

Enhancing Heart Disease Prediction Using Ensemble Techniques

Wasilah Sada^{1,2}, Celinus Kiyee², Baththama Bello Alhassan³, Aminat Bolatito Yusuf^{1,4}, Fatsuma Jauro², and Maria Abur³

¹Department of Computer Science, National Mathematical Centre Abuja, Nigeria

²Department of Computer Science, Ahmadu Bello University, Zaria, Nigeria

³Iya Abubakar Institute of ICT, Ahmadu Bello University, Zaria, Nigeria

⁴Department of ICT, Usmanu Danfodiyo University, Sokoto, Nigeria

wasilahsada@gmail.com^{1,2}, ckiyea@gmail.com², bbalhasaan8510@gmail.com³, aminat.bola@udusok.edu.ng^{1,4},
fatijauro@gmail.com², mmabur1@gmail.com³

Abstract

Background: Cardiovascular diseases are recognized generally to be among the number one illness causing death across the globe. Predicting heart disease using a computer-aided technique makes it easier for medical practitioners to diagnose and thereby saving lives and reducing costs. Feature selection has become an essential component for developing Machine learning models. It chooses the most relevant features from the available dataset, thereby shortening the training period, making the model easier to train, improving generalization and decreasing overfitting without necessarily compromising the system's accuracy. **Aim:** The purpose of this work is to design and build an optimal model for the prediction of heart diseases, especially at an early stage by considering certain features that are most relevant for the prediction without compromising the system's accuracy. **Method:** The Cleveland UCI dataset with 303 instances were used in training the model and the findings show that selectKBest is an effective tool in improving the prediction of heart diseases. The performance metrics Accuracy, Sensitivity, Precision were measured. **Results:** the study found that when hybridizing k-Nearest Neighbor Bagging, Decision Tree Bagging, Gradient Boosting generated the highest accuracy of 90%, 85% and 88% respectively.

Keywords: — Heart Diseases, feature selection, ensemble technique, prediction

1. Introduction

Data mining is incredibly significant in medical sciences and other related fields. Its ability to extract underlying knowledge from datasets of this domain helps in early detection and possible prevention of some diseases. Heart disease, or sometimes referred to as Cardiovascular Disease (CAD) is the major cause of death in the United States, affecting people of all races and nationalities. (Soni et al., 2011) mentioned that there are more cases of heart diseases in the United States than any other industrialized or developing country. According to the report of (CDCP, 2020), Cardiovascular disease is said to claim the lives of one person every 36 seconds. About 655,000 Americans die annually from heart ailments. Heart diseases cause one-third of all fatalities worldwide. According to the World Health Organization data of 2018, Nigeria records a Heart Disease Deaths of about 108,578 annually. This is equal to about 5.60% of the total deaths in the country. The age adjusted Death Rate is 197.37 per 100,000 people of the population. As such, the country is ranked the 31st in global records.

Heart illness can cause a variety of symptoms, including chest pain, back pain, dizziness, shortness of breath, lightheadedness or fainting. At an early stage, it may include snoring, sweating, indigestion and Jaw pain. A fundamental factor that may trigger risk of heart disease or increasing one's risk of a damaged or narrowed artery of the heart is age. These arteries ensure the supply of blood and hence oxygen to the heart. As a result, any

artery failure has an impact on the heart's general function as well as the function of other organs such as the brain. When compared to men, women around age 40s have a lesser risk of heart disease, but their risk factor grows over time (Yekkala& Dixit, 2018). Genetic factors, tobacco, poor diet, hypertension, high cholesterol, diabetics are factors that also exposes the heart to ailments. There are various forms of heart disorders ranging from heart failure to the less severe ones like Heart Rhythms, structural heart diseases and others as shown in Table 1.

Table 1: The different types of heart disease and their implications as in(Yusuf et al., 2021).

Types of Heart Disease	Description
Coronary artery diseases (CAD)	This occurs when arteries in the heart become blocked. It is the most common type of cardiovascular disease
Heart failure	This is when the heart is completely damaged. Two most common reasons of heart failures are Heart attacks and high blood pressure.
Vascular disease (Angina pectoris)	This is a reduction of the amount of blood that flows into the heart
Congestive heart failure	A situation when the heart stops pumping blood to the body
Congenital heart disease	It is a difference in one or more aspects of one's heart structure from birth.
Structural heart disease	This is when the muscles or valves of the heart becomes disorganized. Structural disease also leads to heart failure
Peripheral artery disease	A situation where thin blood vessels of the heart decreases the flow of blood to the limbs.

Researchers across the globe have tested several techniques from traditional methods to statistical methods for predicting heart diseases. The oldest amongst them is the Angiography method. This method is used in finding the localization of heart vessels' stenosis. Although this method has been helpful in detecting the condition, it does have some drawbacks, including the fact that it is expensive, takes a long time to get results, and the disease is quite invasive. These motivated researchers to develop systems that can automatically interpret information gathered through a set of clinical process such as data from past treatments, outcomes of such treatment and also the latest medical research on such ailments and apply them to the prediction of future outcomes.

Machine Learning (ML) techniques have so far effectively identified heart diseases when trained on proper dataset. Artificial intelligence on the other hand has complemented the predictions and thus made it easier for scholars to develop the right models with the available overwhelming databases (Ghosh et al., 2021). Consequently, the purpose of this study is to design an intelligent system for accurate cardiac disease detection and prediction on Cleveland heart disease dataset. This is accomplished by picking the best subset from a given set of techniques. This thus, lower medical expenses and improve treatment quality. Some of these techniques that are mostly tested include Decision Tree (DT) which is one of the most popular and effective technique used especially among medical practitioners, Random Forest (RF), K-Nearest Neighbor (KNN) which is a

nonparametric method, Adaptive Boosting (AdaBoost or AB) is a Boosting algorithm that is used for binary classification. Hist Gradient Boost (HGB) which is Histogram Gradient Boosting and several other techniques.

Feature selection is a data mining pre-processing approach for removing redundant variables from a dataset by discarding extraneous features. It is a technique that selects the features in a data that have the greatest impact on the target variable. In another words, selection of the best predictors for the target variable was explored in this study. In practice, it is uncommon for all of the variables in a dataset to be useful in the building of a ML model. According to 'Ockham Razor's Law of Parsimony,' the best explanation for a problem is one that requires the fewest assumptions. As a result, feature selection has become an essential component for developing ML models and has many advantages, amongst them are shortening the training period, making the model easier to train, improving generalization and decreasing overfitting.

2. Related Works

Researchers across several fields of study, especially in the health sectors are now dwelling into Artificial intelligence and ML techniques.

(Yusuf et al., 2021) designed a model for predicting heart diseases using University of California Irvine (UCI) heart disease dataset. KNN, DT, RF, AB, Gradient Boosting (GB) and HGB were hybridized with bagging and the result obtained performed better than the traditional model designed without Bagging. The work was validated using 5-fold cross validation and the highest accuracy recorded with bagged RF and HGB with 90% accuracy. This proposed work hopes to extend (Yusuf et al., 2021) by including feature selection in the methodology.

(Ghosh et al., 2021)suggested a framework for forecasting heart problems using different prediction techniques. Five different datasets from UCI were used in the experiment conducted. Relief and Least-Absolute-Shrinkage-Selection-Operator (LASSO) feature selection methods were used to select the most appropriate features. In order to create hybrid models based on KNN, DT, and RF, the bagging technique was utilized. Also, AB and GB boosting methods were used. The proposed model with the RF Bagging approach attained the greatest accuracy of 99.05% in the experiments. The high level of missing values in the dataset proved a barrier to their work.

A metaheuristic algorithm was proposed by (Abdollahi& Nouri-Moghaddam, 2021)to forecast heart disease. Genetic algorithm was used as part of the proposed ensemble learning methods alongside statistical analysis and ML methods. A wrapper feature reduction using genetic method was used for selection of most appropriate features. Specific classifiers used include Multilayer Perceptron (MLP), DT, KNN, Radial Basis Function (RBF), Gaussian Naïve Bayes (GNB), Support Vector Machines (SVM), and Single Conjunctive Rule Learner (SCRL). Both bagging, stacking, and boosting methods have been used. Experiments conducted on UCI Stat log heart disease datasets proved that the stacked-genetic method achieved the highest accuracy of 97.57%.

Dissanayake & Johar (2021) evaluated the performance of heart disease prediction models using various feature selection methods. The researchers considered ten different feature selection techniques, namely: exhaustive, Relief, ANOVA, mutual information. Others also include LASSO regression, Recursive Feature Elimination (RFE), Chi-square, ridge regression forward and backward feature selection. The classifiers used were SVM, DT, GNB, RF, KNN, and Linear Regression (LR). Based on experiments with the Cleveland heart disease dataset, the backward feature selection method with DT achieved the highest accuracy of 88.52%.

Hasan & Bao (2021) proposed a model for predicting heart disease based on filter-wrapper-embedded feature selection approaches. The top five ML classifiers were used (KNN, RF, XGBoost, SVM, and GNB). The Artificial Neural Network (ANN) technique was also applied to enhance heart disease prediction. Light GBM with RFE, RF, LASSO, and Chi-square were the specific methods applied for feature selection. The experiment was conducted on dataset obtained from the Kaggle repository. Results from the experiment showed that the XGBoost classifier with wrapper techniques produced the best prediction.

Similarly, Ishaq et al., (2021) looked at the heart survivors in a 299-patient dataset. The aim was to identify important features and data mining approaches that could increase the accuracy of cardiovascular patient prediction DT, AB, LR, RF and GB classifiers. Synthetic Minority Oversampling Technique (SMOTE) was used to address the imbalance problem. The models were trained using RF to identify the highly rated features. The findings were compared to those obtained using all the features from ML methods. The findings of the experiments showed that the additional tree classifier outperformed other models and had the highest accuracy with SMOTE in predicting cardiac patient survival outcomes.

Li et al., (2020) proposed a 'ML based heart diagnostic system'. Different classifiers including LR, SVM, GNB, ANN, KNN and DT were used to design the system. At the feature selection stage, a novel feature selection algorithm named Fast-Conditional-Mutual-Information feature selection algorithm (FCMIM) was used alongside other four popular feature selection namely: LASSO, Local-Learning Based-Features-Selection (LLBFS), Relief, Minimal-Redundancy-Maximal-Relevance (MRMR). To achieve a good hyper parameter setting, the LOSSO cross-validation, was adopted. The best accuracy was 92.37% when SVM was combined with the novel feature selection technique FCMIM. Other performance metrics used include sensitivity, specificity, and Matthews Correlation Coefficient (MCC).

Singh & Kumar, (2020) also conducted a research using the UCI Cleveland dataset for predicting heart diseases. The ML techniques tested were NB, DT, RF, and KNN. The four classifiers had accuracy values of 88.157%, 80.263%, 86.84%, and 90.789%, respectively. The limitation to their work is that no feature selection technique was used.

(Shah et al., 2020) tested the performance of ML models on UCI Cleveland heart disease dataset for predicting heart diseases. Pre-processing was done to reduce null values and data unbalancing. Python was used for simulation of the SVM, DT, LR, KNN. After training and testing, it was discovered that KNN has higher accuracy of 87%, SVM 83%, DT 79%, and linear regression with 78%.

Muhammad et al., (2020) in their work argued that ML techniques are the most effective methods for predicting heart diseases. Therefore, the research was carried out on the UCI dataset of 504 records. Four feature selection algorithms were applied to remove noise, redundant features and variant feature selection. The Fast-Correlation-Based-Filter (FCBF), MRMR, Relief, and LASSO were used. Also, different ML classifiers such as KNN, DT, RF, LR, NB, Extra Tree (ET), ANN, SVM, AB, and GB were used to enhance the performance of the model. Several metrics were used for the assessment the system performance. These include: Accuracy, specificity, precision and sensitivity. The findings revealed that ET was the best classifier, with a precision of 94.41%, and a precision of 93.36% when paired with the Relief selection approach.

Yadav & Pal, (2020) presented a novel approach on ML procedures for predicting heart diseases. The work tested the performance of the system on UCI repository dataset with 1025 instances and 14 attributes. Four tree-based classification methods, including M5P, Random Tree, and Reduced Error Pruning (REP) and RF ensemble techniques were put to the test. Pearson Correlation, Recursive Features Elimination and LASSO Regularization feature selection techniques were each tested differently on all the tree-based classification methods. The performance of the system was measured based on accuracy, precision and sensitivity. The result showed that RF ensemble classifier with Pearson correlation and LASSO Regularization predicted better with a result value of 99% accuracy.

Tama et al., (2020) in their work "Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble", tested the performance of three ensemble learners: RF, GB, and extreme GB. The metrics adopted in their model were accuracy, F1, and Area Under Curve (AUC). The performances recorded were 98.13%, 96.6%, and 98.7%, respectively

Ahmed et al., (2019) proposed a real-time heart disease detection system using ML techniques. The study was conducted on four different ML algorithms. These algorithms were compared with each other with the aim of selecting the best for heart disease prediction. The selected ML algorithms were: RF, DT, LR, and SVM. To solve the feature selection problem, the relief and univariate feature selections were used. The experimental result proved that the RF classifier is the best among the selected classifier with an accuracy of 94.9%.

Bashir et al., (2019) proposed a model for heart disease prediction. This model performed a heart disease prediction using feature selection by applying various data mining methods. These methods include: NB, LR, DT, RF, and LR-SVM. LR-SVM got the highest accuracy value of 84.85% from the experiments conducted using the UCI heart disease dataset.

Amin et al., (2018) discussed critical risk factors for predicting cardiovascular disease. Neural Network (NN), KNN, DT, GNB, LR, SVM, and an ensemble NB-LR were tested. The comparative analysis of the work by integrating some selected features obtained an accuracy of 87.41% from the hybridization.

Liu et al., (2017) presented an ensemble classification technique tested on heart disease dataset. The feature Selection based on Relief-Function-and-Rough-Set (RFRS), on an ensemble classifier; C4.5 decision tree was implemented.

There are various ML techniques designed for the clinical industry as reviewed in this paper. Several of these have achieved excellent performances as regarded to the prediction and diagnosis of cardiovascular diseases. These various techniques incorporated Bagging and Boosting approaches mostly with feature selection. Again, most of them do not give regards to missing values or overfitting. The goal of the study is to determine the accuracy and error rates of the algorithms when missing values are inputted and to determine other performance measures like Precision and ROC.

3. Feature Selection Techniques

There are numerous approaches to feature selection in the literature, which are classified under the three (3) major groups. A summary of some feature selection techniques with their merits and demerits is presented in Table 2.

Filter Method: This technique is based on a ranking procedure. A threshold is set for all the variables in the dataset. Features are eliminated or accepted based on their relationship to the threshold value. Features ranked below the threshold are being filtered out or rejected. This technique enjoys some advantages like being simple, fast, scalable and independent (Firdaus et al., 2021). While the setbacks to this technique includes ignoring of feature dependencies and lack of interaction with the classifiers. The filter methods include Chi-Square Test, Fisher's Score and Information Gain.

Wrapper Method: The wrapper method works by generating a subset of selected features and evaluates the subset using a supervised learner. Then features are added and subtracted based on the model's output and the model is re-trained again. It uses a greedy approach to form the subsets and evaluates the accuracy of all possible feature combinations. This has helped researchers in solving the problems with the filter method because it interacts with the classification algorithm during training and thus reason over dependencies (Mwadulo, 2016). However, the wrapper method also suffers high computation cost and are relatively complex (Firdaus et al., 2021). Forward Selection, Backward Elimination and Genetic Algorithm are instances of wrapper methods of feature selection.

Embedded Method: This method searches the subset for the features that best fit a particular algorithm when building the classification model (Firdaus et al., 2021). This technique interacts with the classifier while paying attention to feature dependencies. Another merit of the embedded method is that it has better computational complexity and less over-fitting ability than the wrapper (Ang et al., 2016). The features employed here are constrained by the classifier's hypothesis and are not applicable to all classifiers. ANN, SVM-RFE and DTs are the most widely used embedded techniques.

Intrinsic Method: This is known to hybridize many methods. It integrates many feature selection algorithms and takes advantage of each one of their strengths to achieve a better result. The best features of the filter and wrapper methods are combined to produce the best subset. This method handles iterative Machine training processes while keeping minimum computation costs. The drawback to this method is that it only matches certain classifiers. LASSO and Ridge Regression, Particle Swarm Optimization-Support Vector Machine (PSO-SVM) are examples.

Table 2: Summary table of the feature selection techniques.

Approach	Advantages	Limitations
Filter	<ul style="list-style-type: none"> ✓ Independent, computational efficiency, reduced computational time and scalable to a large dataset 	Ignore feature dependencies and interaction with classifiers
Wrapper	<ul style="list-style-type: none"> ✓ Performance accuracy better than that of filter, classifier interactive and feature dependencies are considered 	Cost intensive and high complexity overfitting with small dataset
Embedded	<ul style="list-style-type: none"> ✓ Classifier interaction, feature dependencies, better computational cost and low over-fitting value compared to a wrapper 	Classifier specific and computationally costly than filter
Hybrid	<ul style="list-style-type: none"> ✓ Better performance accuracy than filter, less computation complexity and over-fitting than that of wrapper 	Classifier specific

4. Research Methodology

Figure 1 describes the overall methodology adopted for modelling this intelligent system for predicting heart diseases. The data collection and preprocessing were the first and second phases. The UCI heart dataset was obtained in an open source and feature selection (SelectKbest) used for data preprocessing. Phase 3; At this stage the various classifiers were tested for the prediction of heart disease. In Phase 4, bagging technique was adapted to test the enhancement achieved in the prediction. Phase 5 contains the models' performance based on accuracy, precision, recall and AUC.

4.1 Phase 1

Dataset: The data on heart disease was collected from UCI's public data repository. It contains 303 instances and 13 attributes plus one target class. It is a categorical dataset with some numeric and nominal data type. The dataset contains 139 patients' records that were diagnosed of the heart diseases (true) and 164 records as patients without heart diseases (false).

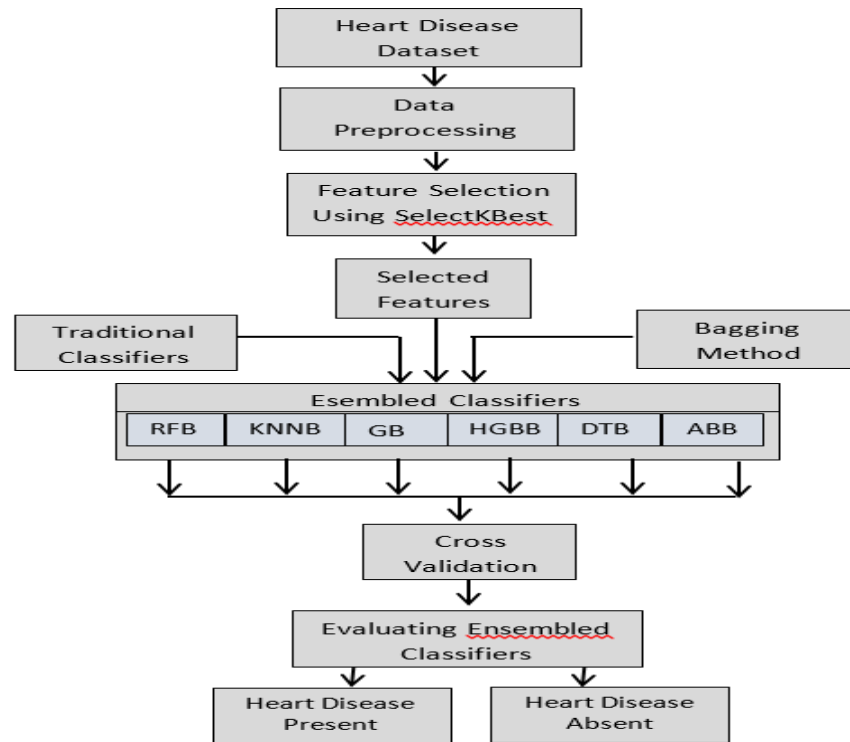


Figure 1: Proposed ensemble architecture for heart disease prediction

4.2 Phase 2

The data reprocessing approach was carried out using the selectKBest feature selection technique. The features with the greatest ratings were maintained, while the features with lower scores were removed. This was done by using the selectKBest feature selection technique. The SelectKBest feature selection technique is a filter-based univariate selection technique as discussed in section 3. Through different univariate statistical tests, features are picked and then a rank ordering mechanism is used to filter out unrelated features in a filter-based feature selection strategy (Srivatsan&Santhanam, 2021). The rankings are based on the statistical scores generated by the features' correlation with the target variable. It is used to select the features with the greatest influence on the output variable. Both numerical and categorical data can be used in this manner (Zulfiker et al., 2021), in which the features with the highest K scores are kept.

To choose K-best features, this study performed univariate statistical tests based on the chi-square(χ^2) test. Other univariate statistical tests, such as Analysis of Variance (ANOVA), F-value, and mutual information approaches can be used to choose K- best features in addition to the χ^2 test. However, because the χ^2 test can only be used on non-negative characteristics, it was used in this investigation. For each non-negative feature and the target feature, it calculates the χ^2 score. Equation (1) can be used to obtain the χ^2 score for n pairs of expected and observed frequencies, as explained by (Zulfiker et al., 2021). Here, OF_i is the frequency which is observed for the i - th value of the feature F and EF_i is the frequency which is expected for the i th value of the feature F .

$$\chi^2 = \sum_{i=1}^n \frac{(OF_i - EF_i)^2}{EF_i} \tag{1}$$

Here the features fbs, sex and slope were removed which were considered to have least K scores. The following is a step-by-step breakdown of the implementation as led by the SelectKBest algorithm:

Step I: Import SelectKBest(), χ^2

Step II Initialize $k = 10$

Step III: Apply the parameters k and χ^2 to SelectKBest()

Step IV: Apply (3) to (X, Y) where X refers to the predictors and Y refers to the target variable.

Step V: An array of scores is returned for each feature.

Step VI: The first 10th k features that have high scores are retained.

4.3 Phase 3

The third phase of the work involves testing of the classifiers this work with bagging. KNN, DT, RF, were considered as the most common techniques in predicting heart diseases, whereas AB, GB and HGB were tagged the least explored techniques (Ghosh et al., 2021). Therefore, this work attempts to explore the performance of these techniques with bagging based on the selected features. Other arguments related to this is that the most popular algorithms are of higher performance (Yadav & Pal, 2020) than the less tested techniques (Bhuvaneswari et al., 2019). Therefore, this work tends to improve the performance across board. This is to devise a solution that will stand as a benchmark in reshaping the current thoughts regarding these techniques. A detailed explanation of each of the selected classifiers is given with the corresponding bagging technique.

a. Traditional Classifiers:

- i. The KNN: is generally regarded as a lazy learner. This is as a result of its ease in training and testing. In this research, the value of $k = 5$, therefore, 5 closest neighbors were considered and the distance between them was calculated using the Euclidean distance metrics as follows. If 'A' has a distance coordinate of A (x_1, y_1) and B has a distance coordinate of B (x_2, y_2) , then the Euclidean distance between A and B is given in equation (2) as follows:

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

The choice of the distance metrics for KNN helps in improving the performance of the model.

- ii. DT: This tree works by first finding the best attribute from the given set, then splits it into branch-like forms and the procedure is being repeated until the best feature is found. The process for selecting the best attribute is by determining the Entropy and Information Gain.
- iii. Random Forest: This employs multiple algorithms to generate the best model and the result therefore is the mean of the total number of algorithms involved. The aggregated bootstraps technique is what enables this process in RF. Now, if the data to be modelled is "X" where $X = \{x_1, x_2, \dots, x_n\}$ and 'Y' is the responses where $Y = \{y_1, y_2, y_3, \dots, y_n\}$, $b = 1$ as the lower limit and B as the upper limit. The prediction for sample x' is the average of the predictions given in equation 3 and 4.

$$\sum_{b=1}^B f_b(x') \text{ for every single tree} \tag{3}$$

$$j = \frac{1}{B} \sum_{b=1}^B f_b(x') \tag{4}$$

The limitation to this technique is that it suffers from high computational cost and complexity as a result of many features involved in it than in the traditional DT.

- iv. Ada Boost: Equation 5 represents the working of the Ada Boost technique.

$$H_k(p) = \left(+ / - \left| \sum_{k=1}^k (a_k h_k) (p) \right. \right) \tag{5}$$

Equation (5) is a linear combination of all the simple learners, where K is the total number of classifiers, $h_k(p)$ is the output of the classifier, t (-1 or 1) and a_k is the weight of the classifier k .

- v. Gradient Boost: The difference between Boosting technique and traditional ML is that the function space excludes optimization. The optimal function $F(X)$ as given in equation (6) is obtained after the m^{th} iterations.

$$F(X) = \sum_{i=0}^m f_i(x) \tag{6}$$

- vi. Hist Gradient Boost: Histogram gradient Boost is almost like the gradient boost. It is commonly referred to as gradient boosting algorithm that supports histograms in modern ML libraries.

b. Bagging Method

Bagging was founded on the concepts of bootstrapping and aggregation. The training dataset in the Bagging classifier is used to build bootstrap datasets. After then, each of these bootstrap datasets is used to train a number of different classifiers. The results of these classifiers are then combined to produce the final forecast. In the bootstrap dataset, misleading training items are typically avoided.(Zulfiker et al., 2021). Furthermore, combining the classifiers frequently yields better results than using a single classifier (Skurichina&Duin, 2002). Because both of these characteristics are integrated in the Bagging classifier, it often outperforms other classifiers. The following are the steps involved in bagging:

Step I: For $n = 1, 2, \dots, N$

- a. Create a bootstrap replicate X^n from the training dataset X .
- b. Using X^n construct a model $C^n(x)$.

Step II: Aggregate the models $C^n(x)$

Step III: End For, where $n = 1, 2, \dots, N$ by majority voting or averaging.

4.4 Phase 4

This section involves testing the system’s performance as an ensemble model. The work integrated the traditional classifiers with bagging technique to test for optimal results. Bagging as a meta-algorithm designed primarily for increasing the stability of algorithms and also for optimal prediction in classification and regression processes.

Bagging was founded on the concepts of bootstrapping and aggregation. The training dataset in the Bagging classifier is used to build the bootstrap datasets, each of these bootstrap datasets is then use in training a number of different classifiers. The results of these classifiers are combined to produce the final forecast. In the bootstrap dataset, misleading training items are typically avoided (Zulfiker et al., 2021). Furthermore, combining the classifiers frequently yields better results than using a single classifier (Skurichina&Duin, 2002). In trying to produce new data for training from the given dataset, bagging minimizes variance and simultaneously avoiding overfitting. Because both of these characteristics are integrated in the Bagging classifier, it often outperforms other classifiers. The following are the steps involve in bagging:

Step I: For $n = 1, 2, \dots, N$

- a. Create a bootstrap replicate X^n , from the training dataset X .
- b. Using X^n construct a model $C^n(x)$.

Step II: Aggregate the models $C^n(x)$ where $n = 1, 2, \dots, N$ by majority voting or averaging.

4.5 Phase 5

Cross-validation is primarily used in the ML to estimate the skills of a model on an unknown data. This is done by taking a sample of the dataset and observing how the model responds. K-fold cross-validation with the choice of $K = 10$ was carried out in this research and the system was evaluated.

The degree of fitness of the models was examined using confusion matrix. Figure 2 represents how the confusion matrix works.

		True class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Figure 2: Confusion Matrix

The green coloration denotes a positive situation in which a patient is not diagnosed with heart disease as a result of training or testing procedures. Meanwhile, the red coloration denotes a negative situation in which a patient is diagnosed with heart disease as a result of training or testing procedures.

Note that Positive prediction occurs when a person is classified as having heart disease and negative when the prediction is not having heart disease. Other measures tested were the accuracy, Precision, Recall and ROC curves. The following formulae were adopted for estimating the performance of the system for heart disease prediction (Shamrat et al., 2020).

TP = True Positive, this is when the model correctly identified a heart disease patient as being positive

TN = True Negative occurs when the model is able to correctly identify a person without heart disease as being negative.

FP = False Positive, this is a misclassification instance where the model classifies a person without heart disease as being positive for the disease.

FN = False Negative is also an incorrect classification. It is an instance where the model tags a positive heart disease patient as being normal or negative. To compare the performance of the system based on Accuracy, Precision and Recall the following formulae from equations (7) to (12) were adopted.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{7}$$

$$\text{Precision} = \frac{(TP)}{(TP+FP)} \tag{8}$$

$$\text{Recall} = \frac{(TP)}{(TP+FN)} \tag{9}$$

$$\text{False Positive Rate} = \frac{(FP)}{(FP+TN)} \tag{10}$$

$$\text{False Positive Rate} = \frac{(FN)}{(TP+FN)} \tag{11}$$

$$\text{Negative Predictive Value} = \frac{(TN)}{(TN+FN)} \tag{12}$$

5. Results Analysis

5.1 Performance of the traditional classifiers:

A model for the various classification algorithms was built on the Cleveland dataset using Jupiter notebook. It is a simulation tool for python programming projects. Some of the traditional algorithms showed promising results while some others were relatively satisfactory. The KNN, GB, HGB and RF had the best accuracy of 90%, 89%, 85% and 85% respectively. This is in relation to the value obtained after computing the outcomes of the prediction obtained from the confusion matrix. Table 3 represents the confusion matrix of the traditional model without hybridizing the six classification algorithms.

Table 3: Confusion matrix for the traditional classifiers

Classifiers	Heart Disease	Predicted Yes	Predicted No
KNN	Does not have	27	2
	Has	4	28
RF	Does not have	26	3
	Has	66	26
DT	Does not have	18	11
	Has	6	26
AB	Does not have	26	3
	Has	7	25
GB	Does not have	27	2
	Has	5	27
HGB	Does not have	26	3
	Has	6	26

After generating a 2×2 confusion matrix on the data model, the different performance metrics such as Accuracy, Precision and Recall, were also obtained. In the like manner, an accuracy value of 85% was obtained for both RF and GB. DT and AB Classifiers generated the lowest accuracies of 72% and 84% respectively. However, the results were significantly better after testing the ensemble classifiers. Table 4 shows the performance of the traditional model before hybridization.

Table 4: Performance evaluation of the traditional models

Classifier	Accuracy	Precision	Recall
KNN	0.90	0.93	0.88
RF	0.85	0.90	0.81
DT	0.72	0.70	0.81
AB	0.84	0.89	0.78
GB	0.89	0.93	0.84
HGB	0.85	0.90	0.81

The relationship between the performance of the six traditional classifiers (accuracy) represented in Table 6 is displayed in Figure 3. From the chart, KNN, GB, HGB and RF were seen to have the highest values and DT stands to be the lowest.

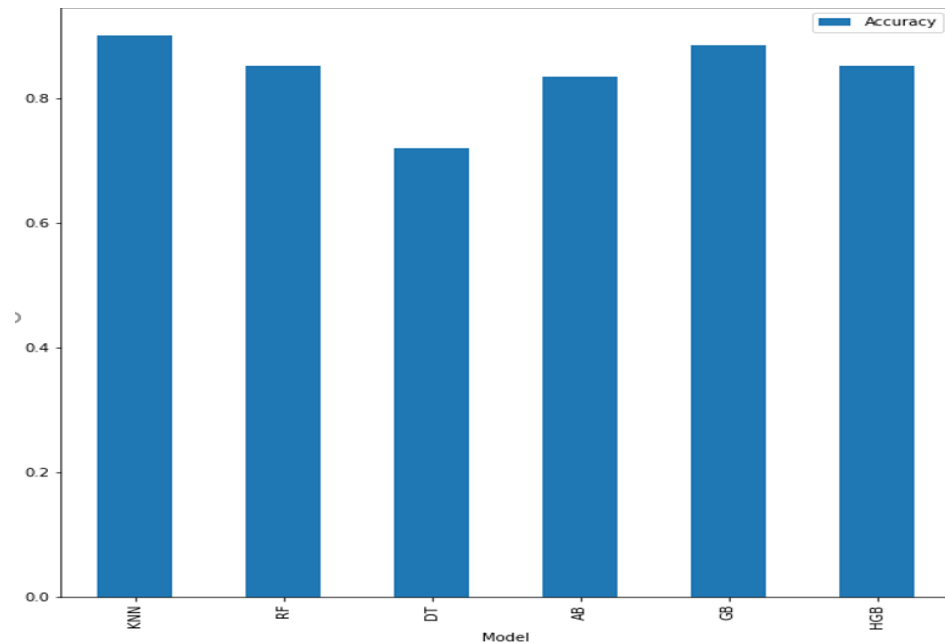


Figure 3: Accuracy of the traditional classifiers

Although accuracy is the most important performance metrics (Ghosh et al., 2021), other measures play significant roles in determining how good models were built. For example, system's sensitivity (Recall) which ensures that the cardiac patients are correctly classified. Both the Precision and Recall of the models were plotted against the accuracy as shown in Figure 4. KNN and GB have the highest precision of 93% while DT has the lowest precision of 70%. KNN and GB recorded the highest recall of 88% and 84% respectively as represented in Figure 4.

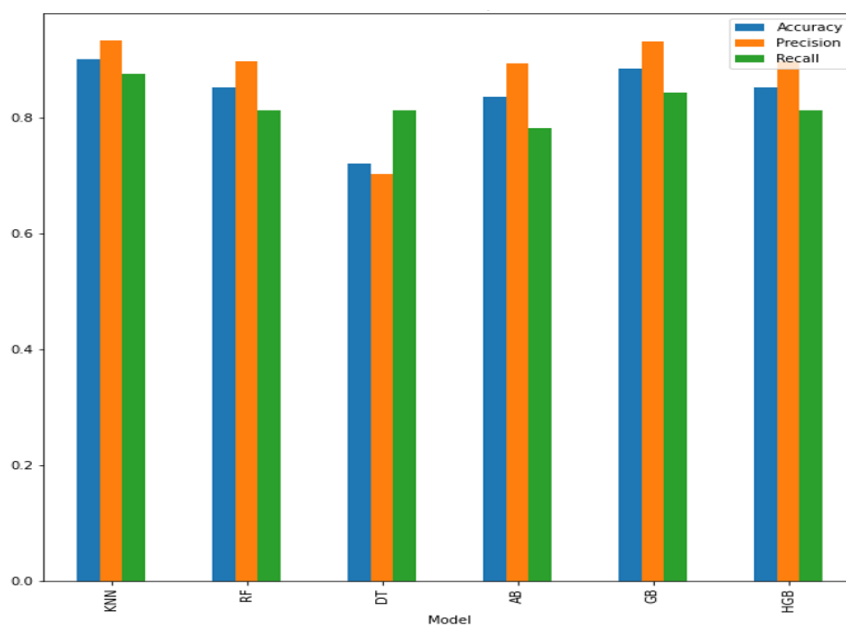


Figure 4: Accuracy, Precision and Recall of the traditional models

A significant improvement was recorded in the overall system in terms of accuracy and other metrics with hybridization.

5.2 Performance of the ensemble classifiers

The selected features were trained with ensembling classifiers. All the six traditional classifiers were ensemble with bagging approach as discussed in phase 3. A 10-fold cross validation was adopted to check for the correctness of the model. Confusion matrix defined by each classifier for the heart disease prediction is shown in Table 5. The performance of the ensemble model when compared to the traditional methods is clearly depicted in Table 5. From the results, the proposed approach reduces the number of cases were healthy patients will be predicted of having heart disease (False Negative) or Patients with heart diseases will be predicted of not having the disease (False Positive).

Table 5: Confusion matrix of the ensembled models

Ensemble Classifier	Heart Disease	Predicted Yes	Predicted No
KNNB	Does not have	27	2
	Has	5	27
RFB	Does not have	27	2
	Has	4	28
DTB	Does not have	26	3
	Has	8	24
ABB	Does not have	25	4
	Has	5	27
GBB	Does not have	25	4
	Has	5	27
HGGB	Does not have	27	2
	Has	4	28

RFB model portrayed a significant impact on illness prediction when compared to the traditional RF model. In which RFB “Has” and “Does not have” of “4 and 27” against “28 and 2” was recorded compared with “66 and 26” against “26 and 3” of the traditional RF. This resulted to the increase in the accuracy of the model. Table 6 depicts the results obtained after hybridizing.

Table 6: Performance of the ensemble model

Classifier	Accuracy	Precision	Recall
KNNB	0.902	0.964	0.844
RFB	0.902	0.933	0.875
DTB	0.852	0.926	0.781
ABB	0.836	0.844	0.844
GBB	0.885	0.931	0.844
HGGB	0.902	0.933	0.875

Table 6 shows that KNNB, RFB and HGGB have the highest accuracy of 90% each followed by GBB with 89% accuracy and DTB with 85%. The least performed model is ABB with an accuracy of 84%. Figure 4 gives the graphical representation of the individual classifier’s accuracies.

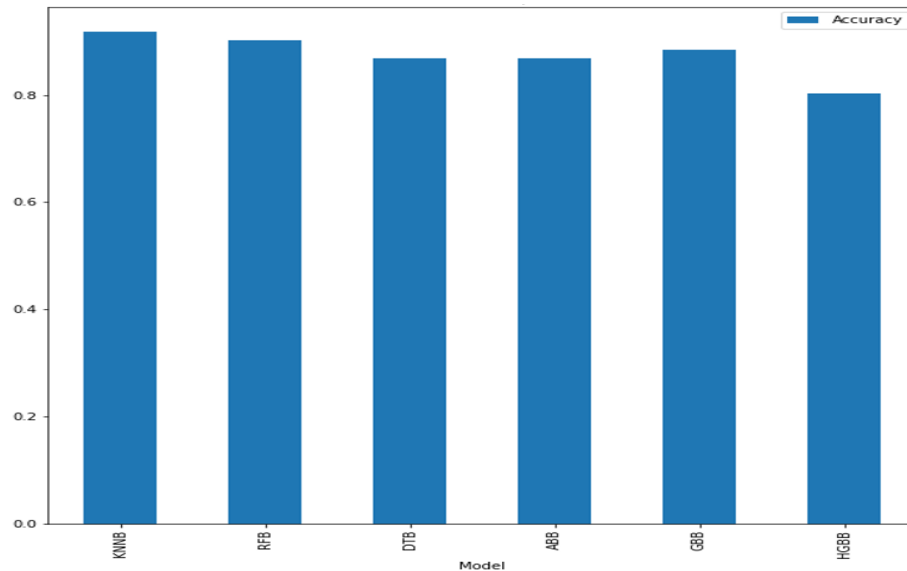


Figure 5: Accuracies of the ensemble models

Also, from the performance in Table 6, it is clear that all the classifiers showed a remarkable increase in accuracy after bagging and feature selection with the exception of ABB. This is graphically displayed in Figure 6.

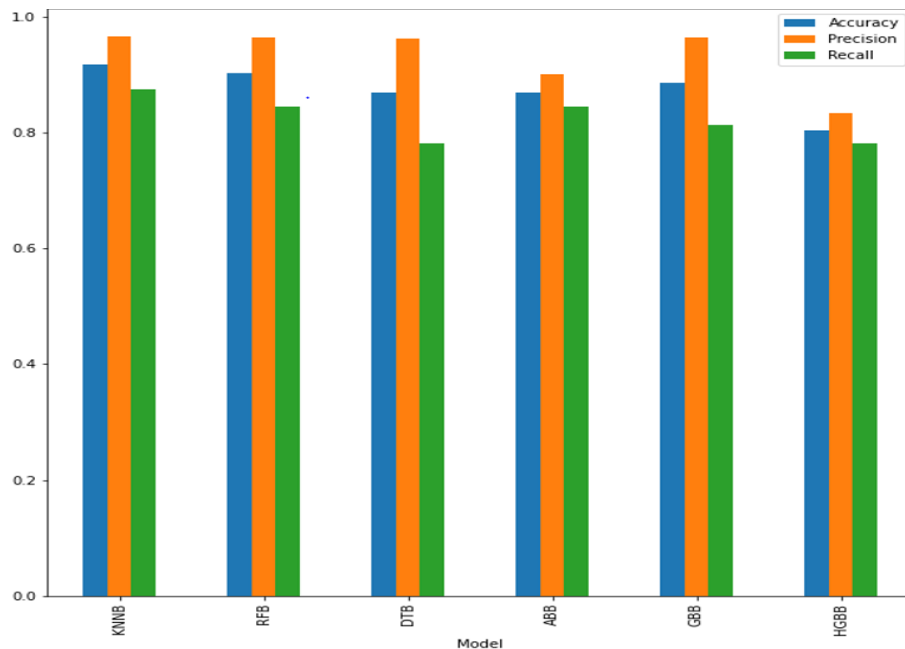


Figure 6: Accuracy, Recall and Precision of the ensemble models.

KNNB scored the highest precision of 96 %, followed by RFB and HGGB, all at 93%, and DTB, 92%. ABB had the lowest precision score of 84%. Similarly, the best recall score of approximately 87% was obtained by RFB

and HGBB classifiers. However, DTB obtained a poor recall of 78%.The results obtained have been compared with existing research of (Yusuf et al., 2021)on the heart disease datasets to show the percentage of improvement achieved.

5.3 Comparison of the proposed model.

The proposed work is an extension of (Yusuf et al., 2021). In (Yusuf et al., 2021), each of the six classifiers were bagged and cross validated using 5-fold cross validation. In the proposed work, important features were selected using selectKbest technique before bagging. A 10-fold cross validation technique was also adopted to test the correctness of the model. Table 7 shows the comparative results obtained from the two models.

Table 7: Comparative results of the two models

Classifier	Proposed model			Yusuf et al., 2021		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
KNNB	0.902	0.964	0.844	0.89	0.93	0.84
RFB	0.902	0.933	0.875	0.90	0.93	0.88
DTB	0.852	0.926	0.781	0.82	0.89	0.75
ABB	0.836	0.844	0.844	0.85	0.87	0.84
GBB	0.885	0.931	0.844	0.85	0.87	0.84
HGBB	0.902	0.933	0.875	0.90	0.93	0.88

The result in Table 7shows the relationship between the proposed model and the work of (Yusuf et al., 2021). The achievement of the proposed work is with KNNB where an accuracy of 90% and a precision of 96% was obtained. For other techniques like RFB, the performance shows a steady increase in accuracy ranging between 90 and 93%.

6. Conclusion

Data analysis for disease prediction or investigation in ML has since grown worldwide in the medical field. This is no different from the fight against heart diseases. Early detection of certain risk factors of heart disease with a significant level of accuracy will not only ease the work of medical practitioners, but also curtail the long-term mortality rate. Different researchers have attempted to predict heart disease using ML techniques for better performances. This research takes no different path. It extended the work of (Yusuf et al., 2021)with the aim of testing for the performance of the models after selecting certain features and dropping the less relevant ones. The results recorded has shown that feature selection is an effective tool in enhancing the performance of ML techniques by selecting features with the highest impact on the target variable. The performance of the ensemble approaches on this model together with feature selection as applied on the heart disease datasets has improved remarkably compared to the work of (Yusuf et al., 2021). GBB was the most improved when only the most relevant features were used, while ABB recorded the least. KNNB, RFB, DTB all recorded a steady increase in their accuracies. This work therefore can be used by medical professionals in detecting heart disease in new patients and for other researchers to build on. For future work, meta-heuristics techniques, or any other relevant techniques can be considered. Moreover, different datasets can be used together to increase the sampling size for a more robust system. To make the model more useful, the application of deep learning can also be examined.

Reference

- Abdollahi, J., & Nouri-Moghaddam, B. (2021). Feature selection for medical diagnosis: Evaluation for using a hybrid Stacked-Genetic approach in the diagnosis of heart disease. 11.
- Ahmed, H., Younis, E. M. G., Hendawi, A., & Ali, A. (2019). Heart Disease Identification from Patients' Social Posts, Machine Learning Solution on Spark. *Future Generation Computer Systems*.
<https://doi.org/10.1016/j.future.2019.09.056>
- Amin, M. S., Chiam, Y. K., & Varathan, K. D. (2018). Identification of Significant Features and Data Mining Techniques in Predicting Heart Disease. *Telematics and Informatics*.
<https://doi.org/10.1016/j.tele.2018.11.007>
- Ang, J. C., Mirzal, A., Haron, H., & Hamed, H. N. A. (2016). Supervised, Unsupervised, and Semi-Supervised Feature Selection: A Review on Gene Selection. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 13(5), 971–989. <https://doi.org/10.1109/TCBB.2015.2478454>
- Bashir, S., Khan, Z. S., Hassan Khan, F., Anjum, A., & Bashir, K. (2019). Improving Heart Disease Prediction Using Feature Selection Approaches. 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), 619–623. <https://doi.org/10.1109/IBCAST.2019.8667106>
- Bhuvaneeswari, R., Sudhakar, P., & Prabakaran, G. (2019). Heart Disease Prediction Model based On Gradient Boosting Tree (GBT) Classification Algorithm. *International Journal of Recent Technology and Engineering (IJRTE)*, 8(2), 41–51. <https://doi.org/10.35940/ijrte.B1008.0982S1119>
- CDCP. (2020). Heart Disease Facts. <https://www.cdc.gov/heartdisease/index.htm>
- Dissanayake, K., & Johar, G. (2021). Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms. *Applied Computational Intelligence and Soft Computing*, 2021, 1–17. <https://doi.org/10.1155/2021/5581806>
- Firdaus, F. F., Nugroho, H. A., & Soesanti, I. (2021). A Review of Feature Selection and Classification Approaches for Heart Disease Prediction. *IJITEE (International Journal of Information Technology and Electrical Engineering)*, 4(3), 75. <https://doi.org/10.22146/ijitee.59193>
- Ghosh, P., Azam, S., Jonkman, M., Karim, A., Mehedi, F. M. J. S., Ignatious, E., Shultana, S., Beeravolu, A. R., & Boer, D. F. (2021). Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques. *IEEE Access*, 9, 19304–19326. <https://doi.org/10.1109/ACCESS.2021.3053759>
- Hasan, N., & Bao, Y. (2021). Comparing different feature selection algorithms for cardiovascular disease prediction. *Health and Technology*, 11(1), 49–62. <https://doi.org/10.1007/s12553-020-00499-2>
- Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques. *IEEE Access*, 9, 39707–39716. <https://doi.org/10.1109/ACCESS.2021.3064084>
- Li, J. P., Haq, A. ul, Din, S. ud, Khan, J., Khan, A., & Saboor, A. (2020). Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare. *IEEE Access*, 8, 107562–107582. <https://doi.org/10.1109/ACCESS.2020.3001149>
- Liu, X., Wang, X., Su, Q., Zhang, M., Zhu, Y., Wang, Q., & Wang, Q. (2017). A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method. *Computational and Mathematical Methods in Medicine*, 2017.

- Muhammad, Y., Tahir, M., Hayat, M., & Chong, K. T. (2020). Early and accurate detection and diagnosis of heart disease using intelligent computational model. *Scientific Reports*, 10, 1–17. <https://doi.org/10.1038/s41598-020-76635-9>
- Mwadulo, M. W. (2016). A Review on Feature Selection Methods For Classification Tasks. *International Journal of Computer Applications Technology and Research*, 5(6), 395–402. <https://doi.org/10.7753/IJCATR0506.1013>
- Shah, D., Patel, S., & Kumar, S. (2020). Heart Disease Prediction using Machine Learning Techniques. *SN Computer Science*, 1(6), 1–6. <https://doi.org/10.1007/s42979-020-00365-y>
- Shamrat, F. M. J. M., Raihan, A., Rahman, A. K. M. S., Mahmud, I., & Akter, R. (2020). An Analysis On Breast Disease Prediction Using Machine Learning Approaches. *International Journal of Scientific and Technology Research*, 9(02), 2450–2455.
- Singh, A., & Kumar, R. (2020). Heart Disease Prediction Using Machine Learning Algorithms. *International Conference on Electrical and Electronics Engineering (ICE3-2020)*, 452–457.
- Skurichina, M., & Duin, R. P. W. (2002). Bagging, Boosting and the Random Subspace Method for Linear Classifiers. *Pattern Analysis & Applications*, 5(2), 121–135. <https://doi.org/10.1007/s100440200011>
- Soni, J., Ansari, U., & Sharma, D. (2011). Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction. *International Journal of Computer Applications*, 17(8), 43–48.
- Srivatsan, S., & Santhanam, T. (2021). EARLY ONSET DETECTION OF DIABETES USING FEATURE SELECTION AND BOOSTING TECHNIQUES. *ICTACT JOURNAL ON SOFT COMPUTING*, 12(01), 12.
- Tama, B. A., Im, S., & Lee, S. (2020). Improving an Intelligent Detection System for Coronary Heart Disease Using a Two-Tier Classifier Ensemble. *BioMed Research International*, 2020. <https://doi.org/10.1155/2020/9816142> Research
- Yadav, D. C., & Pal, S. (2020). Prediction of Heart Disease Using Feature Selection and Random Forest Ensemble Method. *International Journal for Pharmaceutical Research Scholars*, 12(4), 56–66. <https://doi.org/10.31838/ijpr/2020.12.04.013>
- Yekkala, I., & Dixit, S. (2018). Prediction of Heart Disease Using Random Forest and Rough Set Based Feature Selection. *International Journal of Big Data and Analytics in Healthcare*, 3(1), 1–12. <https://doi.org/10.4018/IJBDAH.2018010101>
- Yusuf, A. B., Baththama, B. A., Wasilah, S., Fatsuma, J., Maria, A., & Kiyee, C. (2021). Improving Heart Disease Prediction Using Hybrid Based Technique. *International Conference on Computing and Advances in Information Technology (ICCAIT 2021) ABU, Zaria*, 9.
- Zulfiker, Md. S., Kabir, N., Biswas, A. A., Nazneen, T., & Uddin, M. S. (2021). An in-depth analysis of machine learning approaches to predict depression. *Current Research in Behavioral Sciences*, 2, 100044. <https://doi.org/10.1016/j.crbeha.2021.100044>