular diseases (CVD) currently account for 17.9 million fatalities annually, making them the leading cause of death worldwide. However, with time, more study data and hospital patient information are becoming accessible. The main muscle in the human body is the heart. In essence, it controls the flow of blood throughout our body. Any heart issue might exacerbate pain in other body areas. Heart disease is a condition that prevents the heart from functioning normally [6]. Each year, more than 10 million people lose their lives, according to the WHO. The only strategies to prevent heart-related disorders are through a healthy lifestyle and early detection. The provision of high-quality services and prompt, accurate diagnosis are the main problems in contemporary healthcare. Research can be done utilizing a variety of computer technologies to precisely identify patients and find this issue early enough to avoid it becoming fatal. The patient's records are available for free from a variety of sources [7]. Engineering [8], computer vision [9], speech recognition [10], and medical diagnostics [11] are just a few of the fields where machine learning (ML) has been shown to excel. Techniques for ML can be used to predict diseases. Numerous studies have looked at the link between CKD and heart disease, but few have looked at how ensemble learning and feature selection might improve the classification of CKD and heart attacks. This study aims to enhance the classification of chronic renal disease and cardiac disease through the application of feature selection techniques and ensemble learning. These methods use the decision tree (DT), logistic regression (LR), Naive Bayes (NB), random forest (RF), support vector machine (SVM), gradient boost tree (GBT), and recurrent neural network (RNN) deep learning algorithms in addition to six ML classifiers. We used the CKD dataset from the UCI ML repository and the heart attack prediction dataset from the Kaggle website for our inquiry. The first issue is that it is difficult to collect two data sets from scattered sites. The second goal is to evaluate both datasets and combine hyperparameter adjustment with the grid search technique. Selection (FS), is a vital preprocessing procedure that chooses the properties in a dataset that are most important. Models can be made simpler and more accurate by removing unneeded and redundant attributes. We used the correlation coefficient feature selection technique in this study. The fourth difficulty is using deep learning. All these challenges are applied to both datasets.

To forecast the diseases, the following hybrid approaches are applied to the two datasets (CKD and heart attack).

1) On full features, six ML techniques are used, and grid search is used as a technique for improving the effectiveness of ML algorithms.

2) Used a feature selection method that compares both the entire collection of features and the chosen features in various ML classification algorithms to choose important characteristics from datasets.

3) Deep learning is the most effective performance-based approach for predicting heart disease and CKD.

2. Literature review

The considerable increase in the risk of clinical events associated with chronic renal disease makes it a severe global public health issue. CKD is viewed as a severe risk to society's health in the current day. CKD is a significant public health issue around the world because it can result in detrimental effects such as renal failure, CVD, and early mortality [12]. Chronic renal failure and heart disease both carry significant health hazards. The many various types of healthcare data are currently collected in both clinical and non-clinical contexts, with the digital record of a patient's medical history acting as the most important source of information for healthcare analytics.

In order to comprehend the problem at hand and the path of action that will be most beneficial towards this endeavor, more than 15 publications were researched for the literature study for this thesis. In addition to comparing past studies on CKD and heart disease, this chapter will provide a thorough examination of numerous ML classifiers and deep learning techniques. A range of ML algorithms are taken into consideration for the project. Along with one deep learning algorithm, six other algorithms will be used since they will help to produce predictions that are better and more accurate. Deep learning, feature selection, and optimization techniques are used on the two data sets. However, we cannot declare that the forecast is correct or appropriate for the circumstances if we only employ one method, one algorithm, or one classifier and have nothing else to compare it against. Even though the algorithm may provide us with excellent accuracy, it might not be the most suitable one to utilize in this situation. The chosen algorithms are as follows: NB, SVM, and RF, LR, RNN, and GBT. The next section discusses how the prior work was carried out, demonstrated, and changed as technology advanced.

In an effort to predict heart disease and CKD, substantial research has been conducted on blood testing. The WHO claims that chronic diseases now pose a serious threat to developing countries. Renal failure develops ultimately if CKD is not treated early. In 2016, 336 million men and 417 million women died from chronic renal illness, accounting for 753 million deaths overall. Chronic renal disease can be prevented from progressing to kidney failure by early detection and treatment. The secret to controlling the chronic renal disease is early detection [13]. CKD can be stopped with early diagnosis and the right medical attention. However, because there aren't enough nephrologists to treat individuals with chronic renal illness, not all of them get the correct diagnosis. Many doctors have more than ten years of experience; however, healthcare practitioners still know very little about CKD. The increasing prevalence of patients with chronic renal illness, the lack of knowledge, and the high costs of diagnosis and treatment, particularly in developing countries, necessitate the use of computer-assisted diagnostics to help physicians' and radiologists' diagnostic assessments [14]. Estimates indicate that one in nine Korean individuals suffers from CKD, making it a very common disease. Similarly, 59% of adult Americans are at high risk of developing renal illness in the future, and it affects around 2.5-11.2% of adult Europeans. The high incidence and prevalence of CKD are attributed to its late diagnosis, particularly in developing nations [15]. Kidney function gradually and persistently declines in chronic renal failure. Early detection and treatment are essential for a good outcome and a long life. However, the decline is irreversible and unnoticeable until the disease reaches one of the later stages. A sign of what's to come for illness diagnostics is the promise of ML algorithms in this field [16]. Additionally, there is no upper age limit for CKD, therefore, it can develop at any age. Furthermore, having CKD raises the risk of a sudden loss of renal function. Procrastination in detecting the disease can have a substantial negative impact on the kidneys. Consequently, early disease detection is crucial for successful treatment. However, early-stage CKD shows no symptoms, making testing the only option for diagnosis [17]. Kidney disease affects people everywhere, yet there are significant regional differences in the prevalence, identification, and treatment of the condition. Renal failure is the biggest cause of death for people in modern society. The illness is exacerbated by several risk factors, such as smoking, binge drinking, high cholesterol, and a plethora of others. [18] While it ranges from 10% to 15% in the US, the prevalence of CKD is 10.8% in China. According to another survey, 14.7% of Mexico's adult population falls into this category. At the end of the disease, this syndrome is characterized by a progressive loss of renal function. At first, there are no signs of CKD. The condition may not be identified until the kidney has lost about 25% of its functionality as a result [19]. WHO estimates show that heart disease accounts for 7.9 million annual fatalities. When arterial plaque accumulates and interferes with the heart's ability to pump blood, a heart attack occurs. A thrombus in an artery that prevents blood from reaching the brain causes a stroke [20]. Numerous studies have focused on categorization and prediction for the diagnosis of heart illness, and numerous ML models are being used. In order to develop an autonomous classifier for forecasting congestive heart failure, Melillo et al. used an ML method called CART (Classification and Regression Tree). The electrocardiogram (ECG) technique is then suggested by Rahhal et al., who used deep neural networks to select the best properties before implementing them [21]. The rise in medical data collection has given doctors a new chance to improve patient diagnoses. In order to enhance decision assistance, practitioners have increased their use of computer technology in recent years. ML is becoming a more important treatment in the healthcare sector to help with patient diagnostics [22]. The prevalence of CVD is increasing daily in the modern environment. According to the WHO, heart attacks and strokes are to blame for 17 million of the world's annual fatalities from CVD. Therefore, it is necessary to note the key behaviors and warning indications of CVD [23]. In the US, coronary artery disease is the major cause of heart attacks. The WHO study found that cardiac diseases were to blame for 24% of deaths in India. Numerous risk factors for heart disease and coronary artery disease have been found by researchers [24]. In this study, multiple readings were carried out using not only various methodologies but also by connecting two or more procedures in order to construct a prediction model. Hybrid methods are a common name for these merged techniques [25]. With the aid of ML, numerous diseases may be recognized, found, and forecasted in the medical industry. Following data analysis, ML approaches will aid in the early identification and prediction of cardiac disease. This study evaluates the performance of various ML techniques for early heart disease prediction, including NB, DT, LR, and RF [26]. An essential part of the human body is the heart. If the heart is not working properly, it will affect the kidney, brain, and other human-like organs. According to WHO figures, heart disease killed one-third of all people worldwide in 2017 and was the leading cause of death in developing nations [27]. In the healthcare sector, there is a huge amount of untapped patient medical data. It is necessary to conduct data analysis on these patient medical records or data [28] because they contain hidden patterns [29]. Because it affects people's physical, emotional, and social well-being, health is one of the most important subjects that requires a lot of attention. This is a result of several diseases that attack people silently yet are quite harmful. One of these silent yet deadly killers that increases the number of fatalities each year is heart disease [30].

3. Proposed methodology

The two diseases that this proposed method is intended to predict are chronic renal disease and heart attack. The first approach employed to predict CKD and heart disease is to use grid search as an optimization tool in combination with ML methods including DT, LR, RF, SVM, NB, and GBT as ensemble learning on full features. The second approach uses feature selection techniques to extract the most important features from the datasets for CKD and heart disease. The RNN deep learning model is the third tactic. The first of the five processes in the proposed system is data gathering. The CKD dataset and the heart attack dataset will be used for this stage. The second step, data preprocessing, is where null values will be handled. To fine-tune the settings for the ML and ensemble learning algorithms, the third phase uses a grid search. In the fourth step, the crucial features will be selected by utilizing the feature selection methods. In the fifth step, deep learning is used.

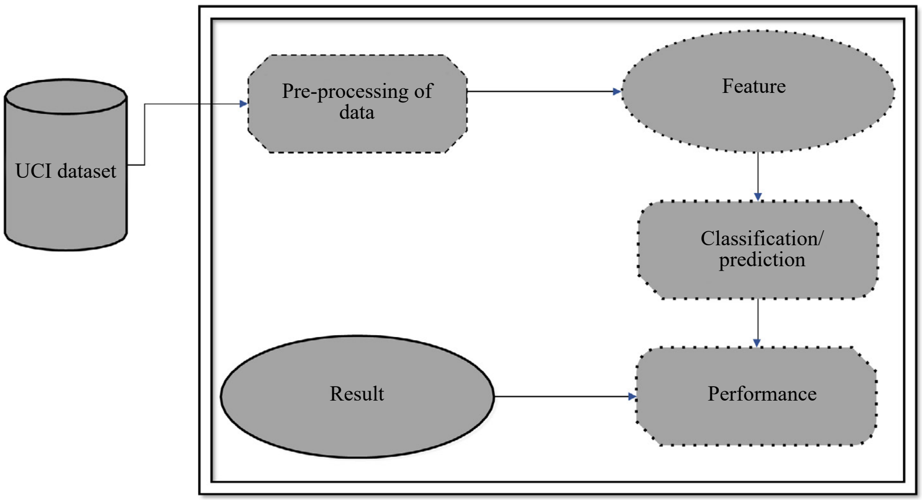

Figure 1. Procedural diagram

3.1 *Data collection*

This study employed two datasets of medical conditions. Heart disease and CKD are two examples. The dataset used in the study on CKD was obtained from the UCI ML repository. The CKD dataset consists of 400 samples, 25 features, and 1 class label. The class label has two values: ckd and notckd.

3.2 *Description of dataset*

1st dataset: From the UCI ML repository.
Title: Chronic kidney disease (CKD)
Number of instances: 400 (250 CKD, 150 notckd)

Number of attributes: 24 + class = 25 (11 numeric,14 nominal)
https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease
2nd dataset: From the Kaggle website

3.3 *Data preprocessing*

Addressing missing and noisy information in the dataset is part of data preparation. The procedure entails cleaning the data, getting rid of errors and faults, and filling in the blanks. Rather than eliminating items, algorithms are employed to infer missing data. The mean is used for numerical attributes, whereas the mode is used to fill in missing values for nominal features. Additionally, during this phase, noise such as outliers is removed, the data is normalized, and the data balance is checked. The objective is to assure accurate and pertinent results by properly preparing the data.

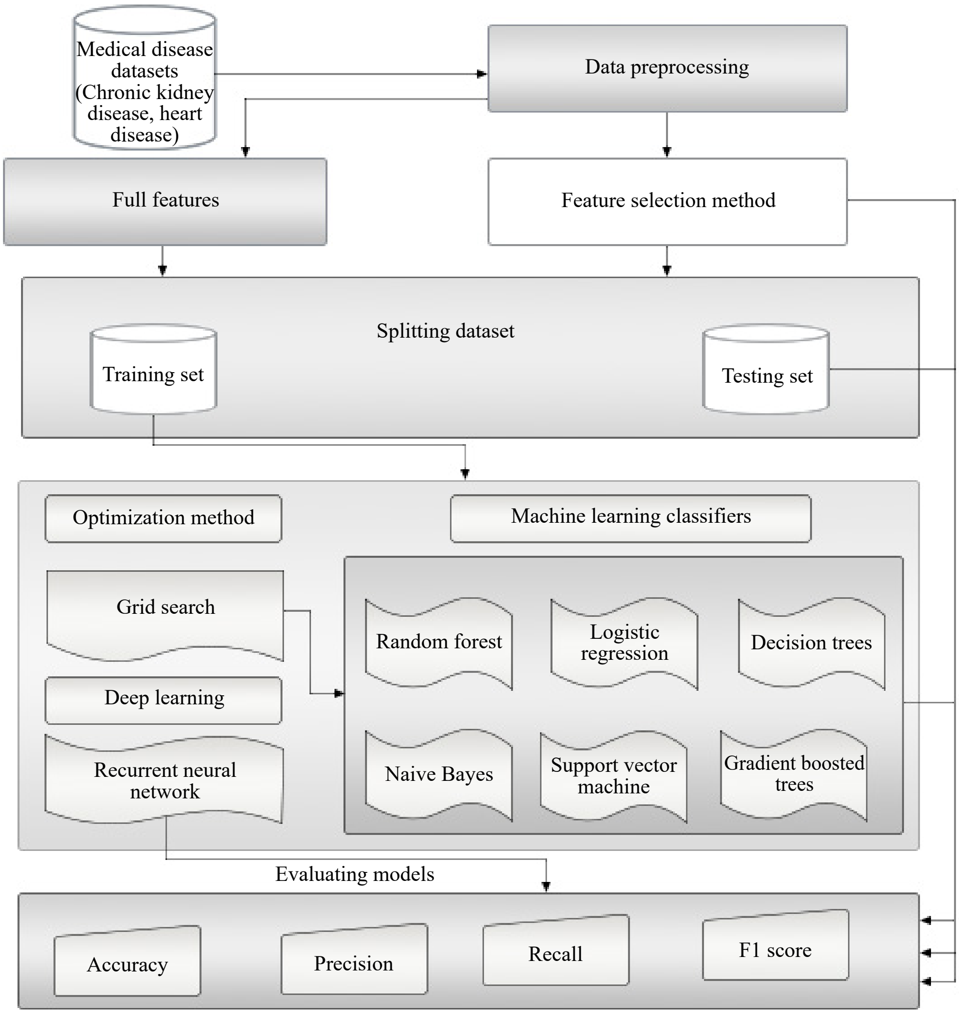

Figure 2. Methodology diagram

3.4 *ML techniques*

A cloud-based Jupyter Notebook environment called Google Colab was used for the exploratory study. To make it simple to access and share notebooks that are tightly connected with Google Drive, the Earth Engine Python API was established in Colab. Colab is a Python 3 program that may be used for a variety of activities, including data science, ML, preprocessing massive datasets, and predictive analysis. A ratio of 80:20 was employed to partition the dataset, with 80% of the data being used for training and the remaining 20% for testing.

3.4.1 *DT*

The DT algorithm has nodes that reflect parameter testing and class labels that are applied to each node. Both discrete and continuous input and output variables are supported. Decision-makers move along the tree, starting at the root node, to identify the most distinct class based on the information gained. Although they are prone to overfitting, DT are effective for managing continuous and categorical characteristics.

3.4.2 *RF*

A supervised ML method called random forest uses different decision trees to increase prediction accuracy. It creates trees dynamically using the bagging technique and online fitting. The final prediction is created by averaging the projected values from each separate tree training. The capacity of each tree and their association with one another affect the generalization error in tree classifiers.

3.4.3 *LR*

Binary categorization is done using the statistical regression analysis technique known as logistic regression. It uses a logistic or sigmoid function to forecast the likelihood of various labels for an unlabeled observation. In LR, the dependent variable is always binary. Prediction and calculating success probability are the main uses of LR.

3.4.4 *SVM*

SVM is a supervised ML approach for classifying data that locates a hyperplane to divide classes using margins and support vectors. It manages both linear and non-linear data, offering precision, the capacity to model complicated boundaries, and lowering overfitting.

3.4.5 *NB*

The Bayes theorem is used by the straightforward yet powerful probabilistic classifier known as Naive Bayes. In contrast to other algorithms, it treats each attribute independently and needs less training data.

3.4.6 *Bernoulli Naive Bayes*

A variant of NB called Bernoulli Naive Bayes (BNB) was created for data with binary-valued features. It works well with discrete data and presupposes a multivariate Bernoulli distribution. The BNB algorithm stands out due to its requirement for binary feature values. Although binary feature vectors are still required, the scikit-learn library provides the BNB class, which enables feature change using a threshold value.

3.4.7 *GBT*

An approach called gradient boosting trees trains a collection of DT sequentially. With each iteration, the model is improved since each tree is optimized using information from previously trained trees. Particularly when employing shallow trees, the ensemble of trees can aid in lowering overfitting. By comparing the predicted labels for training samples to the actual labels, GBT constantly trains multiple DT.

3.5 *Deep learning*

RNNs are one type of artificial neural network that is essential to deep learning, a branch of ML. The linked organization of the human brain is mimicked by these network topologies. Deep learning models analyze data using layered algorithms to reach conclusions, much like a human brain. By utilizing the neural network of the human brain as inspiration for their architecture, artificial neural networks outperform conventional ML techniques.

3.5.1 *RNN*

A sort of deep learning method called RNN is used to represent sequential data, like time series and spoken language. Prior to the popularity of attention models, they were commonly utilized. In a deep feedforward model, RNNs may manage variable parameters for each element of a sequence. They are used in fields like self-driving cars and high-frequency trading. In an RNN, the results from each layer are fed back into the input to forecast the results from the following layers. The success rate of a model's predictions is measured by its accuracy, and a loss function determines the discrepancy between predicted and expected results. For binary classification issues, binary cross-entropy is frequently used.

3.6 *Optimization methods*

In ML, grid search is a popular method for hyperparameter optimization. In order to determine the combination that produces the optimum model performance, it entails methodically assessing each potential value for the hyperparameters. To optimize models and improve their performance, hyperparameters are crucial. Tools for hyperparameter optimization are provided by the scikit-learn Python machine-learning package, enabling the configuration of a model with the most appropriate hyperparameters.

3.6.1 *Grid search*

Grid search is a method for finding the best model parameters by methodically analyzing various parameter value combinations from a specified list. It applies to different models and automates the process of experimenting with different parameter choices to discover the most effective ones. Grid search assesses each combination by generating a grid of hyperparameter values and is helpful, especially for examining widely used successful parameter combinations.

3.7 *Feature selection method*

3.7.1 *Correlation coefficient*

Correlation can be used to determine whether two or more variables are linearly related. Correlation allows us to predict one variable based on another; a correlation can be used to select features. If there is a relationship between two variables, we can predict one variable from another. If two features are connected, the model only genuinely assesses one of them, as the other one does not provide any new information.

3.8 *Evaluating the methods*

Accuracy, recall, and F1 score, where FP stands for false positive FN is for false negative TP is for true positive, and TN is for true negative. These are the four standard metrics used to assess the models.

3.8.1 *Confusion matrix*

A categorization algorithm's performance is evaluated using an M×M matrix called a confusion matrix, where M is the number of target classes. This matrix distinguishes between target values that were attained and anticipated values produced by the ML algorithm. In most cases, binary classification issues are solved using a 2×2 matrix.

Accuracy = TP + TN/TP + FP + TN + FN

Precision = TP/TP + FP

Recall = (TP/TP + FN)

F1 = 2/(1/precision+1/recall)

3.8.2 *Accuracy*

Accuracy is defined as the proportion of data points that correctly predict the data.

3.8.3 *Precision*

The precision benchmark is used to calculate how many accurate positive predictions were produced.

3.8.4 *Recall*

This characteristic represents the fraction of correct positive predictions that were generated out of all possible positive predictions. Simply explained, it is the prediction of a positive class based on all positive classes.

3.8.5 *F1 score*

A model's accuracy on a dataset is evaluated using the F-score, often known as the F1 score. It is employed to assess classification methods that label examples as "positive" or "negative".

4. Experimental results and discussion

The section analyzes the dataset using correlation coefficient feature selection and evaluates the performance of ML algorithms (SVM, LR, NB, RF, DT, GBT classifier) and the deep learning method RNN with full and selected features.

4.1 *Accuracy of models with all features*

We have used SVM, DT, RF, Gaussian NB, BNB, LR, and gradient boosting. In our initial test, we used a tree classifier and a RNN on two datasets that had all the features.

4.1.1 *DT*

Results from Dataset 1: The DT confusion matrix (Figure 3) indicates that the system accurately identified 34 true positive class data points and correctly detected all 46 true negative class data points. There were no instances where negative class data points were falsely identified as the positive class (false positives) or positive class data points were missed (false negatives).

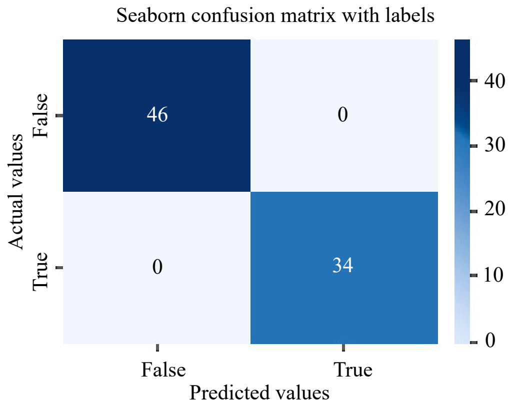

Figure 3. Confusion matrix for DT (Dataset 1)

Results from Dataset 2: The DT confusion matrix (Figure 4) reveals that the algorithm accurately identified 33 true positive class data items and correctly detected all 518 true negative class data items. However, it misclassified 111 data points as false positives and identified 70 positive class data points as false negatives, with only one negative class data point.

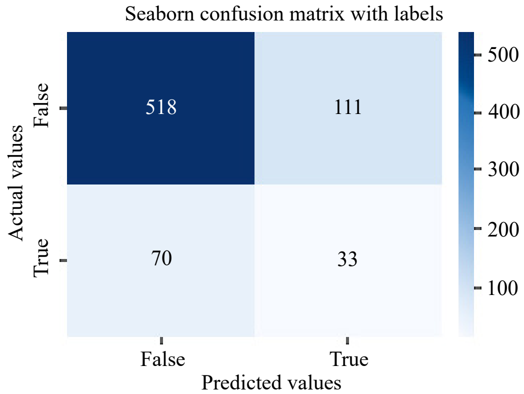

Figure 4. Confusion matrix for DT (Dataset 2)

4.1.2 *RF*

Results from Dataset 1: The RF confusion matrix (Figure 5) shows that the algorithm accurately detected 25 true positive class data points and correctly identified 55 true negative class data points. There were no false positives and no false negatives.

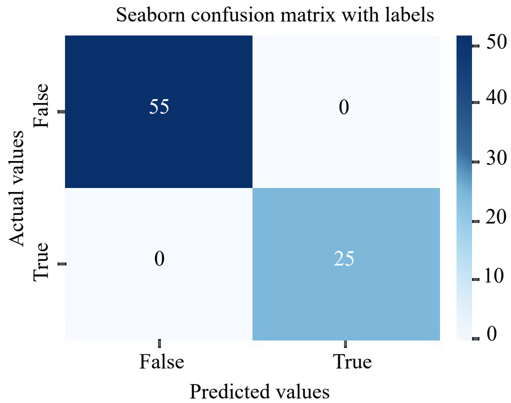

Figure 5. Confusion matrix for RF (Dataset 1)

Results from Dataset 2: The RF confusion matrix (Figure 6) indicates that the algorithm correctly detected all 616 true negative class data points. However, there were no true positive class detections, 2 false positive matches, and 114 false negatives with value.

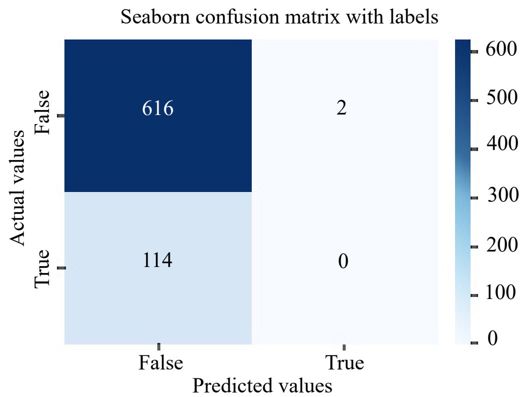

Figure 6. Confusion matrix for RF (Dataset 2)

4.1.3 *LR*

Results from Dataset 1: The LR confusion matrix (Figure 7) shows that the system accurately detected 34 true positive class data points and correctly identified all 46 true negative class data points. There were no false positive matches or false negatives.

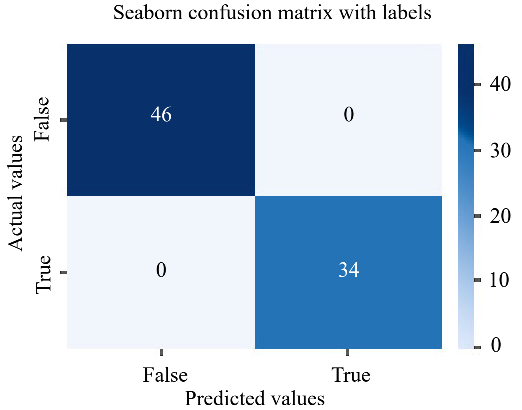

Figure 7. Confusion matrix for LR (Dataset 1)

Results from Dataset 2: The LR confusion matrix (Figure 8) shows that the system accurately detected four true positive class data points and correctly identified all 640 true negative class data points. There were two false positive matches and 86 false negatives.

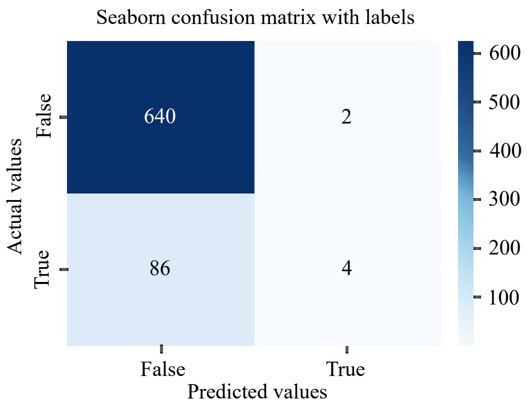

Figure 8. Confusion matrix for LR (Dataset 2)

4.1.4 *SVM*

Results from Dataset 1: The SVM confusion matrix (Figure 9) indicates that the system accurately detected 27 true positive class data points and correctly identified all 52 true negative class data points. There was one false positive match, and there were no false negatives.

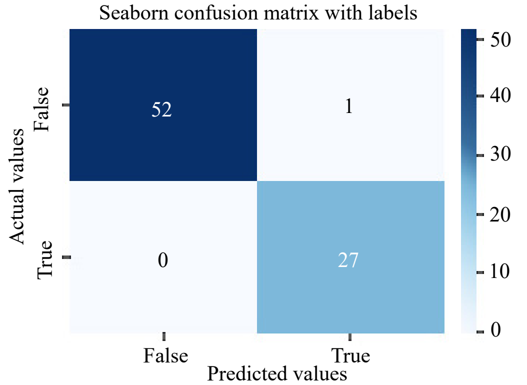

Figure 9. Confusion matrix for SVM (Dataset 1)

Results from Dataset 2: The SVM confusion matrix (Figure 10) reveals that there were no true positive class data points detected, and all 620 true negative class data points were correctly identified. There were no false positive matches, but there were 112 false negatives with value along with 112 false negatives with negative class data points.

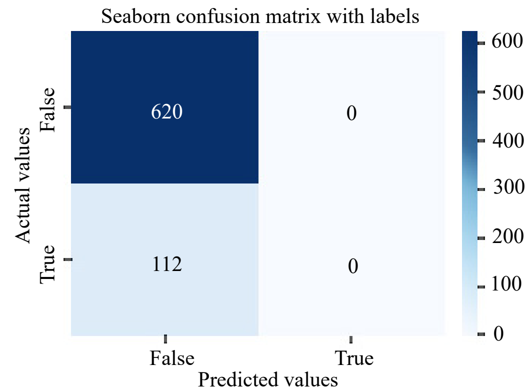

Figure 10. Confusion matrix for SVM (Dataset 2)

4.1.5 *NB*

Results from Dataset 1: The NB confusion matrix (Figure 11) indicates that the system accurately detected 27 true positive class data items and correctly identified 50 true negative class data items. There were no false positive matches. However, 3 false negatives were negative class data points.

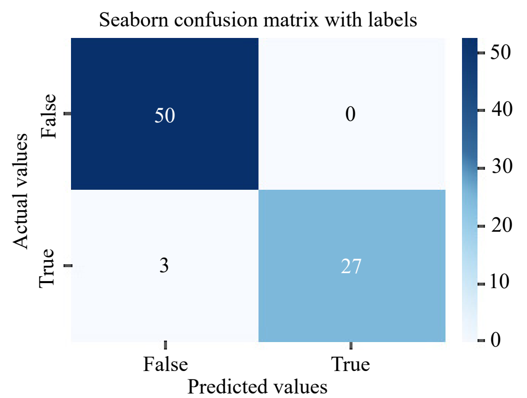

Figure 11. Confusion matrix for NB (Dataset 1)

Results from Dataset 2: The NB confusion matrix (Figure 12) indicates that the system correctly detected 21 true positive class data items and accurately identified all 573 true negative class data items. There were 46 false positive matches and 92 false negatives with values, where positive class data items were mistaken for negative class data points.

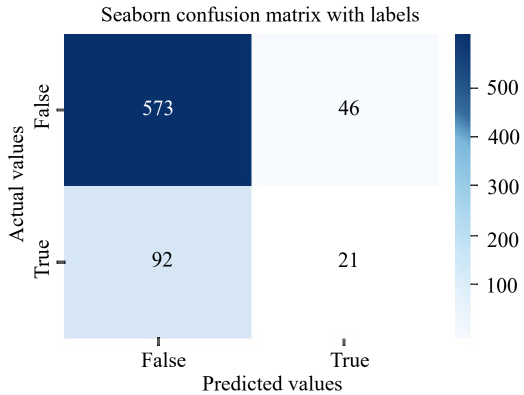

Figure 12. Confusion matrix for NB (Dataset 2)

4.1.6 *BNB*

Results from Dataset 1: The BNB confusion matrix (Figure 13) indicates that the algorithm correctly detected 30 true positive class data items and accurately identified all 50 true negative class data items. There were no false positive matches, and there were no false negatives.

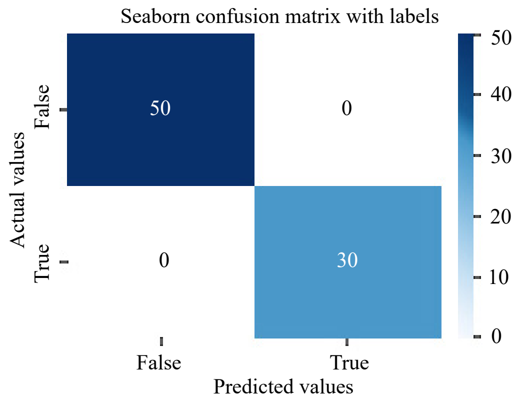

Figure 13. Confusion matrix for BNB (Dataset 1)

Results from Dataset 2: The BNB algorithm correctly identified 33 true positive class data items and 553 true negative class data items (Figure 14). However, it had 66 false positive matches and 80 false negatives, including cases where positive class data was mistaken for negative class data.

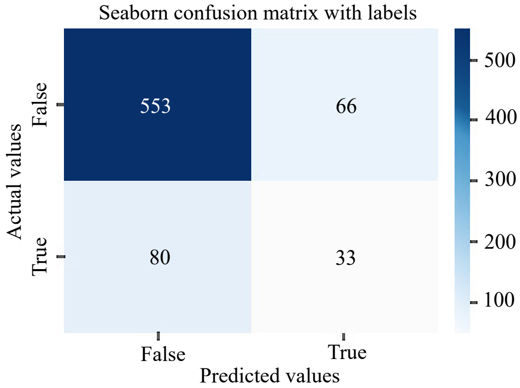

Figure 14. Confusion matrix for BNB (Dataset 2)

4.1.7 *GBT*

Results from Dataset 1: The GBT confusion matrix (Figure 15) indicates that the algorithm correctly detected 32 true positive class data items and accurately identified all 48 true negative class data items. There were no false positive matches, and there were no false negatives.

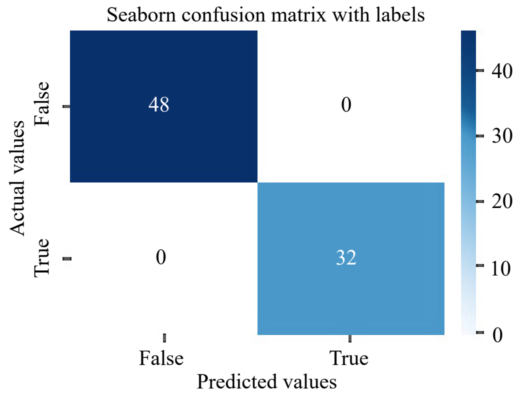

Figure 15. Confusion matrix for GBT (Dataset 1)

Results from Dataset 2: The GBT confusion matrix (Figure 16) indicates that the algorithm correctly detected 17 true positive class data points and accurately identified all 579 true negative class data points. There were 48 false positive matches and 88 false negatives with values, where positive class data was mistaken for negative class data.

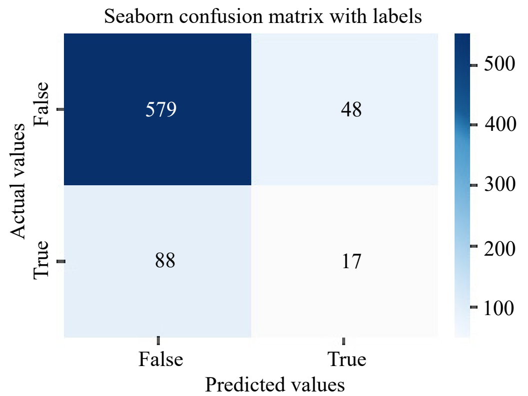

Figure 16. Confusion matrix for GBT (Dataset 2)

4.1.8 *RNN*

Results from Dataset 1: The loss function (Figure 17), is used to measure the quantitative loss across all data items in an epoch. The loss is presented as a curve across iterations for a portion of the dataset. Epochs represent one-loop structures for training the neural network, and multiple epochs are usually needed. The goal is to minimize the loss, which is a scalar value, during model training.

Accuracy (Figure 18) is an important metric for evaluating classification models, representing the proportion of accurate predictions. It provides a clear measure of algorithmic performance. Generally, accuracy increases as loss decreases, although they are defined differently and are not mathematically related.

When the model is put together, the following values are returned: loss= 'binary_crossentropy', optimizer="adam", epoch 90/90 and 320/320, loss: 0.2865, accuracy: 0.9344, val_loss: 0.6257, val_accuracy: 0.8750, 900ms/epoch, and step: 3ms.

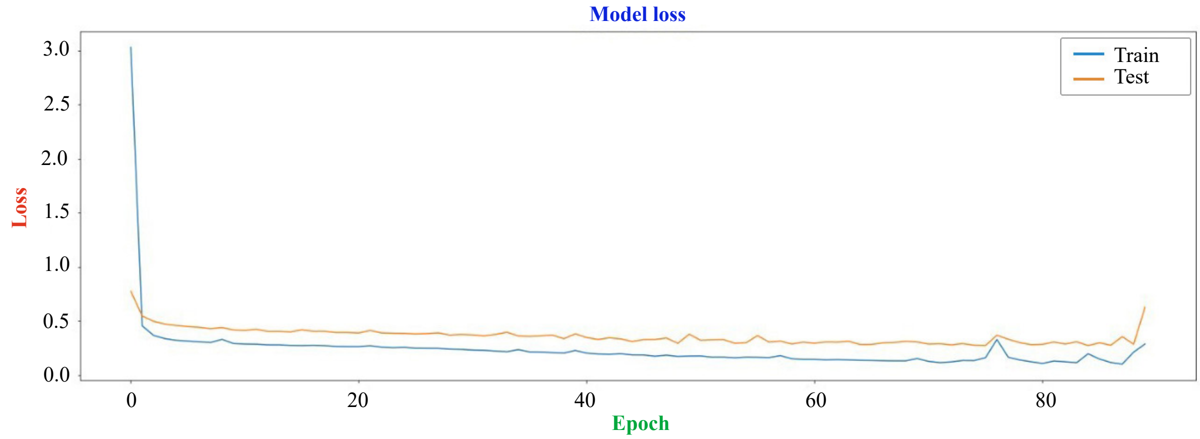

Figure 17. RNN for model loss (Dataset 1)

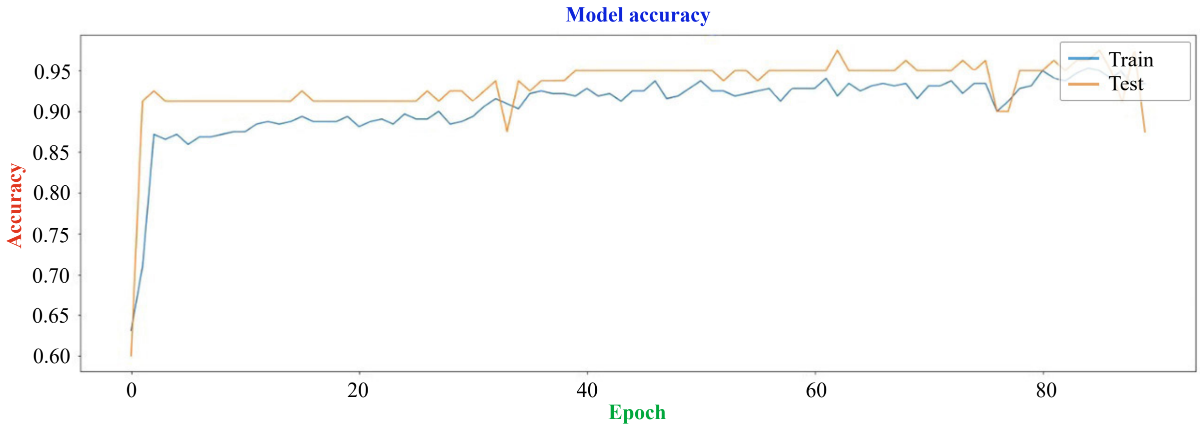

Figure 18. RNN for model accuracy (Dataset 1)

Results from Dataset 2: The following values are returned in (Figure 19) and (Figure 20) when the model has been built: Loss ='binary_crossentropy', optimizer='adam', epoch 90/90 and 320/320, loss: 0.2865, accuracy: 0.9344, val_loss: 0.6257, val_accuracy: 0.8750, 900ms/epoch, and step: 3ms.

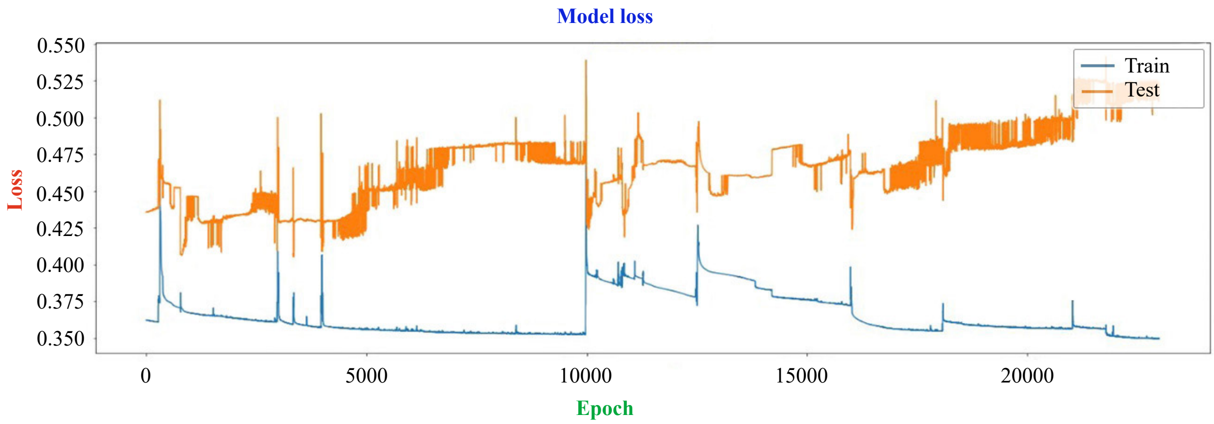

Figure 19. RNN for model loss (Dataset 2)

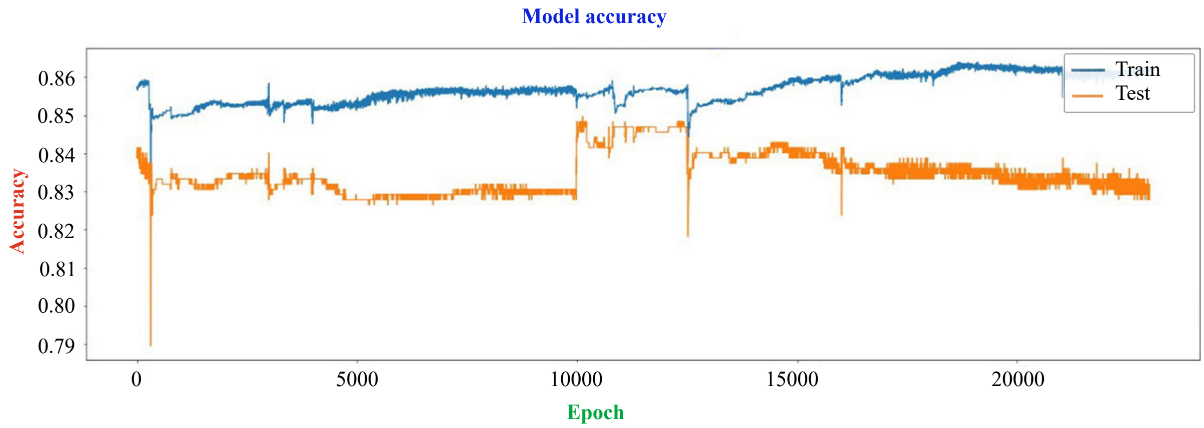

Figure 20. RNN for model accuracy (Dataset 2)

Table 1 and Table 2 display the various algorithms' performance for the two datasets. Results of classifiers after applying Grid search are shown in Tables 3 and 4.

Table 1. Performance of the algorithms (Dataset 1)

Classifier	Accuracy	Precision	Recall	F1 score
RF	100%	100%	100%	100%
SVM	98%	100%	98%	99%
NB	96%	94%	100%	97%
BNB	99%	-	-	-
GBT	100%	100%	100%	100%
DT	100%	100%	100%	100%
LR	100%	100%	100%	100%
RNN	93%	-	-	-

Table 2. Performance of the algorithms (Dataset 2)

Classifier	Accuracy	Precision	Recall	F1 score
RF	84%	84%	99%	91%
SVM	84%	84%	100%	91%
NB	81%	86%	92%	89%
BNB	80%	-	-	-
GBT	81%	86%	92%	89%
DT	75%	88%	82%	85%
LR	87%	87%	99%	93%
RNN	86%	-	-	-

Table 3. Accuracy of the algorithms (Dataset 1)

Classifier	RF	SVM	BNB	GBT	DT	LR
Accuracy	100%	99%	99%	99%	90%	99%

Table 4. Accuracy of the algorithms (Dataset 2)

Classifier	RF	SVM	BNB	GBT	DT	LR
Accuracy	85%	85%	80%	84%	84%	84%

4.2 *Feature engineering*

Utilizing important features significantly impacts algorithm accuracy as selecting fewer traits enables faster training and leveraging connections between critical features can yield unexpected improvements. While linear relationships between certain qualities can lead to model overload, feature selection improves algorithm accuracy by selecting crucial features and it could reduce training time and mitigate overfitting.

4.3 *Feature importance*

"Feature importance" determines which attributes have the greatest impact on predictions. Some dataset properties may have a minimal impact on forecasts. A limited number of factors can decrease model accuracy. Using the right attributes is crucial for optimal results. Grid search optimization is used to assess the classifier's accuracy. This approach helps evaluate accuracy variations.

4.4 *Accuracy of models with selected feature*

Grid search is used to optimize model performance, followed by feature selection techniques. Information gain, Fisher's score, and correlation coefficient are employed for feature selection. The most important features from both datasets are selected for comparison and prediction assessment.

4.5 *Feature selection method*

The objective of feature selection in ML is to identify the optimal set of characteristics for modeling the phenomenon under study.

4.5.1 *Information gain*

Figures 21 and 22 are the Fisher's score for datasets 1 and 2, respectively.

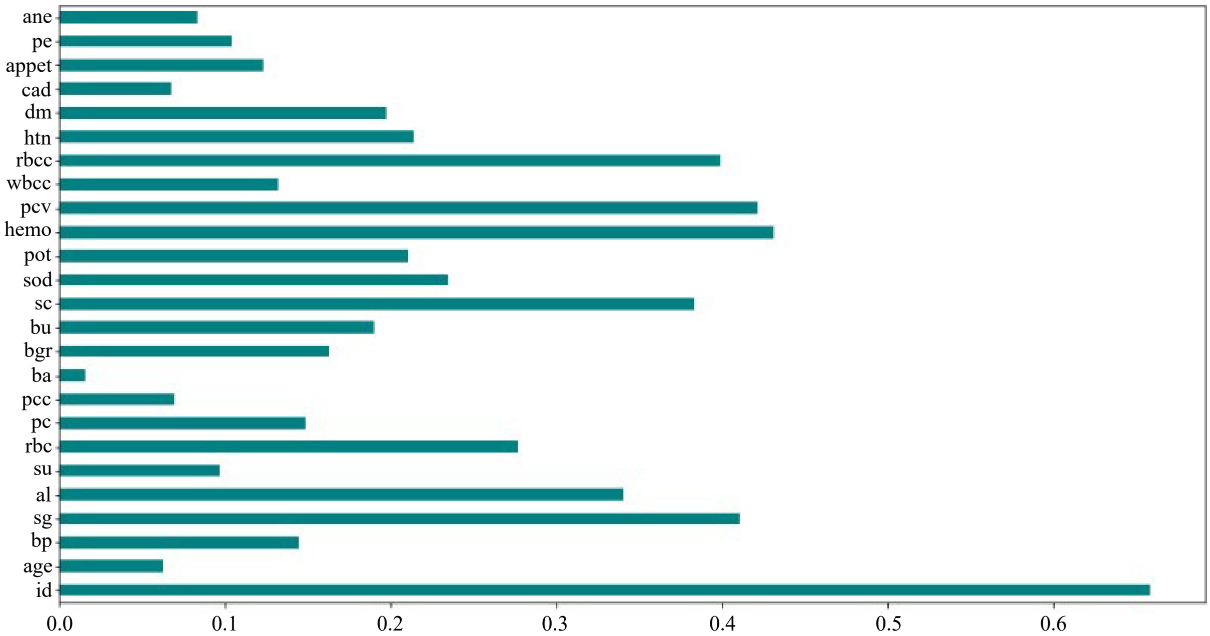

Figure 21. Feature ranking applying information gain (Dataset 1)

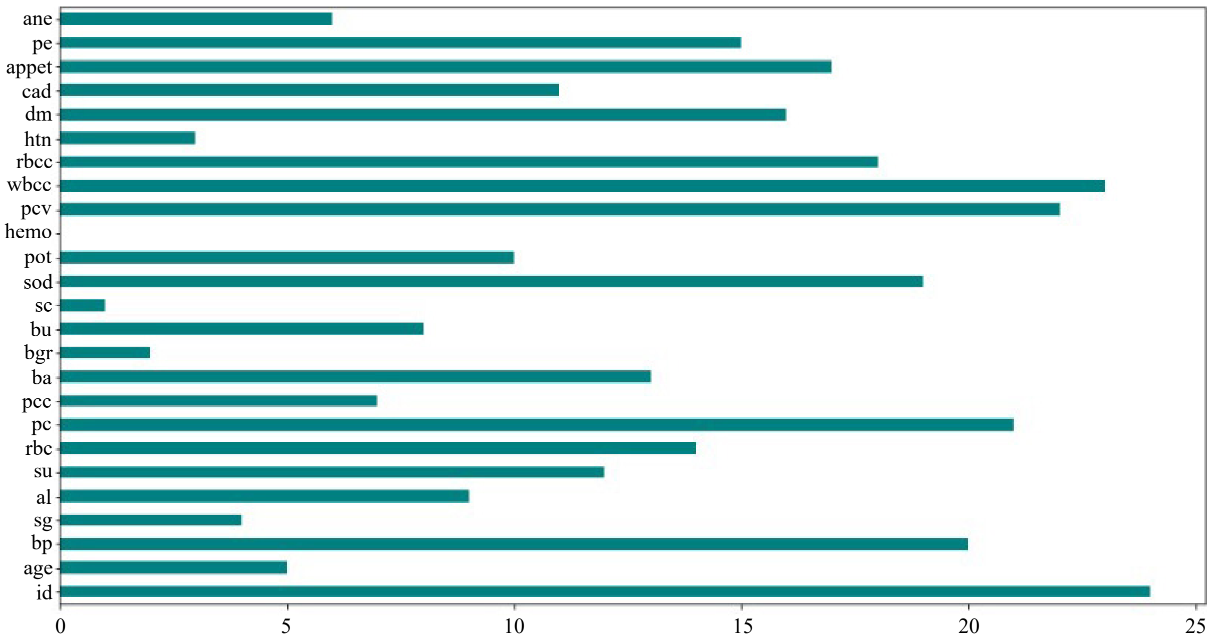

Figure 22. Feature ranking applying information gain (Dataset 2)

4.5.2 *Fisher's score*

Figures 23 and 24 are the Fisher's score for datasets 1 and 2, respectively.

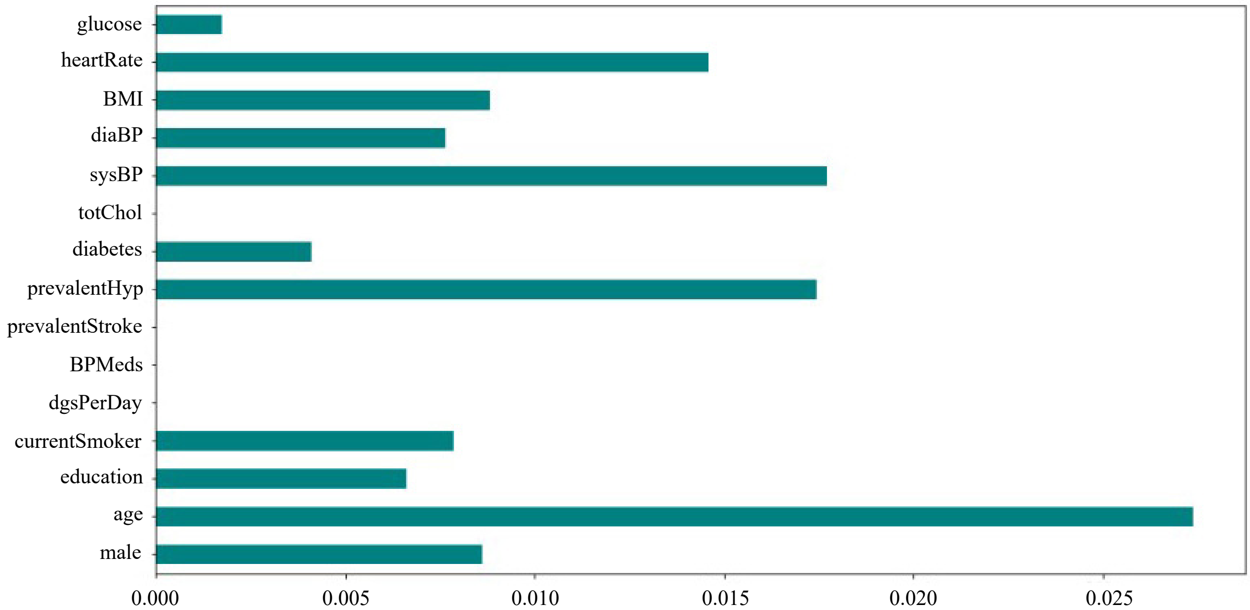

Figure 23. Feature ranking applying Fisher's score (Dataset 1)

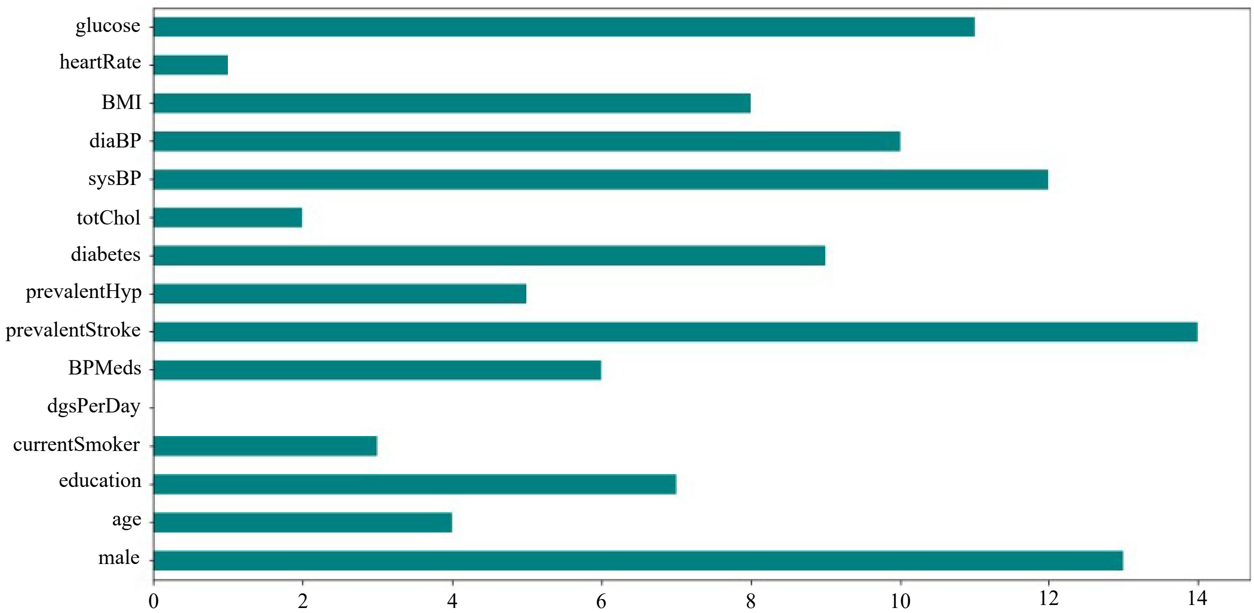

Figure 24. Feature ranking applying Fisher's score (Dataset 2)

4.5.3 *Correlation coefficient*

Figures 25 and 26 are the correlation coefficient for datasets 1 and 2, respectively.

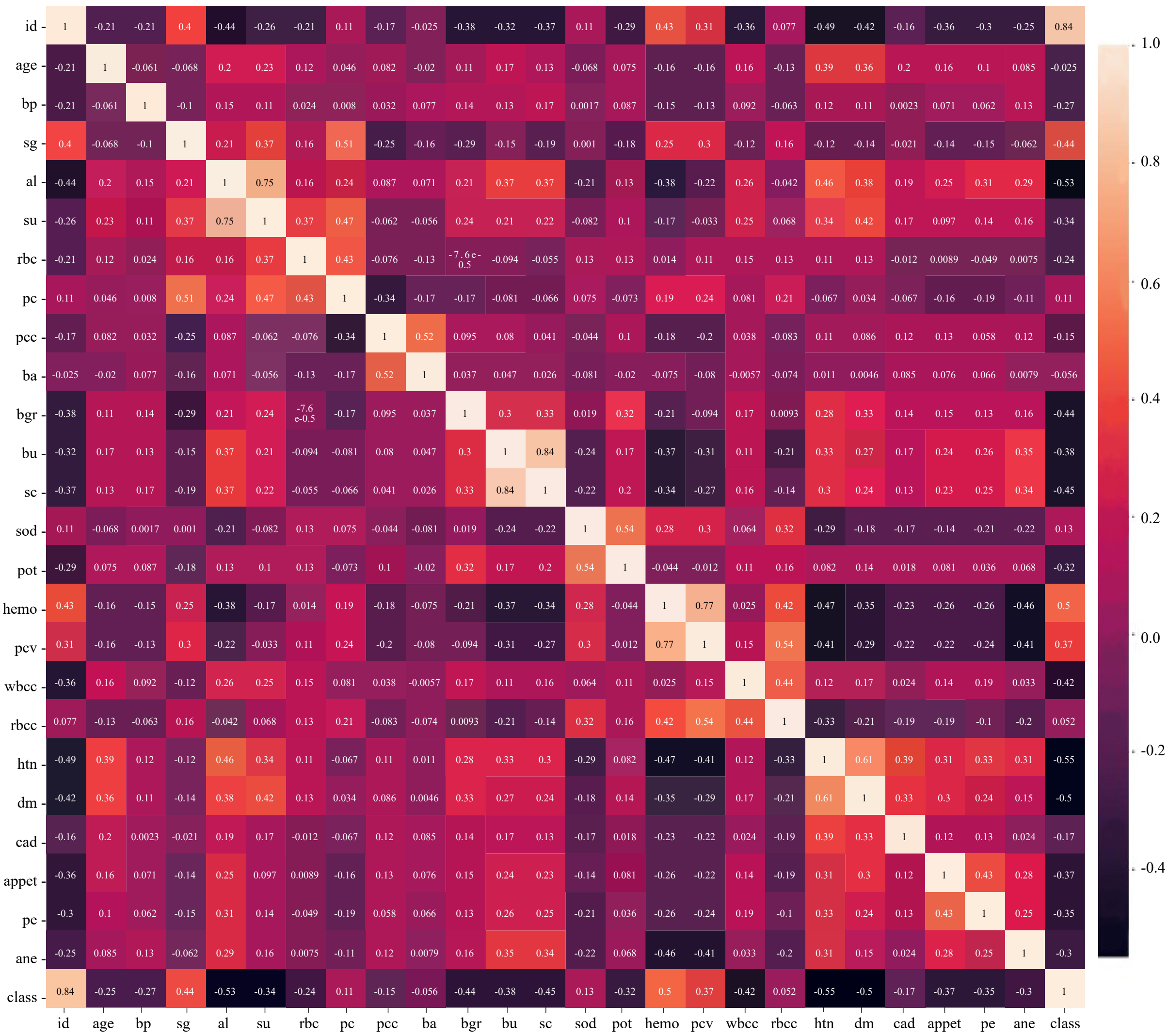

Figure 25. Feature ranking applying correlation coefficient (Dataset 1)

Figure 26. Feature ranking applying correlation coefficient (Dataset 2)

4.5.4 *DT*

Results from Dataset 1: The DT confusion matrix results revealed that the system accurately identified 6 positive class data points (Figure 27). It correctly detected all 31 negative class data points as true negatives. However, there were 15 instances where the algorithm misclassified negative class data points as positive, resulting in false positive matches. Additionally, 28 positive class data points were mistakenly classified as negative, leading to false negatives.

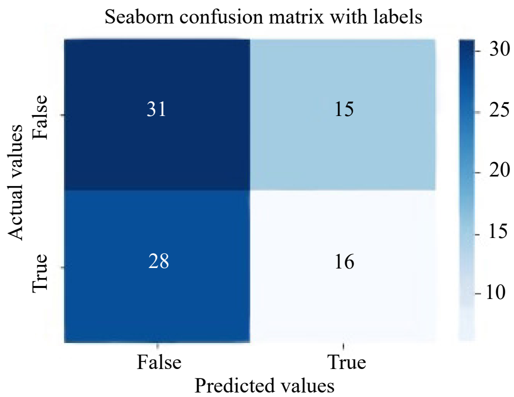

Figure 27. Confusion matrix for DT (Dataset 1)

Results from Dataset 2: The DT confusion matrix (Figure 28) indicates that the algorithm correctly classified 0 positive class data items. It accurately identified all 627 negative class data items as true negatives. There were two instances of false positive matches, where negative class data points were mistakenly classified as positive. Additionally, 103 valuable positive class data points were mistakenly classified as negative.

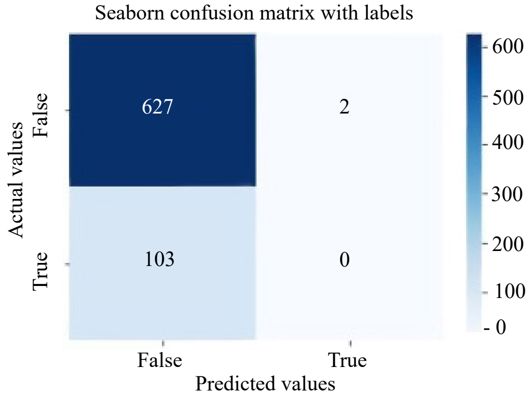

Figure 28. Confusion matrix for DT (Dataset 2)

4.5.5 *GBT*

Results from Dataset 1: The GBT confusion matrix (Figure 29) revealed that the algorithm accurately detected 35 positive class data points. It correctly classified all 37 negative class data points as true negatives. There were 7 false positive matches, where negative class data points were mistakenly classified as positive. Additionally, there was one false negative with value, involving the misclassification of a positive class data point as negative and vice versa.

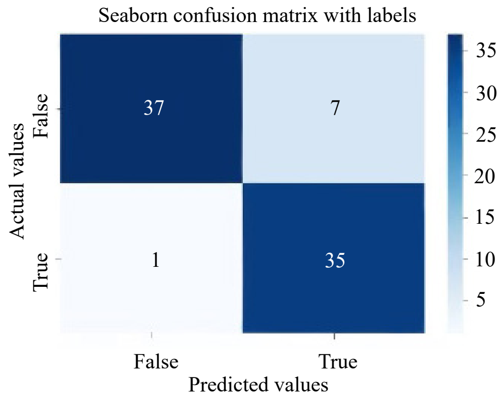

Figure 29. Confusion matrix for GBT (Dataset 1)

Results from Dataset 2: The GBT confusion matrix (Figure 30) showed that the algorithm accurately detected 10 positive class data points. It correctly classified all 589 negative class data points as true negatives. There were 38 false positive matches, where negative class data points were mistakenly classified as positive. Additionally, there were 95 true positive class data points classified as false negatives with a value.

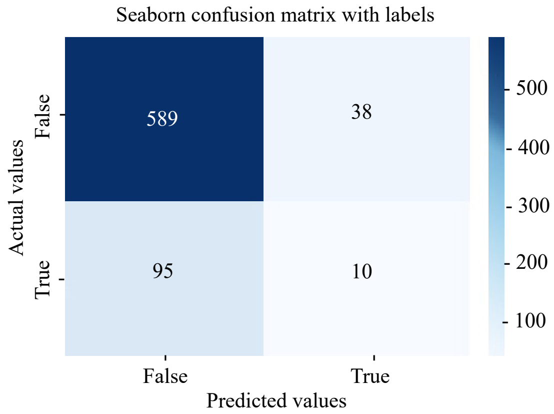

Figure 30. Confusion matrix for GBT (Dataset 2)

4.5.6 *LR*

Results from Dataset 1: The LR confusion matrix (Figure 31) shows that the algorithm detected 36 positive class data points correctly. It accurately identified all 42 negative class data points as true negatives. Two data points from the negative class were mistakenly classified as positive, resulting in false positive matches. There were no false negatives in terms of positive and negative class data points, but false negatives with a value were present.

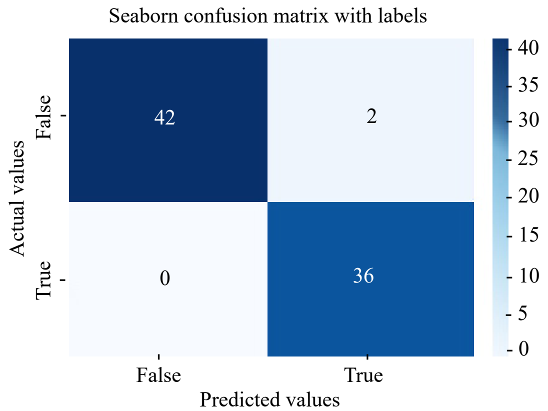

Figure 31. Confusion matrix for LR (Dataset 1)

Results from Dataset 2: The LR confusion matrix (Figure 32) shows one true positive class data point correctly detected by the algorithm. It accurately classified all 600 negative class data points as true negatives. False positive matches occurred when 14 negative class data points were misclassified as positive. Additionally, there were 117 positive class data items classified as false negatives with a value.

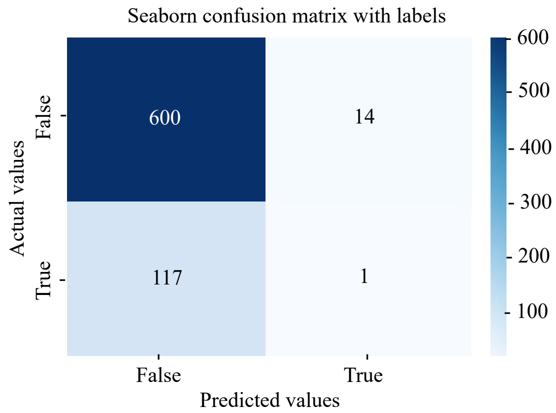

Figure 32. Confusion matrix for LR (Dataset 2)

4.5.7 *RF*

Results from Dataset 1: The RF confusion matrix (Figure 33) results indicate that the system accurately detected 27 genuine positive class data points and correctly identified all 34 data points in the negative class. However, there were 10 instances where the algorithm mistakenly categorized negative class data points as positive, resulting in false positive matches. Additionally, 9 positive class data points were incorrectly classified as negative, leading to false negatives.

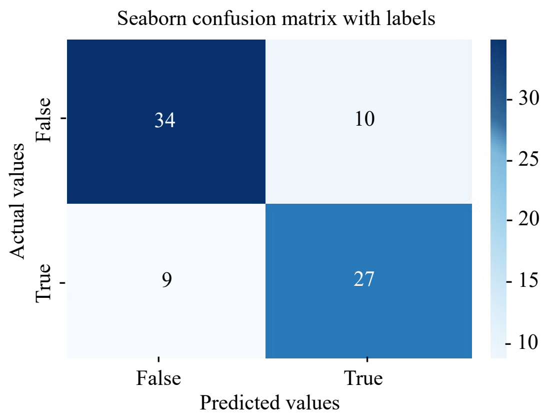

Figure 33. Confusion matrix for RF (Dataset 1)

Results from Dataset 2: In the RF confusion matrix (Figure 34), the algorithm correctly detected 9 genuine positive class data points and accurately identified all 581 negative class data points as true negatives. However, it mistakenly classified 42 negative class data points as positive class data points, resulting in false positive matches. Additionally, 100 positive class data points were incorrectly labeled as negative class data, leading to false negatives.

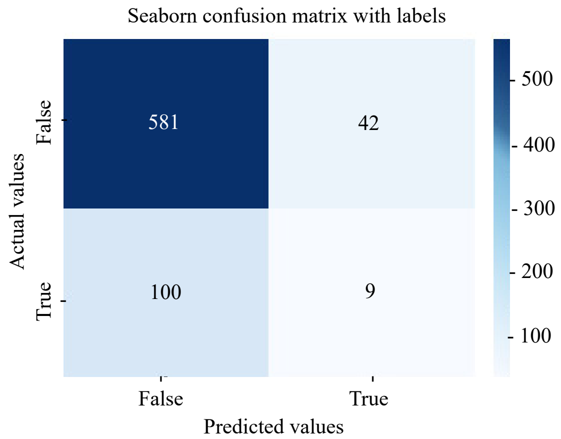

Figure 34. Confusion matrix for RF (Dataset 2)

4.5.8 *SVM*

Results from Dataset 1: In the SVM confusion matrix (Figure 35), there were 36 true positive class data points correctly detected by the system. The algorithm also accurately identified all 38 negative class data points as true negatives. However, it had 6 false positive matches, where negative class data points were wrongly categorized as the positive class. There were no false negatives, indicating no misclassification of positive or negative class data points.

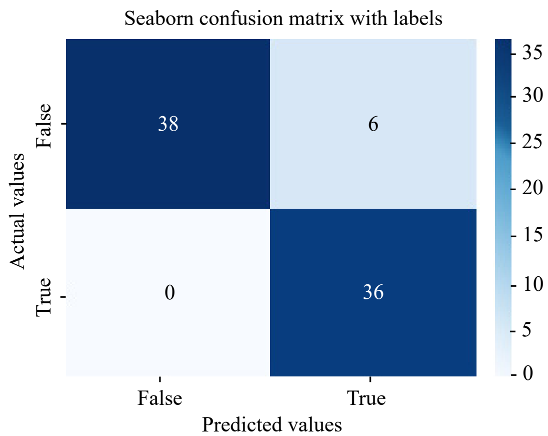

Figure 35. Confusion matrix for SVM (Dataset 1)

Results from Dataset 2: In the SVM confusion matrix (Figure 36), there were no true positive class data points detected by the algorithm. It correctly identified all 623 negative class data points as true negatives. There were no false positive matches, indicating the accurate classification of negative class data. However, there were 109 false negatives, where positive class data points were mistakenly classified as negative class.

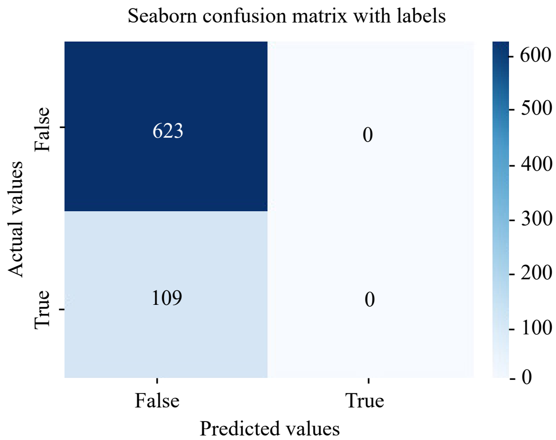

Figure 36. Confusion matrix for SVM (Dataset 2)

4.5.9 *NB*

Results from Dataset 1: In the NB confusion matrix (Figure 37), the algorithm correctly detected 36 true positive class data items. It accurately identified all 41 data points in the negative class as true negatives. There were three false positive matches, where negative class data points were mistakenly classified as the positive class. There were no false negatives with values, indicating the accurate classification of positive and negative class data points.

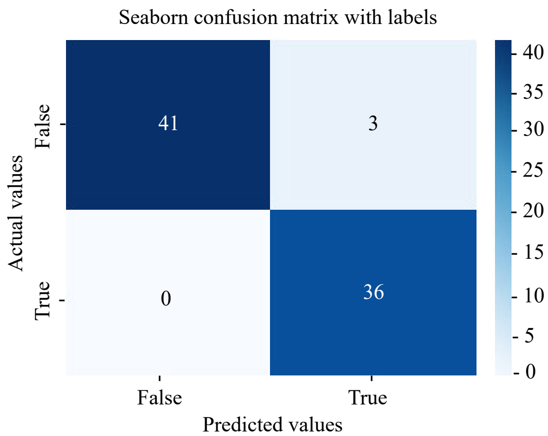

Figure 37. Confusion matrix for NB (Dataset 1)

Results from Dataset 2: The NB confusion matrix (Figure 38) indicates that the algorithm correctly detected 3 true positive class data items. It accurately identified all 609 data points in the negative class as true negatives. There were 10 false positive matches, where negative class data points were mistakenly classified as the positive class. Additionally, there were 110 instances of false negatives, where positive class data points were misclassified as negative class, and vice versa.

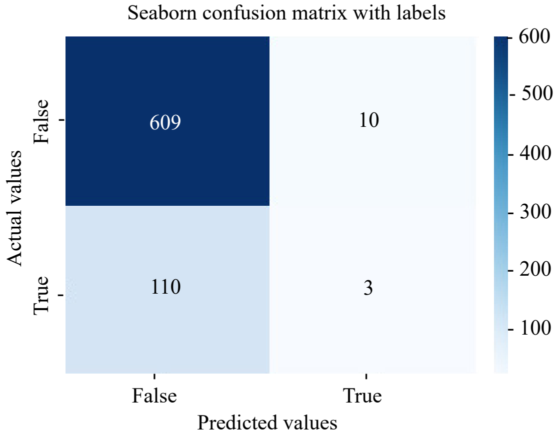

Figure 38. Confusion matrix for NB (Dataset 2)

The results of the feature selection method (information gain, Fisher's score, correlation coefficient) in both datasets are shown in Tables 5 to 10.

Table 5. Results of the information gain for Dataset 1

Classifier	Accuracy	Precision	Recall	F1 score
RF	85%	88%	84%	86%
SVM	93%	97%	90%	94%
NB	97%	97%	97%	97%
GBT	82%	96%	70%	81%
DT	62%	64%	72%	68%
LR	98%	97%	100%	98%

Table 6. Results of the information gain for Dataset 2

Classifier	Accuracy	Precision	Recall	F1 score
RF	82%	85%	95%	90%
SVM	85%	85%	100%	91%
NB	85%	85%	98%	91%
GBT	85%	86%	98%	91%
DT	76%	85%	87%	86%
LR	85%	85%	100%	91%

Table 7. Results of the Fisher's score for Dataset 1

Classifier	Accuracy	Precision	Recall	F1 score
RF	85%	88%	84%	86%
SVM	93%	97%	90%	94%
NB	97%	97%	97%	97%
GBT	82%	96%	70%	81%
DT	62%	64%	72%	68%
LR	98%	97%	100%	98%

Table 8. Results of the Fisher's score for Dataset 2

Classifier	Accuracy	Precision	Recall	F1 score
RF	83%	85%	96%	90%
SVM	85%	85%	100%	91%
NB	83%	85%	96%	90%
GBT	83%	85%	96%	90%
DT	74%	85%	84%	85%
LR	84%	85%	98%	91%

Table 9. Results of the correlation coefficient for Dataset 1

Classifier	Accuracy	Precision	Recall	F1 score
RF	76%	79%	77%	78%
SVM	92%	100%	86%	92%
NB	96%	100%	93%	96%
GBT	90%	97%	84%	90%
DT	63%	62%	84%	71%
LR	97%	100%	95%	97%

Table 10. Results of the correlation coefficient for Dataset 2

Classifier	Accuracy	Precision	Recall	F1 score
RF	80%	85%	93%	89%
SVM	85%	85%	100%	91%
NB	84%	85%	98%	91%
GBT	82%	85%	94%	89%
DT	84%	85%	99%	91%
LR	86%	86%	99%	92%

4.5.10 *Modeling and predicting with ML*

The main goal of the entire project is to test several classification systems in order to accurately predict the prevalence of heart disease and CKD. This section summarizes all study data, highlights the greatest performance in terms of accuracy metrics, and highlights a few algorithms that are often employed in classification approaches to address supervised learning difficulties. Determine whether the model adequately or insufficiently fits the data by comparing the accuracy of the training and test sets. The data is divided to run all algorithms, and models are then tested and trained in an 80:20 ratio.

4.5.11 *Finding the result*

A summary of the various accuracy percentages of the various approaches is provided below, where more complex algorithms like NB, GBT, DT, and LR produced better results than the earlier ones in dataset 1. NB uses a similar strategy to predict the likelihood of different classes based on numerous attributes. The posterior probability of class (c, target) given predictor (x, characteristics) is denoted in NB as P(c|x). The prior probability of the class is P(c). The likelihood, or P(x|c), of a predictor for a particular class is referred to. P(x) is the probability of the prior predictor. Based on ('ccp_alpha': 0.0, 'criterion': 'friedman_mse', 'init': None, 0.5 is the learning rate. 'loss': 'deviance', 'n_estimators': 100, the parameters ('random_state': None,'). Based on ('ccp_alpha': 0.0; 'class_weight': None; 'criterion': 'gini'; 'random_state': None, DT produces the best results. parameters ('splitter': 'best'). Based on ('C': 1.0, True, 'penalty': 'l2', 'random_state': None, LR produces the best results. settings ('solver': 'lbfgs', 'tol': 0.0001, 'verbose': 0). However, using the settings ('C': 1.0, 'penalty': 'l2', 'random state': 12345, and 'solver': 'lbfgs'), LR produces the best results for dataset 2. It is important to stress that careful parameter adjustment is frequently necessary to get reliable findings from these processes. Simpler techniques also proved their worth by yielding reasonable results. The future of deep learning and ML in medicine is quite promising. Imagine a location without any available experts on these ailments. We can anticipate whether a disease will manifest or not with a high degree of accuracy with just a little knowledge of a patient's medical history.

5. Conclusion

The study aimed to develop a reliable system for predicting cardiac diseases and chronic renal conditions associated with increased mortality rates. It analyzed two datasets related to CKD and heart attack using various hybrid ML techniques. Ensemble learning methods such as GBT classifier, SVM, DT, LR, NB, RF, and RF, along with the RNN deep learning algorithm, were employed to identify the most efficient algorithm for disease diagnosis. Feature selection methods and data preprocessing techniques were used to enhance the dataset quality. Grid search optimization was utilized to fine-tune the ML parameters. The models' performance was evaluated using accuracy, precision, recall, and F1 measure metrics. The study aimed to create an accurate and efficient approach for predicting CKDs and cardiac disorders by analyzing the effectiveness of different algorithms and strategies.

6. Future work

In future efforts, three important suggestions are proposed. First, the system can be enhanced by developing a tool that assesses a patient's risk of chronic renal illness and heart attack based on general symptoms and medical history. Second, extensive analysis of data can be achieved by utilizing various data mining techniques and tree approaches like time series, clustering, and association rules. Lastly, employing text mining to extract valuable information from unstructured healthcare data holds potential for further improvement.

Conflict of interest

The authors have no conflicts of interest to disclose.

References

[1] Gárate-Escamila AK, Hajjam El Hassani A, Andrès E. Classification models for heart disease prediction using feature selection and PCA. *Informatics in Medicine Unlocked*. 2020; 1(9): 100330. https://doi.org/10.1016/j.imu.2020.100330

[2] 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. *Computational Intelligence and Neuroscience*. 2021; 2021: 8387680. https://doi.org/10.1155/2021/8387680

[3] Khourdifi Y, Bahaj M. Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. *International Journal of Intelligent Engineering and Systems*. 2019; 12(1): 242-252. https://doi.org/10.22266/ijies2019.0228.24

[4] Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*. 2019; 7: 81542-81554. https://doi.org/10.1109/ACCESS.2019.2923707

[5] Ramprakash P, Sarumathi R, Mowriya R, Nithyavishnupriya S. Heart disease prediction using deep neural network. In: *2020 International Conference on Inventive Computation Technologies (ICICT)*. Coimbatore, India: IEEE; 2020. p.666-670. https://doi.org/10.1109/ICICT48043.2020.9112443

[6] Garg A, Sharma B, Khan R. Heart disease prediction using machine learning techniques. *IOP Conference Series: Materials Science and Engineering*. 2021; 1022: 012046. https://doi.org/10.1088/1757-899X/1022/1/012046

[7] Ali MM, Paul BK, Ahmed K, Bui FM, Quinn JMW, Moni MA. Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. *Computers in Biology and Medicine*. 2021; 136: 104672. https://doi.org/10.1016/j.compbiomed.2021.104672

[8] Abdel-Fattah MA, Othman NA, Goher N. Predicting chronic kidney disease using hybrid machine learning based on apache spark. *Computational Intelligence and Neuroscience*. 2022; 2022: 9898831. https://doi.org/10.1155/2022/9898831

[9] Karthick K, Aruna SK, Samikannu R, Kuppusamy R, Teekaraman Y, Thelkar AR. Implementation of a heart disease risk prediction model using machine learning. *Computational and Mathematical Methods in Medicine*. 2022; 2022: 1-14. https://doi.org/10.1155/2022/6517716

[10] J. SK, S. G. Prediction of heart disease using machine learning algorithms. In: *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*. Chennai, India: IEEE; 2019. p.1-5. https://doi.org/10.1109/ICIICT1.2019.8741465

[11] Obasi T, Omair Shafiq M. Towards comparing and using machine learning techniques for detecting and predicting heart attack and diseases. In: *2019 IEEE International Conference on Big Data (Big Data)*. Los Angeles, CA, USA: IEEE; 2019. p.2393-3402. https://doi.org/10.1109/BigData47090.2019.9005488

[12] Senan EM, Al-Adhaileh MH, Alsaade FW, Aldhyani THH, Alqarni AA, Alsharif N, et al. Diagnosis of chronic kidney disease using effective classification algorithms and recursive feature elimination techniques. *Journal of Healthcare Engineering*. 2021; 2021: 1004767. https://doi.org/10.1155/2021/1004767

[13] Imran Ali S, Ali B, Hussain J, Hussain M, Satti FA, Park GH, et al. Cost-sensitive ensemble feature ranking and automatic threshold selection for chronic kidney disease diagnosis. *Applied Sciences*. 2020; 10(16): 5663. https://doi.org/10.3390/app10165663

[14] Qin J, Chen L, Liu Y, Liu C, Feng C, Chen B. A machine learning methodology for diagnosing chronic kidney disease. *IEEE Access*. 2020; 8: 20991-21002. https://doi.org/10.1109/ACCESS.2019.2963053

[15] Almansour NA, Syed HF, Khayat NR, Altheeb RK, Juri RE, Alhiyafi J, et al. Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study. *Computers in Biology and Medicine*. 2019; 109: 101-111. https://doi.org/10.1016/j.compbiomed.2019.04.017

[16] Boukhatem C, Youssef HY, Nassif AB. Heart disease prediction using machine learning. In: *2022 Advances in Science and Engineering Technology International Conferences (ASET)*. Dubai, United Arab Emirates: IEEE; 2022. p.1-6. https://doi.org/10.1109/ASET53988.2022.9734880

[17] Wibawa MS, Maysanjaya IMD, Putra IMAW. Boosted classifier and features selection for enhancing chronic kidney disease diagnose. In: *2017 5th International Conference on Cyber and IT Service Management (CITSM)*. Denpasar, Indonesia: IEEE; 2017. p.1-6. https://doi.org/10.1109/CITSM.2017.8089245

[18] Nishat M, Faisal F, Dip R, Nasrullah S, Ahsan R, Shikder F, et al. A comprehensive analysis on detecting chronic

kidney disease by employing machine learning algorithms. *EAI Endorsed Transactions on Pervasive Health and Technology*. 2018; 7: 170671. https://doi.org/10.4108/eai.13-8-2021.170671

[19] Almasoud M, Ward TE. Detection of chronic kidney disease using machine learning algorithms with least number of predictors. *International Journal of Advanced Computer Science and Applications*. 2019; 10(8): 89-96. https://doi.org/10.14569/IJACSA.2019.0100813

[20] Poonia RC, Gupta MK, Abunadi I, Albraikan AA, Al-Wesabi FN, Hamza MA, et al. Intelligent diagnostic prediction and classification models for detection of kidney disease. *Healthcare*. 2022; 10(2): 371. https://doi.org/10.3390/healthcare10020371

[21] Debal DA, Sitote TM. Chronic kidney disease prediction using machine learning techniques. *Journal of Big Data*. 2022; 9(1): 109. https://doi.org/10.1186/s40537-022-00657-5

[22] De Guia JD, Concepcion RS, Bandala AA, Dadios EP. Performance comparison of classification algorithms for diagnosing chronic kidney disease. In: *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*. Laoag, Philippines: IEEE ; 2019. p.1-7. https://doi.org/10.1109/HNICEM48295.2019.9073568

[23] Dritsas E, Trigka M. Machine learning techniques for chronic kidney disease risk prediction. *Big Data and Cognitive Computing*. 2022; 6(3): 98. https://doi.org/10.3390/bdcc6030098

[24] Almustafa KM. Prediction of chronic kidney disease using different classification algorithms. *Informatics in Medicine Unlocked*. 2021; 24: 100631. https://doi.org/10.1016/j.imu.2021.100631

[25] Menshawi A, Hassan MM, Allheeib N, Fortino G. A hybrid generic framework for heart problem diagnosis based on a machine learning paradigm. *Sensors*. 2023; 23(3): 1392. https://doi.org/10.3390/s23031392

[26] Haq AU, Li JP, Memon MH, Nazir S, Sun R. A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mobile Information Systems*. 2018; 2018: 3860146. https://doi.org/10.1155/2018/3860146

[27] Ghosh P, Azam S, Karim A, Hassan M, Roy K, Jonkman M. A comparative study of different machine learning tools in detecting diabetes. *Procedia Computer Science*. 2021; 192: 467-477. https://doi.org/10.1016/j.procs.2021.08.048

[28] Ma F, Sun T, Liu L, Jing H. Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network. *Future Generation Computer Systems*. 2020; 111: 17-26. https://doi.org/10.1016/j.future.2020.04.036

[29] Vasquez-Morales GR, Martinez-Monterrubio SM, Moreno-Ger P, Recio-Garcia JA. Explainable prediction of chronic renal disease in the Colombian population using neural networks and case-based reasoning. *IEEE Access*. 2019; 7: 152900-152910. https://doi.org/10.1109/ACCESS.2019.2948430

[30] Boukenze B, Haqiq A, Mousannif H. Predicting chronic kidney failure disease using data mining techniques. In: El-Azouzi R, Menasche DS, Sabir E, De Pellegrini F, Benjillali M. (eds.) *Advances in Ubiquitous Networking 2. UNet 2016*. Lecture Notes in Electrical Engineering, vol 397. Singapore: Springer; 2016. p.701-712. https://doi.org/10.1007/978-981-10-1627-1_55