Intelligent System for the Prediction of Heart Diseases Using Machine Learning Algorithms with Anew Mixed Feature Creation (MFC) technique

Rawia Elarabi[1], Abdelrahman Elsharif Karrar[2*], Murtada El-mukashfi El-taher[3]
relarabi@jazanu.edu.sa
Department of Computer Science, Jazan University, Jazan, Saudi Arabia[1]
College of Computer Science and Engineering, Taibah University, Medina, Saudi Arabia[2]
Department of Computer Science, Faculty of Science, University of Bakht Alruda, Elduwiam, Sudan[1,3]

Abstract
Classification systems can significantly assist the medical sector by allowing for the precise and quick diagnosis of diseases. As a result, both doctors and patients will save time. A possible way for identifying risk variables is to use machine learning algorithms. Non-surgical technologies, such as machine learning, are trustworthy and effective in categorizing healthy and heart-disease patients, and they save time and effort. The goal of this study is to create a medical intelligent decision support system based on machine learning for the diagnosis of heart disease. We have used a mixed feature creation (MFC) technique to generate new features from the UCI Cleveland Cardiology dataset. We select the most suitable features by using Least Absolute Shrinkage and Selection Operator (LASSO), Recursive Feature Elimination with Random Forest feature selection (RFE-RF) and the best features of both LASSO RFE-RF (BLR) techniques. Cross-validated and grid-search methods are used to optimize the parameters of the estimator used in applying these algorithms. and classifier performance assessment metrics including classification accuracy, specificity, sensitivity, precision, and F1-Score, of each classification model, along with execution time and RMSE the results are presented independently for comparison. Our proposed work finds the best potential outcome across all available prediction models and improves the system's performance, allowing physicians to diagnose heart patients more accurately.

Keywords:
machine learning; Classification; random forest; SVM; feature selection; cardiovascular; heart disease; LASSO feature selection; RFE-RF feature selection; K-nearest neighbors; AdaBoost

1. Introduction

Data mining (DM) is the process of extracting usable information from big data sets and employing strategies like classification, clustering, and association to forecast or describe the data. In the machine learning (ML) process, classification can be described as a supervised learning algorithm. It uses previous knowledge of the class to which the data records belong to give class labels to data objects. A set of data records is separated into training and test data sets in classification. The classification framework is designed using the training data set, and the model is validated using the test data record. After that, the model is used to categorize and forecast a fresh collection of data records that is distinct from both the training and test data sets. [1]

Because it has prior knowledge of the class labels of data records, supervised learning algorithms (like classification) are preferred over unsupervised learning algorithms (like clustering) since they make feature selection simple and lead to improved prediction of classification accuracy. [2][3].

Today Hospitals are becoming more numerous, patients are becoming much more numerous, and data is becoming increasingly abundant. A Hospital Information System is used by most hospitals to maintain track of patient information and health treatment. This data, on the other hand, is rarely used to make decisions. The suggested machine-learning-based decision support system will assist clinicians in diagnosing patients more efficiently if we process this data using Artificial Intelligence and Machine Learning techniques. Data mining algorithms have been used to research diabetes, asthma, cardiovascular disease, AIDS, and other disorders. Various data mining approaches, such as naïve Bayesian classification, artificial neural networks, support vector machines, decision trees, logistic regression, Ensembles, and others, have been used in healthcare research to develop models, classifiers, and hybrid models [4] [5] [6][7].

Heart disease has long been considered the most serious and deadly disease. Heart disease is growing more widespread, with a high death rate, providing a significant risk and burden to global healthcare systems. In impoverished nations, diagnosing and treating heart disease is extremely difficult. [8] Patient prediction and treatment are affected by a shortage of diagnostic technology, as well as a scarcity of doctors and other resources. [9]. Heart disease is responsible for one-third of all deaths worldwide, according to WHO statistics. According to the European Cardiology Society (ESC), heart disease affects 26 million individuals worldwide, with an additional 3.6 million people diagnosed each year. Half of all patients diagnosed with heart disease die within 1-2 years and treating heart

Manuscript received May 5, 2023
Manuscript revised May 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.5.17

disease accounts for around 3% of total health expenditure [10].

When employing invasive methods to diagnose cardiac illness, medical practitioners look at the patient's medical history, physical examination report, and analysis of concerned symptoms. Because of human error, all of these processes result in incorrect diagnoses and, in many cases, delays in diagnosing findings. It's also expensive and computationally demanding, and evaluating it takes time. [11]. Recent research on heart disease in adults and children has emphasized the importance of lowering heart disease-related mortality. Datasets on heart disease are publicly available for comparing prediction models. Researchers can use machine learning and artificial intelligence to construct the optimal prediction models possible by utilizing the massive databases available. [1], [7]. Because the existing clinical datasets are unreliable and duplicated, preprocessing is essential. It's crucial to pick the right key qualities to use as risk variables in prediction models. To build successful prediction models, it's important to pick the right combination of features and machine learning algorithm's [5], [12], [13],[14]. Since this algorithm rely on the consistency of the training and test data, data mining, Relief selection, FER-RF, and LASSO can help prepare the data so that a greater accurate prediction can be made. [15], [16].

The contribution of the proposed research is to develop a medical intelligent decision support system based on machine learning for the diagnosis of heart disease. We have used a mixed feature creation (MFC) technique to extract and derive new features from the UCI Cleveland Cardiology dataset. Three selection techniques. LASSO, RFE-RF, and best features from LASSO, RFE-RF (BLR), to choose the most strongly correlated features that have a significant impact on the predicted value of the target, this also aids in the resolution of machine learning overfitting and underfitting issues. In this study, various supervised models, such as

- K-Nearest Neighbors (KNN)
- Support Vector Machines(SVM)
- Stochastic Gradient Descent (SGD)
- Decision Tree Classifier(DT)
- Random Forest Classifier (RF)
- AdaBoost Classifier (AB)
- Logistic regression (LR)
- Gradient Boosting(GB)

Methods of Cross validation and grid search over a parameter grid is used to optimize the parameters of the estimator used to implement these algorithms. and classifier performance evaluation metrics such as classification accuracy, specificity, sensitivity, precision, and F1 Score, of our model, together with execution time and RMSE the results are presented separately for comparison.

The following are the steps that must be completed:

1. All classifiers' results were evaluated on full features, including accuracy, specificity, sensitivity, precision, and F1-Score, of each classification model, along with execution time and RMSE

2. Three selection techniques. LASSO, RFE-RF and best features from LASSO, RFE-RF combined (BLR),to choose the most strongly correlated features that provide a substantial effect on the predicted outcome of the target, This also aids in the resolution of machine learning overfitting and underfitting issues

3. The study evaluates which algorithm and classifier are suitable for constructing a powerful intelligent system for heart disease that can effectively discriminate between those who have heart disease and those who are normal.

The structure of this article is as follows. Section 2 presents the related work. Section 3 contains the description of the dataset and methods used in this study. Section 4 presents the results and discussion. The conclusion and potential future work are presented in section 5.

II. LITERATURE REVIEW

There are related works that use classification techniques to diagnose various diseases and phenomena, such as Automated Diagnosis of Thalassemia, Diabetes, Heart Disease, Breast Cancer, Liver Disease, and Hepatitis, among others. Research in this area is significant because of its ability to develop and select models with the best accuracy and efficacy. Over the past two decades, artificial intelligence and data mining techniques have been used to predict heart disease. There is a lot of research that has applied artificial intelligence and classification techniques using a variety of patient databases worldwide. Hybrid approaches that combine various machine learning algorithms with information systems offer promise for diagnosing diseases. [17], [18]. in [19] Researchers applied a ML-based ensemble approach to improve the accuracy of prediction because of hybrid modelling experiments, they used bagging and boosting algorithms to use the majority voting of C 4.5, Multilayer Perceptron, Naive Bayes, Bayes Net, Random Forest (RF), and PART classifiers. The designed model had an accuracy of 85.48 percent. There was an improvement in accuracy of weak classifiers, and the accuracy for predicting heart disease risk was acceptable.

In [10] The authors used seven popular classification algorithms with Python to develop a ML-based hybrid intelligent system framework for predicting heart disease patients in their paper. Among them are KNN, ANN, DT, SVM, NB, LR, and MLP. To train and test the model, 303 Cleveland instances with 76 features were used the researchers applied a 10-fold cross validation procedure on the data. Selecting features by algorithms, including Relief,

Minimal-Maximal-Relevance (mRMR) and Least Absolute Shrinkage and Selection Operator (LASSO) were applied to select optimal heart disease correlated features. A pre-processing step removed the records with large missing values from the data. The data collection was reduced to 297 records with only 14 features. After applying the feature selection algorithms, the features were reduced to 6 only related to heart disease. To find the best performing classifier, they tested each with different feature selection algorithms. They found that SVM with LASSO feature selection algorithm provided the best results, as compared with other feature selection algorithms and classifiers. Limiting the heart disease features to only 6 results in insufficient classification accuracy, since more relevant characteristics are not included.

In [20] They compared different classification algorithms for the prediction of heart disease using the correlation coefficient methodology. All four algorithms were implemented and tested on the Cleveland Heart Disease dataset. The accuracy of various classification methods varies. When we use a KNN algorithm with a correlation factor of nearly 94%, we get the best accuracy.

In algorithms such as in [21] The accuracy of SVM, neural networks, Bayesian classification, decision trees, and logistic regression were evaluated. The SVM model of 102 cases had the highest accuracy rate (90.5 percent), followed by the neural network model at 88.9%, the Bayesian algorithm at 82.2 %, the decision tree algorithm at 77.9%, and the logistic regression algorithm at 73.9 %. [22] [10] demonstrated a machine learning-based algorithm to detect heart disease using sampling methods that dealt with unbalanced data. A variety of sampling methods are employed, including Random Over-Sampling, Synthetic Minority Over-Sampling (SMOTE), and Adaptive Synthetic Sampling (ADASYN). For the algorithm training and testing, the Framingham datasets from the Kaggle website were used, which contain 4239 instances with 15 features. Based on the features, the aim was to predict whether a patient had a 10-year risk of future coronary heart disease. The machine learning techniques used include LR, KNN, AdaBoost, DT, NB, and RF. Precision, recall, and accuracy were used to measure and evaluate the performance of these classification models. Depending on how the samples were collected, each parameter varied. 99% accuracy was achieved with the SVM classifier paired with the Random Over-sampling technique. However, RF had 91.3% accuracy with the SMOTE model, while the DT and RF had 90.3% accuracy with the ADASYN model. Hence, this approach's classification accuracy is solely based on sampling techniques, which are not always necessary in all datasets.

[23] It is based on a dataset of 1159 healthy individuals, 405 negative angiography results, and 782 positive results. The model can determine risk factors with an accuracy of 94% by using 10 variables out of 12. In [24] A comparative study on heart disease classification and prediction was conducted using ML techniques. Rapid-Miner uses NB, DT, RF, SVM, and LR algorithms. A dataset of 303 cases and 14 attributes from the UCI machine learning repository was used. To test the model, the 10-fold cross validation procedure was used. Based on the results of the experiments, the DT model scored the greatest in heart disease detection accuracy, then SVM at 93.19 and 92.30, correspondingly. In [25]A framework was presented for addressing privacy issues and overcoming problems inherent in limited medical research datasets by using synthetic data. The researchers investigated the use of surrogate datasets including synthetic observations to model the system. Based on the preliminary observations, the data were generated and compared with the results of LR, DT, and RF. Cleveland heart disease data used in this study consists of 303 instances with 76 features, which were pre-processed to become 279 instances with 14 features. The experiment was divided into three stages. In the first stage, baseline models were generated, and their results were compared to those in previous studies to validate and validate the accuracy and stability of the proposed models. The similar original data Cleveland: 279 records and 14 attributes was utilized in the second stage to produce 50,000 records, which were then used to train and test the existing LR, DT and RF algorithms. Using the same data set, 60,000 records were generated and used to train and test the ANN model of the forward and backward propagation algorithm type. Using 10-fold cross validation with the traditional models (LR, DT, and RF), they achieved an improvement in prediction stability within 2% variance at around 81%. Using the ANN with surrogate data, they improved the accuracy of heart disease prediction by nearly 16% to 96.7% while maintaining 1% accuracy. According to [15] A study was conducted using ML techniques (K-NN, DT, NB, LR, SVM, Neural Network, and a hybrid of voting with NB and LR) to identify significant risk indicators, and a comparative analysis was conducted. According to their findings, the hybrid model's accuracy was 87.41% when combined with the selected attributes.

The methodology presented by [12] used the mean Fisher score feature selection algorithm (MFSFSA) in conjunction with SVM classification. The MCC was calculated using a validation method using an SVM to obtain the targeted feature subset. The features were selected based on a Fisher score that was higher than normal. Combining MFSFSA and SVM yielded an accuracy of 81.19%, a sensitivity of 72.92%, and an accuracy of 88.68%. In [25] Using enhanced bagging and weighing, we offer an ensemble technique with multilayer categorization. By leveraging a seven heterogeneous classifier ensemble, the suggested model, known as HMBagMoov, can overcome the disadvantages of typical performance bottlenecks. in [26] This experiment was conducted on the Cleveland dataset 303. A decision tree was performed at 75.55% accuracy. in. [27] employed the same method using a boosted hybrid technique, achieving a 75.9% accuracy rate. The performance of the boosting ensemble technique was also evaluated using the UCI laboratory dataset. [28] established a rule-based

classification system for cardiac disease prediction that was 86.7% accurate. In [29] created an electronic medical record probabilistic principal component analysis (PPCA) based on the principal components of probabilistic medicine. The primary function of the PPCA is to identify the most significant predictors of heart disease. [30] provided a heart disease classification approach that included the PSO with SVM, with a classification accuracy of 84.36 percent.

In a recent paper [31] developed a machine intelligence framework using factor analysis of mixed data (FAMD) and RF-based MLA]. By employing the FAMD to discover the appropriate traits, RF was utilized to forecast illness. The proposed approach had a precision of 93.44 percent, a sensitivity of 89.28 percent, and a specificity of 96.96 percent. In the research work in [32]. They proposed a heart disease prediction framework using a set of five datasets containing 1190 cases associated across 14 variables, which were used for testing and training the models. Two feature selection algorithms, which include Relief and Least Absolute Shrinkage and Selection Operator (LASSO) techniques, Applying the feature selection algorithms reduces the features. With LASSO, there are 11 features, and with Relief, there are 10. Develop novel hybrid classifiers such as the Decision Tree Bagging Method (DTBM), Random Forest Bagging Method (RFBM), K-Nearest Neighbors Bagging Method (KNNBM), AdaBoost Boosting Method (ABBM), and Gradient Boosting Boosting Method (GBBM) that combine classical classifiers with bagging and boosting techniques. Each of the classifiers was tested using different performance metrics together with different feature selection algorithms to find the most efficient one. This hybrid classifier, Random Forest Bagging with Relief features, achieved 99.05% accuracy.

Based on the extensive literature review, it can be concluded that most of the studies used the Cleveland Heart Disease Dataset, which contains only 303 cases with 14 features, which is a small and limited number of features. In some studies that used other data sources, a data set with limited cardiac features was also used. Therefore, it is not possible to generalize the different classification accuracy obtained in predicting heart disease. Since the CHD Cleveland dataset is a mixed dataset with both numeric and categorical variables, mixed feature generation (MFC) is a convenient feature extraction technique to derive new features and label coding for categorical features. Through the selection techniques used to extract the best features, the extracted and derived features are analysed. The selection techniques used in this research included LASSO and RFE-RF and the best of both LASSO and RFE-RF (BLR). This helps extract the best features. With more features to develop more efficient machine learning models—and this is the primary goal of our research—more efficient classification and early prediction of heart disease will be achieved, thereby reducing CVD morbidity and mortality.

A. THE ANALYSIS AND STATISTICAL OF THE AVAILABLE FEATURES OF CLEVELAND DATASET.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
count	302	302	302	302	302	302	302	302	302	302	302	302	302	302
mean	54.42	0.68	0.96	131.6	246.5	0.15	0.53	149.57	0.33	1.04	1.4	0.72	2.31	0.54
std	9.05	0.47	1.03	17.56	51.75	0.36	0.53	22.9	0.47	1.16	0.62	1.01	0.61	0.5
min	29	0	0	94	126	0	0	71	0	0	0	0	0	0
25%	48	0	0	120	211	0	0	133.25	0	0	1	0	2	0
50%	55.5	1	1	130	240.5	0	1	152.5	0	0.8	1	0	2	1
75%	61	1	2	140	274.75	0	1	166	1	1.6	2	1	3	1
max	77	1	3	200	564	1	2	202	1	6.2	2	4	3	1

III. METHODOLOGY

A. Dataset Specification and Data Description

To acquire the best results from ML algorithms, the first and most critical component is data. Therefore, it must be displayed, analysed, and its characteristics and qualities accurately understood. The Cleveland dataset from the University of California (UCI) is the most commonly used benchmark for CHD prediction. It can be obtained through the "UCI machine learning repository," which is one of the most well-known data repositories[38] . The dataset was utilized in this study to develop a machine learning based heart disease diagnosis framework. A set of 14 attributes is commonly used for research purposes.13 attributes are used as diagnosis inputs, while the "target" attribute is used as an output. For the "target" output label, there are two types:

heart patients and normal-

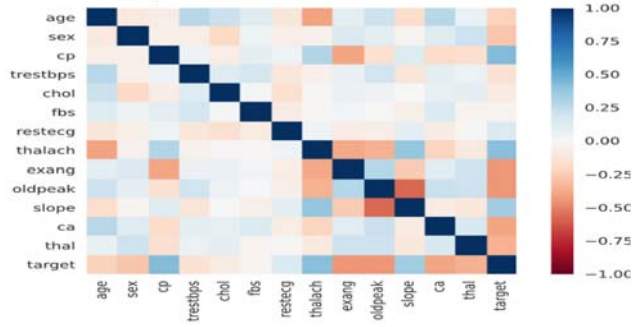


Fig.1 Pearson Correlation Coefficient Heat Map of Cleveland CHD Features.

development adults. During the analysis of the dataset, we found one duplicate row and no missing values. Table 1 shows the analysis and statistics for the Cleveland dataset's available features. Table 2. Statistical overview of the features of the Cleveland dataset The correlation coefficient of the features of the Cleveland CHD dataset.

A Cleveland CHD dataset includes information on medical history as well as clinical characteristics. These features appear to be dependent on each other. The heatmap in Fig. 1 shows the correlations within the features in the data set. To calculate the heatmap, we used Pearson's correlation coefficient. Correlation between features is represented by a scaled heatmap, , where (−1.00 and 1.00) represent the negatively correlated (red colour) and positively correlated (blue colour) features, respectively.

B. Feature Engineering

1. Mixed feature creation

This dataset CHD contains both categorical and numeric features, in statistics, categories are qualitative, while numerical features are quantitative. Mixed feature creation (MFC) This is a factorial approach that considers both types of features, First, derivation of new numerical features using Factor Analysis. Secondly, concatenated the categorical and the new numerical features, respectively. Among the benefits of this technology, categorical characteristics are transformed into a disjunctive representation of crisp coding and scaled using the selective format, while numerical variables are scaled down to unit variance. The pseudo-code of MFC is shown as Algorithm1. Our original data set contained 14 individual traits after using a Mixed feature creation MFC method that extracted new features bringing the number of features to 61 [39] [40].

TABLE 2 FEATURES DESCRIPTION AND VALUE RANGE IN DATASET.

No.	feature code	Description	Data Type	Value Range	Missing
1	age	age in years	Integer	29 to 77	No
2	sex	Female=0	Integer	0 and 1	No
3	cp	Type of chest pain 0 = asymptomatic 1 = typical angina 2 = atypical angina 3 = non-anginal pain	Integer	0,1,2,and 3	No
4	trestbps	resting blood pressure (in mm Hg on admission to the hospital)	Integer	94–200	No
5	chol	serum cholestoral in mg/dl	Integer	126 - 564	No
6	fbs	fasting blood sugar > 120 mg/dl (1 = true; 0 = false)	Integer	0,1	No
7	restecg	restingelectrocardiograp results (0 = hypertrophy = normal; 2 = having ST-wave abnormality)	Integer	0,1 and 2	No
8	thalach	maximum heart rate achieved	Integer	71-202	Np
9	exang	exercise induced angina (1 = yes; 0 = no)	Integer	0,1	No
10	oldpeak	ST depression induced by exercise relative to rest	Real	0-6.2	No
11	slope	slope of the peak exercise ST segment (0 = downsloping; 1 = flat; 2 = upsloping)	Integer	0,1,and 2	No
12	ca	number of major vessels (0-3) colored by fluoroscopy	Integer	0,1,2,and 3	No
13	thal	Thallium scan(1 = fixed defect; 2 = normal; 3 = reversible defect	Integer	1,2,and 3	No
14	target	Diagnosis of heart disease	Integer	0,1	NO

Algorithm 1 MFC

begin
$C_i \leftarrow$ categorical feature and $i \leftarrow$ 1,2,...,n
$N_j \leftarrow$numerical feature and $p \leftarrow$ 1,2,...m
$P_j \leftarrow$ Factor analysis of N_j
for j =1 to m do
Q_j= new numerical feature N_j/P_j
end for
for R in list of categorical features $[C_i]$ do
for J in list of new numerical feature $[Q_j]$do
MFC=$[C_i]$+ $[Q_j]$
end for
end for
return mixed feature creation (MFC)
end

2. *Preprocessing and Cleaning Data*

Dataset preprocessing is an important step before machine learning algorithms can be applied, in which the dataset is examined to detect and avoid duplicate or missing values. In the case of real-time data, there are methods to preprocess or normalize the data. To reduce the unbalancing effect of illness classes, we employed the following methods: Label encoding for categorical features, data standardization using z-score normalization to normalize the features. Standardization converts the data to a mean of 0(μ)and a standard deviation (∑) of 1. The conversion formula of (8) is given below [41]:

Standardization, X=(X-μ)/σ

C. *Methodology of the Proposed System*

The proposed system has been developed with the aim of building an intelligent machine learning system for heart disease diagnosis. It is to optimize the system for predicting heart disease to increase the survival rate of patients by accurate, precise and early detection of the disease. This is included in the framework. Fig. 2 illustrates the workflow. Dataset preprocessing is an important step in which the dataset is examined to detect and avoid duplicate or missing values. Because the CHD Cleveland dataset is a mixed-type dataset with both numeric and categorical variables, mixed feature creation (MFC) is an appropriate technique for extracting features for the derivation of new features, and label encoding for categorical features. Through the selection techniques used to extract the best features, the extracted and derived features are analyzed. The selection techniques used in this research included: LASSO and RFE-RF and the best of both LASSO & RFE-RF(BLR). This helps in extracting the best features.

The performance of classifiers with the features identified by these techniques is examined, in addition to the original features. After the feature is identified, the dataset is divided into two parts: training and testing. Based on typical learning rates, 80% of the data is assigned to the training phase, and the remaining 20% is used for the testing phase. All ensemble models and traditional classifiers were applied in the system. The model's validation and different performance evaluation metrics were obtained, comparing all algorithms applied in the system.

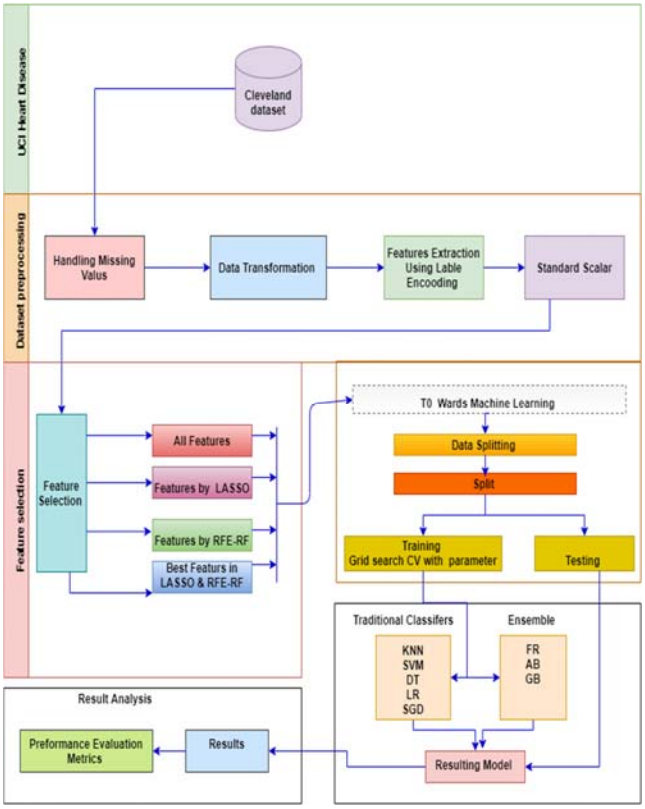


Fig 2. Flow Diagram of Proposed Model.

D. *Evaluation Metrics*

To evaluate the performance of the proposed framework, the effectiveness and accuracy of the machine learning method can be evaluated using performance indicators. The confusion matrix is computed, which gives an idea of the machine learning approach's learning extent and ability for accurate classification. The prime components of the confusion matrix are

True Positive (TP)=when the model is correctly Identified as having heart disease (HD).

True Negative (TN)= (when the model correctly identified the opposite class, such as patients truly having no heart issues).

False Positive (FP)= (when the model incorrectly identifies HD patients, i.e., identifying non-HD patients as HD patients).

False Negative (FN)= (when the model incorrectly identifies the opposite class, such as HD patients, as normal patients).)

$$\text{classification accuracy (Acc)}=\frac{(TP+TN)}{(+TN\,TP+FP+FN)}\text{x } 100\% \quad (1)$$

$$\text{Precision}=((TP))/((TP+FP))\text{x } 100\% \quad (2)$$

$$\text{Recall or Sensitivity (Sen)}=\frac{(TP)}{(TP+FN)}\text{x } 100\% \quad (3)$$

F1-score = $\frac{2(Precision \times Recall)}{(Precision + Recall)}$x100% (4)

Specificity of true Negative Rate TNR=$\frac{(TN)}{(TN + FP)}$x100% (5)

ROC and AUC: The performance metric AUC - ROC curve is used to represent the classification using the curve area. Receiver operating characteristics (ROC) represent a probability curve and the area under the curve (AUC) represents the degree of separability of the classifier to accurately classify instances between classes. The higher the value of AUC, the better the model's ability to distinguish between heart patients and normal subjects.

E. Feature Selection Techniques

Because irrelevant features might impair the classification performance of a machine learning classifier, feature selection is required for the machine learning process. Using feature selection strategies, you may boost classification accuracy while also cutting down on model execution time. To pick features for our system, we have selected two algorithms: Recursive Feature Elimination with Random Forest feature selection (RFE-RF) and the Least Absolute Shrinkage and Selection Operator (LASSO) and the best of both LASSO and RFE-RF.

1) Least absolute shrinkage and selection operator(LASSO)

Least absolute shrinkage and selection operator (LASSO) selection features are based on modifying the absolute value of the feature coefficient. Some feature coefficient values become zero, and these zero coefficient features are removed from the feature subset. With low coefficient feature values, the LASSO performs admirably. Feature subsets with high coefficient values will be included in selected feature subsets. Some irrelevant features may be selected, and a subset of selected features may be included in LASSO [42]. Moreover, the reliability of this feature can be enhanced by repeating the above procedure many times, eventually taking the most frequently found features as the most important ones. This is called the randomized LASSO feature, which was introduced by [43].

2) Recursive Feature Elimination with Random Forest feature selection (RFE-RF)

Recursive Feature Elimination with Random Forest feature selection RFE is based on the idea of building a machine learning model (for example, an SVM or RF model) and selecting the best or worst performing feature based on coefficients, then setting the feature aside and continuing the process with the remaining features. This process is repeated until the dataset's features have been exhausted. The features are then graded based on when they were removed. AS a result, identifying the best performing subset of characteristics is a greedy optimization[44]. The type of model utilized for feature ranking in each iteration has a big impact on RFE's stability.[45].

3) Best features from LASSO, RFE-RF combined (BLR)

The process of extracting a set combines the most important features of the two feature selection methods, Random Forest feature selection (RFE-RF) and the Least Absolute Shrinkage and Selection Operator (LASSO).

F. Machine learning Classifier

This section explains the machine learning methodologies utilized in this study to create an intelligent heart disease prediction system.

1) Logistic Regression (LG)

A logistic regression is a classification algorithm [46] [47] [47]. For binary classification or multi- classification problems, to predict the value of predictive variable y when y $\in$ [0, 1] In a binary classification, 0 represents the negative class and 1 represents the positive class.in multi-classification to predict the value of y when y $\in$ [0, 1, 2, 3]. To classify binary classes 0 and 1, a hypothesis hypothesis $h(\theta) = \theta^T X$ will be designed and the threshold classifier output is $h\theta(x)$ at 0.5. If the value of hypothesis $h\theta(x) \geq 0.5$, it will predict y = 1 which means that the person has heart disease and if the value of $h\theta(x) < 0.5$, then predict y = 0 which shows that the person is healthy. As a result, under the constraint $0 \leq h\theta(x) \leq 1$, logistic regression prediction is performed. The sigmoid function for logistic regression can be written as follows:

$$h\theta(x) = g(\theta^T X) \quad (6)$$

where $g(z) = \frac{1}{(1+x^{-z})}$ and $h\theta(x) = \frac{1}{(1+x^{=z})}$.

Similarly, the cost function for logistic regression can be defined as follows:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} cost(h\theta(x^{(i)}), y^{(i)}). \quad (7)$$

2) Decision Tree (DT)

One of the most powerful and well-known prediction instruments is the Decision Tree (DT)method [48], which has only two Classes. A decision tree shape is just a tree where every node is a leaf node or decision node, every interior node in the structure of a Decision Tree refers to testing a property, every branch corresponds to a test outcome, and each leaf node is a separate class. Both classification and regression issues can be solved with this algorithm [49][50].

3) Random Forest (RF)

The Random Forest (RF) classifier is an ensemble algorithm or bagged decision trees [51],This means that it is made up of multiple algorithms. Usually It consists of many DT algorithms in this situation. During the training section, RF constructs a full forest from many uncorrelated and random Decision Trees. Multiple learning algorithms are used in ensemble learning approaches to create an ideal predictive model that can outperform any of the individual models' predictions [52] .The computational complexity of the RF may increase as more features are used than in a standalone DT, but it has a higher accuracy when dealing with unseen datasets. The Random Forest algorithm produces the same meaning as the total number of Decision Tree algorithms. To achieve the optimum outcome, the Random Forest ensemble classifier constructs and integrates many decision trees. It primarily refers to tree learning via bootstrap aggregation. Allow the data given to be used $X = \{x1, x2, x3, \ldots\ldots, xn\}$ with responses $Y = \{ x1, x2, x3, \ldots\ldots, xn\}$ with a lower limit of b = 1 and an upper limit of B: The prediction for sample x^i s made by averaging the predictions $\sum_{b=1}^{B} f_b(x^i)$from every individual tree for x^i that is shown using (8).

$$j = \frac{1}{B}\sum_{b=1}^{B} f_b(x^i) \quad (8)$$

For huge data analysis, the Random forest (RF) classifier, which combines multiple distinct tree predictors, is often used [53]. It is a learning method for grouping, regression, and other functions in an ensemble.

4) Support Vector Machine (SVM)

SVMs are machine learning classification techniques that tend to be used in solving classification problems [54] [55]. By employing the maximum margin method, SVM was able to solve a challenging quadratic programming problem. Because of SVM's high classification performance, it has been widely used in a variety of applications [15].In a problem of binary classification, A hyperplane is used to separate the instances $w^T x + b = 0$, where $\mathcal{W}$ and d are dimensional coefficient vectors, which are normal to the hyperplane of the surface, $b$is offset value from the origin, and x is data set values. The SVM receives the $\mathcal{W}$ and $b$results. In the linear example, $\mathcal{W}$ can be solved by using Lagrangian multipliers. The data points on borders are called support vectors. The solution of $\mathcal{W}$ can be written as: $w = \sum_{i=1}^{n} \alpha_i y_i x_i$, where n is the number of support vectors and y_i are target labels to x. After calculating the values of $\mathcal{W}$ and b, the linear discriminant function can be represented as follows:

$$g(x) = sgn(\sum_{i=1}^{n} \alpha_i y_i x_i^T x + b) \quad (9)$$

For the kernel trick and decision function, the nonlinear scenario can be expressed as follows:

$$g(x) = sgn(\sum_{i=1}^{n} \alpha_i y_i K(x_i^T x + b) \quad (10)$$

5) K-Nearest Neighbours (KNN)

KNN is one of the most widely used categorization methods in supervised learning. Previously, it was used to diagnose heart disease. Since KNN does not rely on data distribution assumptions, it is classified as nonparametric. When a new dataset is compared to an existing dataset, KNN places it in the class closest to the existing data. It is used to solve both regression problems and recognition problems. Since KNN does not rely on data distribution assumptions, it is classified as nonparametric. As it compares the new data to existing data, KNN classifies it in the class closest to the available data. It is sometimes known as the "lazy algorithm" [56] because it does not learn from a set of training data as soon as it is collected. KNN calculates the Euclidean distance between new A (x1, y1) data and the previously available B (x2, y2) data, using the formula (11) [57].

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (11)$$

The Euclidean distance between two data points (x2, x1) and (y2, y1) can be determined in two-dimensional space. KNN adds new data points to the class with the least Euclidean distance.

6) AdaBoost or Adaptive Boosting (AdaBoost)

AdaBoost is a binary classification algorithm that combines several weak classifiers into a stronger one [58]. AdaBoost predicts accuracy by analyzing 1000 samples. According to [59], the training dataset entries are weighted with beginning weights.

$$\text{Weight (Xi)} = 1/N \quad (12)$$

Using the decision stump, output is produced for a each input value, when N is the frequency of the training sample, and *xi* is the *ith* training sample. A failure rate can be determined by a formula.

$$\text{Error}=(\text{correct}-N)/N \quad (13)$$

N This value indicates the number of training sample. As a result, the classifiers concentrate on results that are harder to classify. Boosting is the process of combining many simple trainers to generate a more accurate prediction. For

both samples and classifiers, AdaBoost fixes the weights that fluctuate. Equation illustrates the final classifier.

$$H_k(p) = +/-(\sum_{k=1}^{k} \quad a_k h_k(p)) \qquad (14)$$

Equation (13) is a linear combination of all the weak classifiers, and *K* is the total amount of weak classifiers (simple learners). $h_k(p)$ is the output of the weak classifier *t*. (It can either be -1 or 1.) a_k is the weight of classifier *k*.

7) Gradient Boosting (GB)

GB is a one-hundred-sample boosting strategy for classification and regression problems [60]. An improved error function, a weak learner for producing prediction, and an additive model to combine weak learners to reduce the loss function are the three main components of GB. GB is a technique for increasing an algorithm's efficiency by reducing overfitting. Use the gradient tree when there is an imbalance between the numbers in each class, boosting to the Tobit model, sometimes known as the "Grabit" model, helps to increase accuracy. Boosting rather than base techniques, also called as regression tree learners, are used to obtain higher projected precision on a wide variety of datasets, although they rely on expertise in a specific region. The difference between the Boosting procedure and ordinary machine learning is that in function space, optimization is not allowed. The optimum function *F(X)* is obtained after m^{-th} *mth* repetitions, and is deduced as follows: (15)

$$F(X) = \sum_{1=0}^{m} f_i\,(X) \qquad (15)$$

where$f_i(X)$ (i = 1, 2,..., M) denotes feature increments, and $f_i(X) = \rho_i x\, gm(X)$ The greatest loss function correlated with negative gradients is the most recent base-learner. The m^{-th} iteration's negative gradient is (16)

$$gm = -[\frac{\partial L(y,(F(X))}{\partial FX)}]\, F(X) = Fm - 1(X) \qquad (16)$$

When *F(X) = Fm −1(X)*, gm is the path along which the loss function reduces the most quickly (X). The goal of a new decision tree is to correct the inaccuracies of the previous base learner. After that, the T model will be changed to (17).

$$f_m(X) - f_m - 1(X) + \rho_m X h_m(X, \alpha_m) \qquad (17)$$

8) Stochastic gradient descent (SGD)

In machine learning techniques, stochastic gradient descent is a well-known and widely used method. Gradient descent is an iterative procedure that begins at a random position on the slope of a function and gradually lowers until it ∇ reaches the function's lowest point. Both statistical estimation and machine learning are used to solve the problem of minimizing an objective function in the form of a sum:

$$Q(w) = \frac{1}{n}\sum_{1=i}^{n} \quad Q_i(w) \qquad (18)$$

The parameter w that minimizes $Q(w)$ must be estimated in this case.

The i-th observation in the data set is normally connected with each summand function displaystyle Q_i (used for training)[61].

A conventional (or "batch") gradient descent method would conduct the following iterations when used to minimize the above function[62]:

$$w := w - \eta \bigtriangledown Q(w) = w - \frac{\eta}{n}\sum_{i=1}^{n} \quad \bigtriangledown Q_i(w) \qquad (19)$$

G. Difference Resources for Machine Learning

The implemented models are written in a colaboratory or “colab notebook”, a product of Google Research. Colab is written in Python and makes use of simple libraries such as Pandas, numpy, Seaborn, Pyplot, and Scikit-learn.

H. Hyper Parameter used

GridSearchCV was employed in our proposed models to get a higher level of accuracy. The following parameters were applied to the algorithms under investigation Take a look at Table.6 for further information.

sklearn.model_selection.*GridSearchCV*(estimator, param_grid, *, scoring=None, n_jobs=None, refit=True, cv=None, verbose=0, pre_dispatch='2*n_jobs', error_score=nan, return_train_score=False)

The model's parameters are chosen to create the best possible accuracy match between the training and validation data.

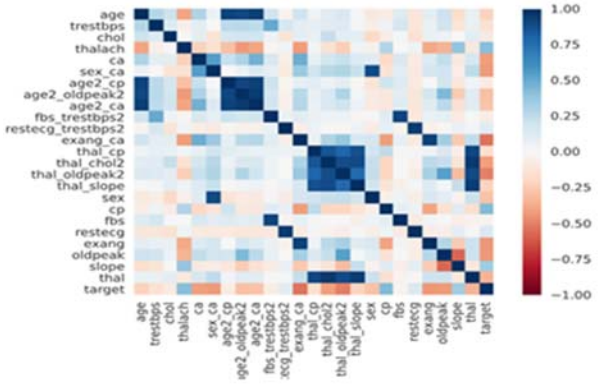
TABLE 3. HYPER PARAMETER USED.

methods	**Parameters**
RF	{'bootstrap': False, 'criterion': 'gini', 'max_depth': 5, 'max_features': 'auto', 'min_samples_leaf': 12, 'min_samples_split': 70, 'n_estimators': 40}
AdaBoost	{'learning_rate': 0.095}
SGD	{'alpha': 0.035}
SVM	{'kernel': 'linear', 'tol': 0.001}
LR	{'C': 0.3}
DT	{'min_samples_leaf': 5}
KNN	{'n_neighbors': 5}
GB	{'learning_rate': 0.07, 'max_depth': 4, 'min_samples_leaf': 5}

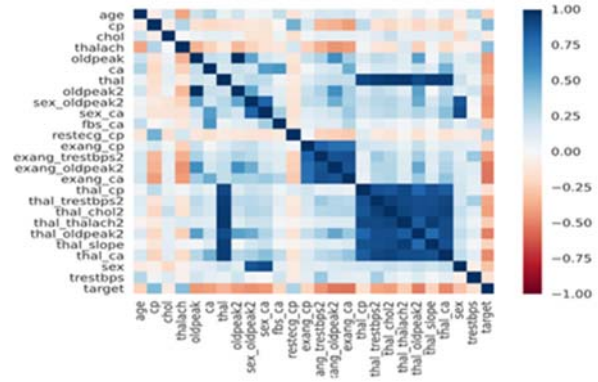
IV. RESULTS AND DISCUSSION

A. Outcomes of feature selection processes

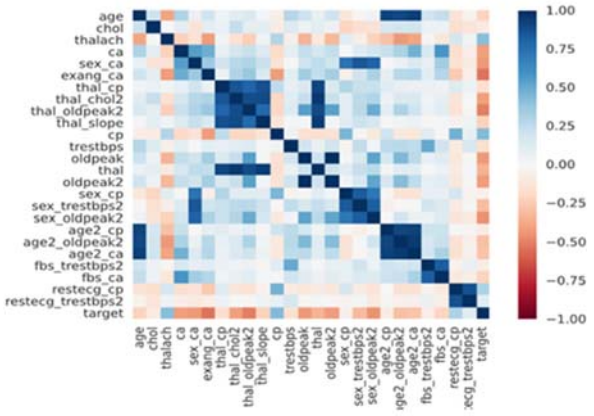
Feature selection techniques are always used to select the most important features from a dataset. In this study, three techniques, LASSO, RFE-RF, and BLR were applied to select different features and select the best 25 features from the outputs of each technology separately. Fig.3 shows a heatmap of 25 features produced using Pearson's correlation coefficient after using selection techniques LASSO, RFE-RF, and BLR, respectively shown in Fig 3. The heatmap shows that all feature selection methods can choose the most relevant features from the data set. A negative correlation carries unique information and aids the model's prediction ability. The scaled heatmap depicts feature correlation, with -1.00 representing negatively correlated (red color) and 1.00 representing favourably correlated (blue color) features, respectively.



(a)



(b)



(c)

Fig.3 Pearson Correlation Coefficient Heat Map of Feature Selection Techniques, (a)LASSO, RFE-RF(b), BLR(c), (25 in total) of The Dataset.

B. Comparison of Various Machine Learning Algorithms on Different Features

In this section, we compare the classification results of different models that use different input features. First, we applied eight machine learning models to all the features of the Cleveland heart disease (CHD) dataset. Second, eight algorithms were applied to the 25 features selected by the Least Absolute Shrinkage and Selection Operator Features (LASSO) section algorithm. Third, the top 25 features selected by the Recursive Feature Elimination with Random Forest feature selection (FFE_RF) feature selection algorithm were input into the classification models. Finally, Best Features of Combined LASSO and RFE-RF (BLR) were implemented to extract 25 relevant features, and the same machine learning classifiers were applied. The expected outcomes are also assessed using a variety of performance metrics. A confusion matrix was created to evaluate and compare all algorithms. A confusion matrix was created. The performance metrics are accuracy, RMSE, AUC, sensitivity, specificity, precision, F1-Score, TP, FP, TN, and FN.

1)Comparison of classification methods based on accuracy

When it comes to evaluating machine learning algorithms, accuracy is widely regarded as the most important criterion. As already stated, we used the eight ML approaches characteristic of the original input attributes, then the 25 input features of the LASSO algorithm, 25 input features of the RFE-RF method, and 25 input features of the BLR method. Fig 4 depicts the accuracy of several classifiers.

Given all the features, SVM and AdaBoost had the highest accuracy of 90.16%. While KNN is 88.52% accurate, LR, SGD, and GB classifiers get 86.89%, 85.25%, 85.25%, and 83.61% respectively. The accuracy of DT and RF are

extremely similar (82.97%). When we applied the twenty-five features selected by LASSO, the RF was the best classifier among all in terms of accuracy, having an excellent accuracy of 96.72%. Classifiers LR, SVM, SGD, AdaBoost, KNN, GB, and DT show good accuracy in achieving 93.44%, 91.80%, 90.16%, 90.16%, 88.52%, 85.25%, and 78.69% respectively. Consider the accuracy of these eight ML approaches with features of RFE-RF, the results of AdaBoost, SGD and RF are similar (90.16%). The SVM, LR, KNN, GB and DT classifiers also provide good accuracy: 86.89%, 85.25%, 83,61%, 83.61%, and 81.97% respectively. When ranking BLR's Top 25 Special Features, the accuracy of the KNN classifier is the lowest (78.69%). We obtained an accuracy of 91.80 for both AdaBoost and SGD. The RF, SVM, and DT achieved 90.16%, 88.52%, and 83.61%, respectively. The accuracy of LR and GB are very similar to each other (85.25%).

Significant improvement in accuracy with the use of features selected by feature selection algorithms especially LASSO, which is observed in the RF model, from 81.97% to more than 96% accuracy.

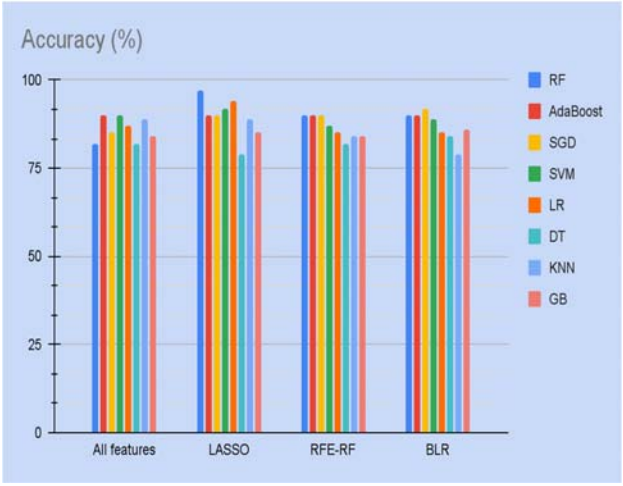


Fig 4. Comparison of Classification Methods Based on Accuracy

1) Comparison of classification methods based on precision

Precision and other performance indicators have also been used to evaluate the performance of classifiers. Evaluate the performance of classification algorithms using precision. Based on All features, AdaBoost produced an exceptional result with a precision score of 92%. GB has the lowest precision score: 84%. A precision score between these values is found in the other models. For LASSO features, RF provided the highest level of precision score (100%), and LR also had an excellent precision score of 94%, while the GB classifier had the lowest precision (84%). When the RFE-RF features are combined, RF and AdaBoost get high precision score of approximately 94%. the lowest precision (86%) for the KNN classifier. On BLR features, the best precision was obtained with the RF (94%), and the lowest with the GB (84%). The precision results are displayed in Fig 5.

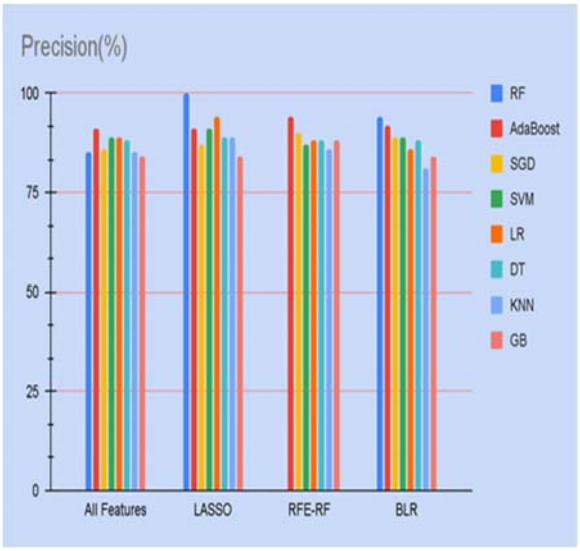


Fig.5 Comparison of Precision.

2)Comparison of classification methods-based Recall or sensitivity score

Performance metrics such as recall, or sensitivity are important because correctly identifying patients with heart disease is important. When applied All features, the highest recall score was reached by KNN (92 %). There was a low recall score (just over 82%) obtained from the RF. For the LASSO features, the DT algorithm produced a relatively low recall score (just over 71%), whereas the SGD algorithm produced the highest recall score (over 97%). The DT algorithm generated poor results (just over 80%), while the LR and KNN algorithms produced the highest recall score (just over 95%) when applied to the RFE-RF features. With the BLR features, the DT and KNN classifiers obtained similar low recall scores of 83%, while the SGD achieved the highest recall score (97%). The recall scores for the various algorithms and feature sets are shown in Fig 6.

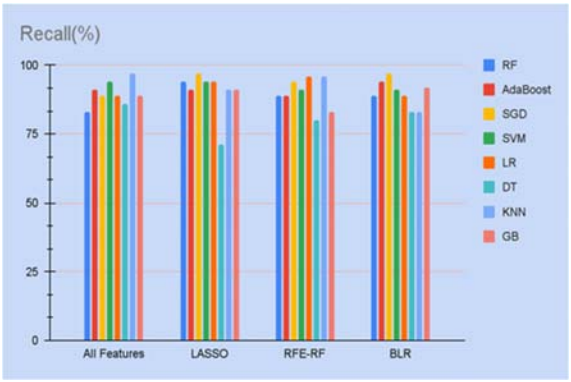


Fig. 6: Showing Recall Scores.

3)Comparison of classification methods based on F1-score

The harmonic means of precision and recall scores is labelled F1-score. For all features, SVM has the highest F1 score (92%), while DT scores are the lowest (84%). when LASSO features, The RF classifier surpassed all other algorithms with the highest F1-score (just over 97 %). DT had the lowest F1 score of (over 79%), and LR, SGD, and AdaBoost also performed well at 94%, 92%, 91%, respectively. With RFE RF attributes, SGD receives the greatest F1 score of (92%) while DT receives the lowest result of (84%). SGD receives the highest F1 score of (over 93%) while KNN receives the lowest result of (82%). Fig. 7 shows F1 scores as percentages.

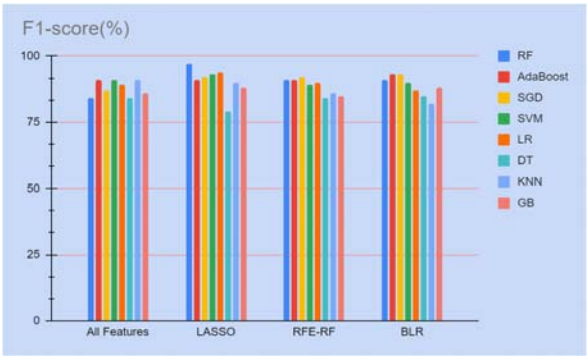


Fig. 7: Displays F1 Scores as Percentages.

4)Comparison of classification methods based on Specificity

Specificity of True Negative Rate (TNR) of the various algorithms has also been evaluated. When RF was applied to the LASSO feature selection, it yielded the maximum TNR (100%). The lowest TNR was recorded for GB (76.9%). For All features, the performance of the classifiers was not so good. The best TNR, for AdaBoost, was only 88.5% and the lowest TNR was recorded for GB and DT (76.9%). The AdaBoost and RF had the highest TNR (92.3%) and KNN had the lowest TNR (80.8%) when applied to the RFE-RF feature selection. For BLR features, receives the highest TNR of (over 92%) while GB and KNN receive the lowest results of (76.9%) and (73.1%) respectively. Fig 8 shown TNR.

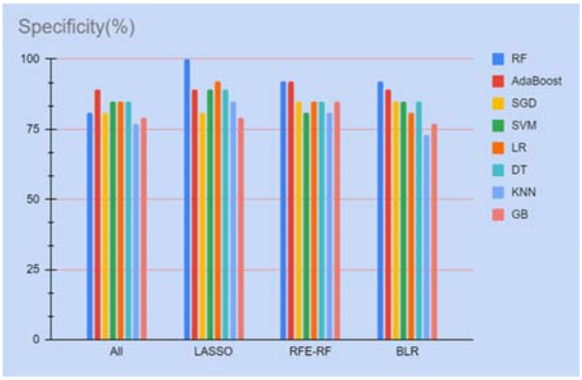


Fig. 8 Specificity of True Negative Rate (TNR).

5)Comparison of classification methods based on ROC-AUC score

The quality of binary classifiers is assessed by the AUC (Area Under the Curve) of the ROC (Receiver Operating Characteristic; default) or PR (Precision Recall) curves. ROC-AUC includes all a model's operational points, unlike accuracy. According to Fig 9, the highest AUC was reached by AdaBoost and SVM (90%) and the lowest AUC was recorded for DT (82%) when applied to all features. The best AUC was obtained by evaluating the LASSO features by the RF, which was (slightly more than 97%). For LASSO features, the lowest AUC recorded for a DT was close to 80%. RF, AdaBoost, and SGD scored the highest, near 91%, when applying RFE-RF features, but the lowest AUC for DT was 82%. However, when applying the BLR features, the results from DT, AdaBoost, and SGD were like the previous results, but the lowest AUC for DT is 79%.

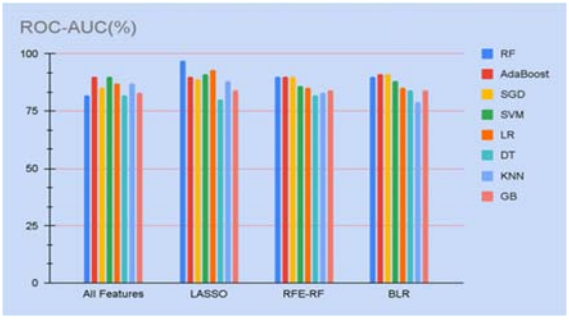


Fig. 9 ROC-AUC (%).

6)Comparison of classification methods based on Root mean square error (RMSE)

Fig 10 displays root mean square error for all features. The highest RMSE was reached by RF and DT (42.47%). There was RMSE (31.36%) obtained from the SVM. The RF provided the best RMSE (18.11) for LASSO features, and the lowest RMSE was recorded for DT (46.16). When using the RFE-RF features, DT got a high RMSE of 42.47, the lowest RMSE of 31.36 for the RF, AdaBoost and SGD classifiers. On BLR features, the best RMSE of 28.63 was recorded for SGD and the lowest obtained from the KNN was 46.16.

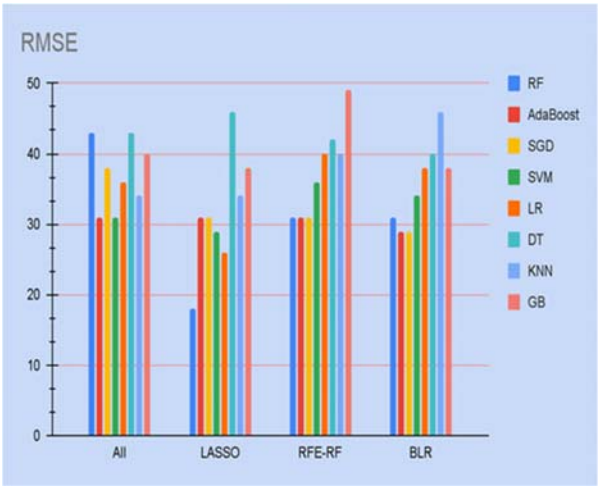


Fig. 10: Root Mean Square Error (RMSE).

7)A Comparative Analysis of Runtime (RT)

The run time is calculated as the unit is in seconds. A comparison of the computation time of the eight ML approaches of the original input features, then the 25 input features of the LASSO approach, the 25 input features of the RFE-RF method, and the 25 input features of the BLR method, is tabulated in TABLE 4. We can clearly notice that the RF model has the longest RT, 56.8 for BLR features. and the lowest RT of 0.569 recorded by SGD for LASSO features.

TABLE 4. COMPARATIVE ANALYSIS OF TIME.

Classifier	RF	AdaBoost	SGD	SVM	LR	DT	KNN	GB
All	54.5	2.27	0.725	0.688	0.747	0.748.	0.713	54.1
LASSO	48.8	1.93	0.569	0.941	2.05	0.639	0.639	31.2
RFE-RF	52.6	2.03	0.658	3.31	0.670	0.663	0.681	34.7
BLR	56.8	2.33	0.674	0.981	0.727	0.705	0.717	38.3

A.A comparison of accuracy recall and specificity between the proposed system and some existing systems on UCI heart disease Cleveland dataset

The performance of our study, the proposed framework for a machine-learning-based intelligent decision support system for the prediction of heart disease, is compared with several core methodologies that have been recently introduced and developed by academics and researchers to aid in the development of a decision support system for diagnosing heart disease. Our research improves the system's overall accuracy by approximately 97%, whereas the highest previous results were a little over 93.4%, in [31]. as presented statistically as well as pictorially in TABLE 5 and Fig 11. it was also noticed that our study had the greatest recall value of 94.29%, whereas [12]. had the second-best score of 92.8% [13]. The maximum score of 100% at Specificity is obtained by both our study and researcher in [14]. The performance of our proposed framework is very good compared to earlier research studies.

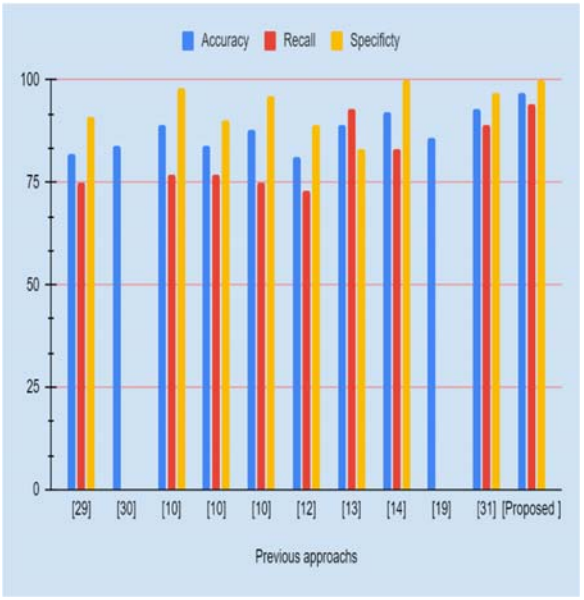


Fig.11 Comparison of Accuracy, Recall, and Specificity of the Proposed System and Some Existing Systems on The UCI Heart Disease Dataset.

B.performance of the best classifier RF approach with different features

Table 5 shows the performance of the RF approach with different features and scales. It can be seen from the confusion matrix of classifiers that the features of RF using LASSO are relatively better for classifying a normal subject as a normal subject and a cardiac patient as a cardiac patient. He has never misclassified normal people and there are only two cases where RF predicts heart disease as normal people.

We also note an increase in RF accuracy from 90.16% for all features to 96.72% after applying LASSO features, as well as a significant improvement in performance and an increase in all rating metrics, as can be seen from Table 6.

TABLE 5. A COMPARISON OF ACCURACY, RECALL AND SPECIFICITY BETWEEN THE PROPOSED SYSTEM AND SOME EXISTING SYSTEMS ON UCI HEART DISEASE CLEVELAND DATASET.

Source	Approach	Accuracy	Recall	Specificity
[29]	PPCA	82.18	75	90.57
[30]	PSO with SVM	84.36	-	-
[10]	Relief +LR	89	77	98
[10]	mRMR+NB	84	77	90
[10]	LASSO+SVM	88	75	96
[12]	RBF kernel-based SVM	81.19	72.92	88.68
[13]	HRFLM	88.7	92.8	82.6
[14]	L1 Linear SVM+L2 Linear RBFSVM	92.22	82.92	100
[19]	Hybrid Approach+Majority vote with NB, BN, RF and MP	85.48	-	-
[31]	FAMD+RF	93.44	89.28	96.96
Proposed	MFC _LASSO+RF	96.72	94.29	100

TABLE 6. PERFORMANCE OF RF CLASSIFIER IN ALL FEATURES AND LASSO FEATURES SELECTION.

features using	TP	FP	TN	FN	Accuracy	Pre cision	Recall	F1-score	Specificity	AUC
All features	29	6	21	5	90.16	93.94	88.57	91.18	80.8	90.44
LASSO	33	2	26	0	96.72	100	94.29	97.06	100	97.14

V. CONCLUSION

The proposed system was created with the aim of developing an intelligent machine learning system for diagnosing heart diseases. By providing accurate, and early detection of heart disease, the system will improve patient survival rates. The proposed system uses properties of (MFC) to extract and extract features from the Cleveland UCI cardiology dataset and train machine learning predictive models to classify cases in addition to predicting heart disease and normal subjects. We also used selection techniques: LASSO, RFE-RF, and BLR.

This study demonstrates how LASSO feature identification technology can create a highly correlated feature set that can be used with a variety of machine learning algorithms. The research also discovered that RF performs well with high-impact properties and generates much higher resolution than similar works. With the features of LASSO, the RF achieved an accuracy of 96.72%.

In future work, by combining a larger dataset with other feature selection algorithms and optimization techniques, we will be able to further improve the accuracy of these predictive classifications for diagnosing heart disease.

REFERENCES

[1] J. Ross, Q. Morgan, and K. Publishers, "Book Review : C4 . 5 : Programs for Machine Learning," vol. 240, pp. 235–240, 1994.

[2] T. Shaikhina, D. Lowe, S. Daga, D. Briggs, R. Higgins, and N. Khovanova, "Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation," Biomed. Signal Process. Control, vol. 52, pp. 456–462, 2019, doi: 10.1016/j.bspc.2017.01.012.

[3] R. Chandra, K. Chaudhary, and A. Kumar, "The Combination and Comparison of Neural Networks with Decision Trees for Wine Classification," no. January, pp. 10–17, 2007.

[4] M. I. Al-janabi, M. H. Qutqut, and M. Hijjawi, "Machine Learning Classification Techniques for Heart Disease Prediction : A Machine learning classification techniques for heart disease prediction : a review," no. October, 2018, doi: 10.14419/ijet.v7i4.28646.

[5] N. Kausar, S. Palaniappan, B. B. Samir, A. Abdullah, and N. Dey, "Systematic analysis of applied data mining based optimization algorithms in clinical attribute extraction and classification for diagnosis of cardiac patients," in Applications of intelligent optimization in biology and medicine, Springer, 2016, pp. 217–231.

[6] Y. Khan, U. Qamar, N. Yousaf, and A. Khan, "Machine Learning Techniques for Heart Disease Datasets : A Survey," Assoc. Comput. Mach. ACM ISBN, pp. 27–35, 2019.

[7] T. Mythili, D. Mukherji, N. Padalia, and A. Naidu, "A Heart Disease Prediction Model using SVM-Decision Trees-Logistic Regression (SDL)," Int. J. Comput. Appl. (0975, vol. 68, no. 16, pp. 11–15, 2013.

[8] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, "Innovative Artificial Neural Networks-Based Decision Support System for Heart Diseases Diagnosis," J. Intell. Learn. Syst. Appl., vol. 05, no. 03, pp. 176–183, 2013, doi: 10.4236/jilsa.2013.53019.

[9] Q. K. Al-Shayea, "Artificial neural networks in medical diagnosis," Int. J. Comput. Sci. Issues, vol. 8, no. 2, pp. 150–154, 2011.

[10] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms," Hindawi Mob. Inf. Syst. accuracy, vol. 2018, 2018.

[11] K. Vanisree and J. Singaraju, "Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks," Int. J. Comput. Appl., vol. 19, no. 6, pp. 6–12, 2011.

[12] S. M. Saqlain et al., "Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines," Knowl. Inf. Syst., vol. 58, no. 1, pp. 139–167, 2019.

[13] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," IEEE access, vol. 7, pp. 81542–81554, 2019.

[14] L. Ali et al., "An optimized stacked support vector machines based expert system for the effective prediction of heart failure," IEEE Access, vol. 7, pp. 54007–54014, 2019.

[15] M. S. Amin, Y. K. Chiam, and K. D. Varathan, "Identification of significant features and data mining techniques in predicting heart disease," Telemat. Informatics, vol. 36, pp. 82–93, 2019.

[16] D. Singh and J. S. Samagh, "A comprehensive review of heart disease prediction using machine learning," J. Crit. Rev., vol. 7, no. 12, pp. 281–285, 2020.

[17] F. Z. Abdeldjouad, M. Brahami, and N. Matta, "A hybrid approach for heart disease diagnosis and prediction using machine learning techniques," in International conference on smart homes and health telematics, 2020, pp. 299–306.

[18] M. Tarawneh and O. Embarak, "Hybrid approach for heart disease prediction using data mining techniques," in International Conference on Emerging Internetworking, Data & Web Technologies, 2019, pp. 447–454.

[19] C. B. C. Latha and S. C. Jeeva, "Informatics in Medicine Unlocked Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," Informatics Med. Unlocked, vol. 16, no. July, p. 100203, 2019, doi: 10.1016/j.imu.2019.100203.

[20] S. Moturi, S. Vemuru, and S. N. Tirumala Rao, "Classification model for prediction of heart disease using correlation coefficient technique," Int. J. Adv. Trends Comput. Sci. Eng., vol. 9, no. 2, pp. 2116–2123, 2020, doi: 10.30534/ijatcse/2020/185922020.

[21] J. Chen, G. Xi, Y. Xing, J. Chen, and J. Wang, "Predicting Syndrome by NEI Specifications : A Comparison of Five Data Mining Algorithms in Coronary Heart Disease," pp. 129–130.

[22] A. Lakshmanarao, Y. Swathi, and P. S. S. Sundareswar, "Machine Learning Techniques For Heart Disease Prediction," Int. J. Sci. Technol. Res., vol. 8, no. 11, pp. 1–4, 2019.

[23] M. Tayefi et al., "hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree algorithm," Comput. Methods Programs Biomed., vol. 141, pp. 105–109, 2017, doi: https://doi.org/10.1016/j.cmpb.2017.02.001.

[24] F. S. Alotaibi, "Implementation of machine learning model to predict heart failure disease," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 6, pp. 261–268, 2019.

[25] A. Sabay, L. Harris, and K. Jaceldo-siegl, "Overcoming Small Data Limitations in Heart Disease Prediction by Using Surrogate Data," vol. 1, no. 3, 2018.

[26] S. I. Sherly, "An Ensemble Basedheart Disease Predictionusing Gradient Boosting Decision Tree," Turkish J. Comput. Math. Educ., vol. 12, no. 10, pp. 3648–3660, 2021.

[27] V. Chaurasia and S. Pal, "Early prediction of heart diseases using data mining techniques," Caribb. J. Sci. Technol., vol. 1, pp. 208–217, 2013.

[28] K. Saxena and R. Sharma, "Efficient heart disease prediction system," Procedia Comput. Sci., vol. 85, pp. 962–969, 2016.

[29] S. M. S. Shah, S. Batool, I. Khan, M. U. Ashraf, S. H. Abbas, and S. A. Hussain, "Feature extraction through parallel probabilistic principal component analysis for heart disease diagnosis," Phys. A Stat. Mech. its Appl., vol. 482, pp. 796–807, 2017.

[30] J. Vijayashree and H. P. Sultana, "A machine learning framework for feature selection in heart disease classification using improved particle swarm optimization with support vector machine classifier," Program. Comput. Softw., vol. 44, no. 6, pp. 388–397, 2018.

[31] A. Gupta, S. Member, and R. Kumar, "MIFH : A Machine Intelligence Framework for Heart Disease Diagnosis," no. Ml, pp. 14659–14674, 2020.

[32] P. Ghosh et al., "Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques," IEEE Access, vol. 9, pp. 19304–19326, 2021, doi: 10.1109/ACCESS.2021.3053759.

[33] M. Fatima and M. Pasha, "Survey of Machine Learning Algorithms for Disease Diagnostic," J. Intell. Learn. Syst. Appl., vol. 09, no. 01, pp. 1–16, 2017, doi: 10.4236/jilsa.2017.91001.

[34] D. H. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," vol. 01, no. 04, pp. 140–147, 2020, doi: 10.38094/jastt1457.

[35] R. Elarabi, F. Alqahtani, A. Balobaid, H. Zain, and N. Babiker, "COVID-19 Analysis and Predictions Evaluation for KSA Using Machine Learning," in 2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM), 2021, pp. 261–266.

[36] R. S. Sutton and A. G. Barto, Reinforcement learning: An introduction. MIT press, 2018.

[37] J. P. Dominguez-Morales, A. F. Jimenez-Fernandez, M. J. Dominguez-Morales, and G. Jimenez-Moreno, "Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors," IEEE Trans. Biomed. Circuits Syst., vol. 12, no. 1, pp. 24–34, 2017.

[38] "Heart Disease UCI exp," UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/Heart+Disease, (accessed Jul. 25, 2018).

[39] F. Developers, "feature _ engine Documentation," 2021.

[40] M. Umair et al., "Main Path Analysis to Filter Unbiased Literature," Intelligent Automation \& Soft Computing , vol. 32, no. 2. 2022, doi: 10.32604/iasc.2022.018952.

[41] A. Acharya, "Comparative study of machine learning algorithms for heart disease prediction," 2017.

[42] R. Tibshirani, "Regression shrinkage and selection via the lasso: a retrospective," J. R. Stat. Soc. Ser. B (Statistical Methodol., vol. 73, no. 3, pp. 273–282, 2011.

[43] C. Zhou and A. Wieser, "Jaccard analysis and LASSO-based feature selection for location fingerprinting with limited computational complexity," in LBS 2018: 14th International Conference on Location Based Services, 2018, pp. 71–87.

[44] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products," Chemom. Intell. Lab. Syst., vol. 83, no. 2, pp. 83–90, 2006.

[45] X. Lin et al., "A method for handling metabonomics data from liquid chromatography/mass spectrometry: combinational use of support vector machine recursive feature elimination, genetic algorithm and random forest for feature selection," Metabolomics, vol. 7, no. 4, pp. 549–558, 2011.

[46] F. E. Harrell, "Ordinal logistic regression," in Regression modeling strategies, Springer, 2015, pp. 311–325.

[47] K. Larsen, J. H. Petersen, E. Budtz-Jørgensen, and L. Endahl, "Interpreting parameters in the logistic regression model with random effects," Biometrics, vol. 56, no. 3, pp. 909–914, 2000.

[48] V. Vapnik, The nature of statistical learning theory. Springer science & business media, 1999.

[49] P. Ghosh, M. Z. Hasan, and M. I. Jabiullah, "A comparative study of machine learning approaches on dataset to predicting cancer outcome," J. Bangladesh Electron. Soc., vol. 18, no. 1–2, pp. 81–86, 2018.

[50] X. Wu et al., "Top 10 algorithms in data mining," Knowl. Inf. Syst., vol. 14, no. 1, pp. 1–37, 2008.

[51] F. M. J. M. Shamrat, Z. Tasnim, P. Ghosh, A. Majumder, and M. Z. Hasan, "Personalization of job circular announcement to applicants using decision tree classification algorithm," in 2020 IEEE International Conference for Innovation in Technology (INOCON), 2020, pp. 1–5.

[52] K. Fawagreh, M. M. Gaber, and E. Elyan, "Random forests: from early developments to recent advancements," Syst. Sci. Control Eng. An Open Access J., vol. 2, no. 1, pp. 602–609, 2014.

[53] R. Banerjee et al., "Time-frequency analysis of phonocardiogram for classifying heart disease," in 2016 Computing in Cardiology Conference (CinC), 2016, pp. 573–576.

[54] A. E. Karrar, "Investigate the Ensemble Model by Intelligence Analysis to Improve the Accuracy of the Classification Data in the Diagnostic and Treatment Interventions for Prostate Cancer," vol. 13, no. 1, pp. 181–188, 2022.

[55] A. E. Karrar, "A Proposed Model for Improving the Performance of Knowledge Bases in Real-World Applications by Extracting Semantic Information," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 2, pp. 116–123, 2022, doi: 10.14569/IJACSA.2022.0130214.

[56] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," Ann. Transl. Med., vol. 4, no. 11, 2016.

[57] P. Cunningham and S. J. Delany, "k-Nearest neighbour classifiers-A Tutorial," ACM Comput. Surv., vol. 54, no. 6, pp. 1–25, 2021.

[58] R. Rojas, "AdaBoost and the super bowl of classifiers a tutorial introduction to adaptive boosting," Freie Univ. Berlin, Tech. Rep, 2009.

[59] K. W. Walker and Z. Jiang, "Application of adaptive boosting (AdaBoost) in demand-driven acquisition (DDA) prediction: A machine-learning approach," J. Acad. Librariansh., vol. 45, no. 3, pp. 203–212, 2019.

[60] C. Bent, A. Cs, and G. Mart, "A Comparative Analysis of XGBoost arXiv : 1911 . 01914v1 [cs . LG] 5 Nov 2019," pp. 1–20.

[61] L. Bottou, "Stochastic gradient descent tricks," in Neural networks: Tricks of the trade, Springer, 2012, pp. 421–436.

[62] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in Neural networks: Tricks of the trade, Springer, 2012, pp. 9–48.

.