IJCSN - International Journal of Computer Science and Network, Volume 9, Issue 3, June 2020
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

MFES-NB Framework for Detection of Heart Disease using Data Mining Technique

[1]Abba Babakura; [2]Mahmud Ahmad Yusuf; [3]Baba Yachilla Alhaji

[1] Department of Computer Science, Usmanu Danfodio University Sokoto,
Sokoto 840231, Nigeria

[2] School of Computing and Engineering, University of Huddersfield,
Huddersfield HD1, United Kingdom

[3] Department of Electrical and Electronics Engineering,
Nile University of Nigeria, Abuja 900211, Nigeria

Abstract - Huge amount of data containing information deemed to be useful for effective decision making are collected and stored by health care industries. The heart disease (HD) has been considered one of the complex and deadliest human diseases in the world. An accurate and timely diagnosis of heart disease is crucial for heart disease prevention and treatment. Machine learning algorithms have proven to be crucial in providing the desired solutions. However, it suffers from finding the best combination of features that improves the classification accuracy of heart disease and maintain a balanced feature selection. In this paper, we propose new machine learning based diagnosing framework that provides an improved feature selection approach and classification algorithms for the prediction of heart disease. The framework combines Multiple Feature Evaluation System (MFES) and Naïve Bayes (NB) algorithm for the experimentation on the heart disease dataset. The result showed that the MFES selects the best features and also, improves the performance of the classification algorithm. In addition, the NB algorithm results showed an improvement in the prediction of accuracy of the heart disease. The framework will assist the doctors in the diagnosing of patients efficiently.

***Keywords** - Heart disease, Naïve Bayes algorithm, Feature selection, Classification*

.

1. Introduction

The heart disease (HD) has been considered as one of the complex and deadliest human diseases in the world. Usually, the heart failure occurs when the heart is unable to push the required amount of blood to other parts of the body to fulfill its normal functionalities [1]. The investigation techniques in early stages used to identify heart disease were complicated, and its resulting complexity is one of the major reasons that affect the standard of life [2]. The heart disease diagnosis and treatment are very complex, especially in the developing countries, due to the rare availability of up-to-date diagnostic apparatus and shortage of physicians (cardiologists) and other resources which affect proper prediction and treatment of heart patients [3]. Due to that, accurate and proper diagnosis of the heart disease risk in patients is necessary for reducing their associated risks of severe heart issues and improving security of heart [4].

The efficiency of data mining largely varies on the techniques used and the features selected. The medical datasets in the healthcare industry are redundant and inconsistent. It is harder to use data mining techniques without prior and appropriate preparations. According to Kavitha and Kannan [5], data redundancy and inconsistency in a raw dataset affect the predicted outcome of the algorithms. As a result, to apply the machine learning algorithms to its full potential, an effective preparation is needed to preprocess the datasets. Furthermore, unwanted or irrelevant features can reduce the performance of the data mining techniques as well [6]. Thus, along with data preparation, a proper feature selection method is needed to achieve high accuracy in heart disease prediction using significant features and data mining techniques.

The performance of data mining techniques used in predicting cardiovascular disease is greatly reduced or shortens without a good combination of key features and also the improper use of the machine learning algorithms [7]. Thus, it is crucial to identify the best combination of significant features that works incredibly well with the best performing algorithm for the attainment of desired output.

In this paper, we propose a machine learning based framework that identifies significant features combined with a classification algorithm to accurately predict heart disease. Our framework adaptively employs multiple feature evaluation system (MFES) and Naïve Bayes algorithm to accurately predict the heart disease. The goal is to reduce the number of insignificant features, as well as reduce the time taken to build a model and improve the classification accuracy performed by the algorithm.

This paper is organized as follows: Section 2 briefly reviews related works, and Section 3 describes our proposed framework. Section 4 presents the performance evaluations and we summarize in Section 5.

2. Related Works

Machine learning literature introduces a number of data analytics and learning classifiers [8][9]. The classifiers have been utilized to solve complex and critical classification problems including Computer-Aided Diagnosis (CAD) applications [10]. This section presents several works that attempt to identify significant features and apply classification methods for heart disease diagnosis.

A hybrid intelligent machine learning based predictive system for the diagnosis of heart disease was proposed by [11]. In this proposed predictive system, classifiers such as Decision tree, Artificial Neural Network, K-Nearest Neighbor, Support Vector Machine, Naive Bayes, Logistic Regression and Random forest are used for the experiment along with three feature selection algorithms namely; Relief, mRMR and LASSO. Cleveland dataset adopted from the machine learning repository was used to test the system. K-fold cross validation technique was adopted to evaluate the performance of the classifiers. The experimental result revealed that Logistic Regression Classifier with relief FS algorithm emerged best with an accuracy score of 89%.

Hybrid classification system based on ReliefF and RS (RFRS) to diagnose heart disease was proposed by [12]. The RFRS was used to carefully select the features efficient enough for the prediction. It uses an ensemble classifier to evaluate the feature selection model to improve the efficiency and effectiveness of the classification performance for the diagnosis of the heart disease. A maximum classification accuracy score of 92.59% was recorded with jackknife cross validation. However, the number of K-nearest Neighbor (K) and threshold (theta) of ReliefF parameters are not stable and needs to be optimized for better feature selection.

An intelligent heart disease prediction system based on genetic algorithm and optimized neural network using some risk factors was proposed [13]. The system utilizes genetic algorithm to optimize initial weight of the neural network to improve its performance. There by making the system to learn faster, more stable and accurate compared to back propagation algorithm. The network uses risk factors of 50 patients collected and experiment were conducted on them. The system resulted in an accuracy score of 96.2%.

Another study presented was an enhanced prediction system of heart disease with feature subset selection using genetic algorithm [14]. The system uses Naive Bayes, Decision Tree, and Clustering classification with genetic algorithm optimally selecting the features. A total of 909 datasets with 13 attributes were used initially. Later, the attributes were reduced to 6 using genetic search and experiments were conducted on all the classifiers. From the result, it was observed that Decision tree with 6 attributes outperformed the other two techniques with 99.2% prediction accuracy score, 0.09s of construction time and mean square error of 0.00016 respectively.

A study was conducted which explains the various data mining techniques deployed in an automated system [15]. The system aimed at reducing the test conducted by patients, as well as saving cost and time for both the analyst and the patients. It considered Naive Bayes, Decision Tree and Neural Network for the experimentation. Based on the experimental result, it is observed that neural network with 15 attributes outperformed the Decision tree and Naïve Bayes with the accuracy score of 100% as against 99.62 and 90.74% obtained from decision tree and Naive Bayes respectively.
The obtained classification results are considerably high. Others went farther and apply feature evaluation methods prior to the classification process and reduce the features of the dataset. The results show that features reduction in general increases the accuracy of the classifiers and reduces computational time for building a model. However, applying combinations of feature evaluation methods might over-reduce or overweight the features.

3. Research Framework

This study intends to find the best combination of features that improve the heart disease classification accuracy results and maintain balanced feature selection. It includes applying multiple feature evaluation and classification methods for improved diagnoses of heart disease. Here, the framework consists of data preparation, class balancing, feature evaluation, methods setting, classification, and classifiers evaluation as shown in figure 1. The work encompasses processing the original heart disease dataset and outcomes a new filtered dataset. The dataset preparation activities include performing a number of preprocessing steps (data cleaning, data filtration, data transformation and data visualization) to check the quality of the data and scanned for feature selection. The feature evaluation includes employment of MFES that combines selected feature evaluation and ranking algorithms to weight the worth of features and select the most significant features. The evaluation outcomes are the heart disease filtered dataset which contains 14 features. The methods for setting activities include implementing the Naïve Bayes algorithm (NB). The classifiers are evaluated by applying the stratified 10-fold cross-validation (to prevent randomly selection of accuracy score) technique to the dataset before and after feature evaluation. The main activities and methods of this work are detailed in the following sections.

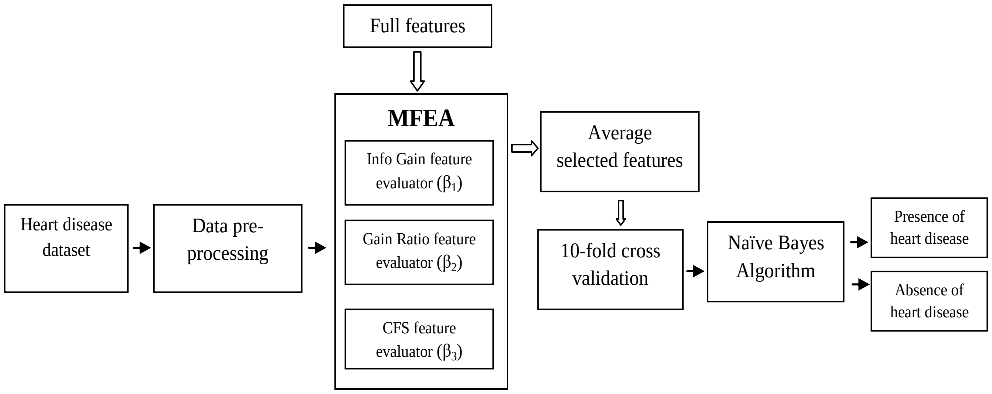

Full features
MFEA
Info Gain feature evaluator (β_1)
Gain Ratio feature evaluator (β_2)
CFS feature evaluator (β_3)
Average selected features
Heart disease dataset
Data pre-processing
10-fold cross validation
Naïve Bayes Algorithm
Presence of heart disease
Absence of heart disease

Fig 1 MFES-NB framework

3.1 Data Description

The preprocessing of data is necessary for efficient representation of data and machine learning classifier which should be trained and tested in an effective manner. At this stage, missing values, data duplications and erroneous records are removed, standard scalar are applied to ensure that every feature has a mean 0 and variance 1, bringing all features to the same coefficient. They are further scanned for feature evaluation. After the data are processed, the full features are now transferred to the MFES for evaluation as seen in the next stage. The dataset contains 303 records with a total of 76 attributes. For this study, a dataset with subset of 14 attributes are obtained from the repository. Table 1 describes the parameters and the associated values with 14 attributes that features in heart disease prediction and one attribute serves as the output or the predicted attribute for the presence or absence of heart disease in a patient.

Table 1: Description of 14 used parameters [16].

S/N	**Parameters**	**Parameter description**	**Values**
1	Age	Age in years	Continuous
2	Sex	Male or female	1= male 0= female
3	Cp	Chest pain type	1= typical angina 2= atypical angina 3= non-anginal pain 4= asymptomatic
4	trestbps	Resting blood pressure	Continuous value (in mm Hg)
5	chol	Serum cholesterol	Continuous value (in mg/dl)
6	Fbs	Fasting blood sugar	1= true 0= false
7	Restecg	Resting electrocardiographic results	0= normal 1= having ST-T wave abnormality 2= definite left ventricular hypertrophy
8	thalach	Maximum heart rate achieved	Continuous value
9	Exang	Exercise induced angina	0= no 1= yes
10	oldpeak	ST depression induced by exercise relative to rest	Continuous value
11	Slope	Slope of the peak exercise ST segment	1= unsloping 2= flat 3= downsloping
12	ca	Number of major vessels colored by fluoroscopy	0–3 value
13	Thal	Defect type	3= normal defect 6= fixed defect 7= reversable defect

14	Num	Diagnosis of heart disease	0: absence of heart disease 1: presence of heart disease

3.2 Class Balancing

This is another important aspect of data preprocessing phase, especially in the heart disease dataset because sometimes you can end up having high percentage of accuracy score with a bad model. For example if 80% of all the observations or instances belong to class presence of heart disease and 20% belong to class absence of heart disease, hence we could say almost everything belong to class presence of heart disease, but we were able to deal with this problem using class balancing (i.e under-sampling technique).

3.3 Feature Evaluation

Multiple feature evaluation system (MFES) is an intelligent agent-based system that jointly solves problems. The system is designed in such a way that it increases the flexibility by segregating its functionality and enabling interaction with its modules during runtime. The MFES encompasses three (3) agents in which each agent operates particular feature evaluation methods. The agents evaluate and rank the features and produce subsets of feature vectors via implementing search algorithm. The process includes weighting each feature and ordering subsets of the feature according to their individual evaluation results. The agents based on the subsets of features perform feature filtering process and produce a number of preliminary copies of feature vectors. Then they collaborate with each other to generate an optimized feature vector. The three agent-based feature evaluation operators are described as follows:

- The first agent, μ_1, operates an Info Gain feature evaluator, β_1. The β_1 evaluates the worth of a subset of features by measuring the information gained with respect to the corresponding class.

- The second agent, μ_2, operates a Gain Ratio feature evaluator, β_2. The β_2 evaluates the worth of a subset of features by measuring the gain ratio with respect to the corresponding class.

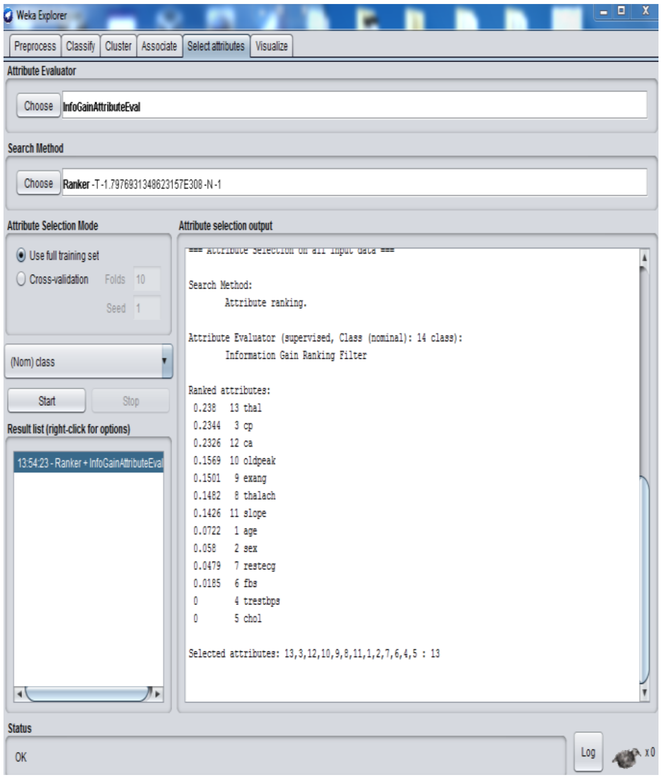
- The third agent, μ_3, operates a CFS feature evaluator, β_3. The β_3 evaluates the worth of a subset of features by measuring the predictive ability of each feature and the redundancy degree between all the features.

The agents collaborate to measure the frequency of the feature's appearance and the mean score value of the features rank. The selected features are the features that have a higher frequency of appearance and lower mean rank values. The agent then combines subsets of the selected features to form the filtered feature vector. This approach ensures selecting the best features that have higher ranks and at the same time prevents overweighting the features. It eliminates the nominal features and produces a feature vector that fit different types of classifiers. All the operators might differently rank the features and considerately omit the weak features. Hence, the MFES is meant to improve the classification process and at the same time avoid the global minimum learning of the classifiers that results from over-fitting some of the classified features.

Table 2 shows the evaluation of ranked features by each agent. Considering the features of the dataset, each of the agents generates its own evaluation of the features, μ_1: X → x_1» - - -»x_n in which the » denotes the permutation process of X. Figure 2, 3 & 4 describes the features evaluated by the three agents.

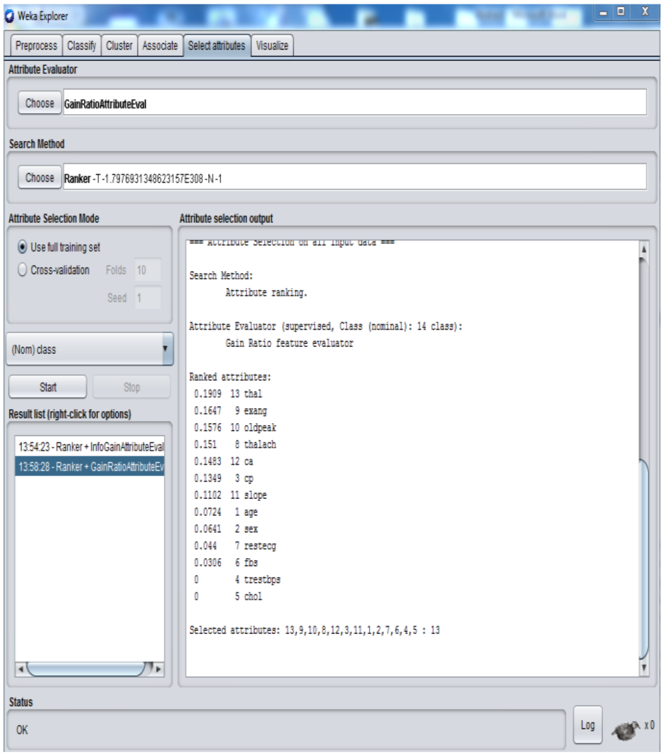
Table 2: Evaluation of features ranked by agents

I	Info gain:μ_1 (β_1)	Gain ratio:μ_2 (β_2)	CFS:μ_3 (β_3)
1	13(0.238)	13(0.1909)	3
2	3(0.2344)	9(0.1647)	6
3	12(0.2326)	10(0.1576)	7
4	10(0.1569)	8(0.151)	8
5	9(0.1501)	12(0.1483)	9
6	8(0.1482)	3(0.1349)	10
7	11(0.1426)	11(0.1102)	11
8	1(0.0722)	1(0.0724)	12
9	2(0.058)	2(0.641)	13
10	7(0.0479)	7(0.044)	
11	6(0.0185)	6(0.0306)	
12	4(0)	4(0)	
13	5(0)	5(0)	



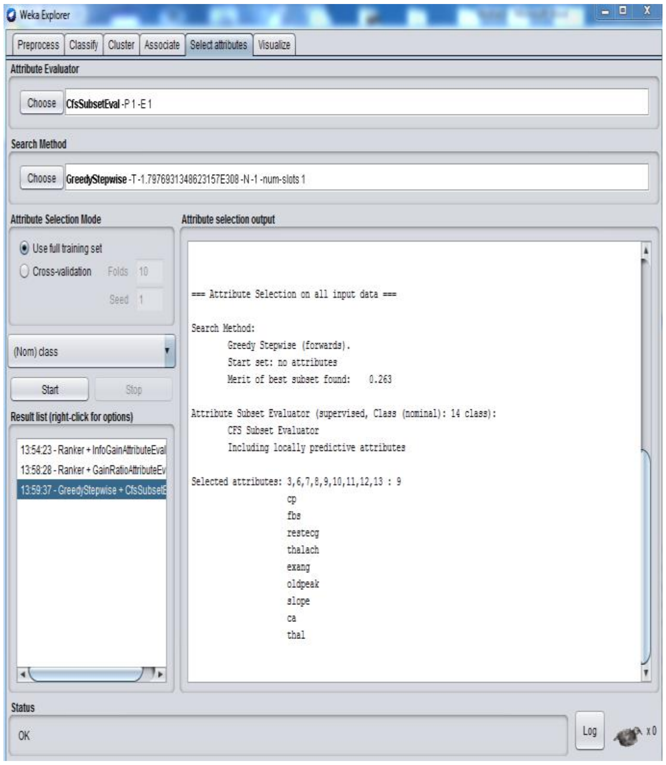
```
Attribute Evaluator
Choose  InfoGainAttributeEval

Search Method
Choose  Ranker -T -1.7976931348623157E308 -N -1

Attribute Selection Mode
● Use full training set
○ Cross-validation   Folds 10
                     Seed 1
(Nom) class

Result list (right-click for options)
13:54:23 - Ranker + InfoGainAttributeEval

Attribute selection output
=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 class):
        Information Gain Ranking Filter

Ranked attributes:
 0.238   13 thal
 0.2344   3 cp
 0.2326  12 ca
 0.1569  10 oldpeak
 0.1501   9 exang
 0.1482   8 thalach
 0.1426  11 slope
 0.0722   1 age
 0.058    2 sex
 0.0479   7 restecg
 0.0185   6 fbs
 0        4 trestbps
 0        5 chol

Selected attributes: 13,3,12,10,9,8,11,1,2,7,6,4,5 : 13

Status
OK
```

Fig. 2 Info Gain ranked features



```
Attribute Evaluator
Choose  GainRatioAttributeEval

Search Method
Choose  Ranker -T -1.7976931348623157E308 -N -1

Attribute Selection Mode
● Use full training set
○ Cross-validation   Folds 10
                     Seed 1
(Nom) class

Result list (right-click for options)
13:54:23 - Ranker + InfoGainAttributeEval
13:58:28 - Ranker + GainRatioAttributeEv

Attribute selection output
=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 class):
        Gain Ratio feature evaluator

Ranked attributes:
 0.1909  13 thal
 0.1647   9 exang
 0.1576  10 oldpeak
 0.151    8 thalach
 0.1483  12 ca
 0.1349   3 cp
 0.1102  11 slope
 0.0724   1 age
 0.0641   2 sex
 0.044    7 restecg
 0.0306   6 fbs
 0        4 trestbps
 0        5 chol

Selected attributes: 13,9,10,8,12,3,11,1,2,7,6,4,5 : 13

Status
OK
```

Fig 3 Gain Ratio ranked features



```
Attribute Evaluator
Choose  CfsSubsetEval -P 1 -E 1

Search Method
Choose  GreedyStepwise -T -1.7976931348623157E308 -N -1 -num-slots 1

Attribute Selection Mode
● Use full training set
○ Cross-validation   Folds 10
                     Seed 1
(Nom) class

Result list (right-click for options)
13:54:23 - Ranker + InfoGainAttributeEval
13:58:28 - Ranker + GainRatioAttributeEv
13:59:37 - GreedyStepwise + CfsSubsetE

Attribute selection output
=== Attribute Selection on all input data ===

Search Method:
        Greedy Stepwise (forwards).
        Start set: no attributes
        Merit of best subset found:    0.263

Attribute Subset Evaluator (supervised, Class (nominal): 14 class):
        CFS Subset Evaluator
        Including locally predictive attributes

Selected attributes: 3,6,7,8,9,10,11,12,13 : 9
                     cp
                     fbs
                     restecg
                     thalach
                     exang
                     oldpeak
                     slope
                     ca
                     thal

Status
OK
```

Fig. 4 CFS ranked features

The collaborative mechanism of the multi-agent system comprises a filter function that receives the X and determines the included feature, x^{in}, from the excluded feature, x^{ex} of the X.

$$filter(X) = \begin{cases} x_{k1}^{in}, & f_i \geq \frac{1}{n}\sum_{i=1}^{n} f_i \wedge rank\left(x_i, \frac{1}{f_i}\sum_{j=1}^{f_i} r_{i,j}\right) \leq t \\ x_{k2}^{ex}, & otherwise \end{cases} \qquad (1)$$

where x_i is a feature, n is the total number of features of the X and i = (1, 2, - - - n) in which $x^{in} + x^{ex} = X$; k1and k2 are the indexes for the x^{in} and x^{ex} respectively; f_i is the appearance frequency of x_i and $r_{i;j}$ is the rank of xi by a **β**j in which the *j* is the index of the corresponding feature evaluator agent; rank is a function that returns a feature's referenced rank as an integer value and t is the threshold of the required number of the x^{in}. Furthermore, the selected best features are used to feed the Naïve Bayes classifier. The features are used to train the model whereby the classifier is built to perform the prediction of either presence of heart disease or absence of heart disease.

3.4 Classification Method

In Naive Bayesian classifier approach, it uses a generative model based on Bayes rule that can predict probabilities of class membership. For instance, it predicts the probability of a given sample belonging to a particular class [17].

The Naïve Bayesian classifier is based on the class conditional independence. This means that an attribute value on the class is independent of other attribute values. It reduces the time involved in the computation of the probability of attribute and class and because of this; it is labeled as “naive”.

According to the Naive Bayes theorem, the computation probability of P(Y|X) can be formulated in terms of probabilities P(X|Y), P(Y), and P(X) as:

$$P(Y|X) = \frac{P(X|Y) * P(Y)}{P(X)} \qquad (2)$$

where:

- X is considered attributes i.e the instances from the data;
- Y is any class attribute i.e the predicting class (presence or absence of heart disease)
- P(Y|X) is the a posteriori probability of Y conditioned on X.
- P(Y) is the a priori probability of Y. A posteriori probability P (Y|X) is based on more information than the a priori probability P(Y), which is independent of X.
- Likewise, P (X|Y) is the a posteriori probability of X conditioned on Y.
- P(X) is the a priori probability of X.

The expression of Y probability that takes over the kth possible value, according to the assumption and Bayes rule, is:

$$P(Y = y_k | x_1 \ldots x_n) = \frac{P(Y = y_k) \prod_i P(x_i | Y = y_k)}{\sum_j P(Y = y_j) \prod_i P(x_i | Y = y_j)}. \qquad (3)$$

To calculate the probability hypothesis given the training data, maximum a posterior hypothesis is used which does not depend on the denominator y_k as follows:

$$Y_{MAP} \leftarrow \underset{y_k}{\operatorname{argmax}} P(Y = y_k) \prod_i P(x_i | Y = y_k). \qquad (4)$$

4. Performance Evaluation

To validate the proposed model, the heart disease datasets were tested which is obtained from [16]. Therefore, to evaluate the performance, the prediction mechanism is derived from the MFES-Naive Bayesian framework, which is measured based on the accuracy and confusion matrix of heart disease predicted results. A stratified 10-fold cross validation is used to train the model and therefore, average value is calculated. The accuracy and confusion matrix are very essential to ensure accurate diagnosis of the heart disease.

The overall accuracy is calculated using the equation:

$$\text{Overall accuracy (ACC)} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (5)$$

A confusion matrix contains information about real and predicted classifications done by a classification system. The data in the matrix are evaluated to know the performance of such systems. The confusion matrix really provide essential picture of how well a classifier algorithm is doing because it helps us to visualize and understand if one class is performing well or not. The confusion matrix contains the following four entries:

- TP (true positive): The number of records classified as true while they were actually true.
- FP (false positive): The number of records classified as true while they were actually false.
- FN (false negative): The number of records classified as false while they were actually true.
- TN (true negative): The number of records classified as false while they were actually false.

All the experiments were conducted in Weka 3.8.0 software with Intel Core-i7, 2.20 GHz CPU and 4 GB RAM.

Table 3, shows the accuracy performance and error rate of the predicted results. It was observed that framework resulted in 0.72% accuracy and an error rate 0.28%. This result is a huge success because of the fact that the features selection agents are been hybridized and also, the proposed model is quick in terms of adaptability and scalability.

Table 3: Accuracy and Error rate results

Accuracy percentage of predicted heart disease		
	Accuracy (%)	Error rate (%)
Values	0.72	0.28

Figure 5 and 6 shows the resulting accuracy and confusion matrix of the predicted heart disease.

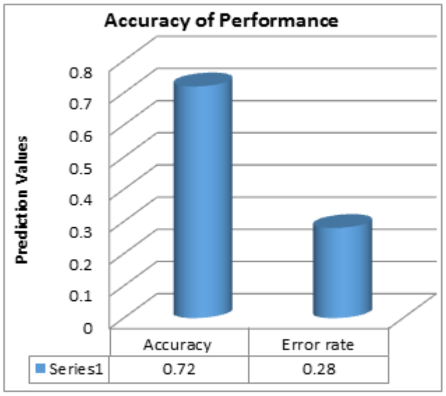
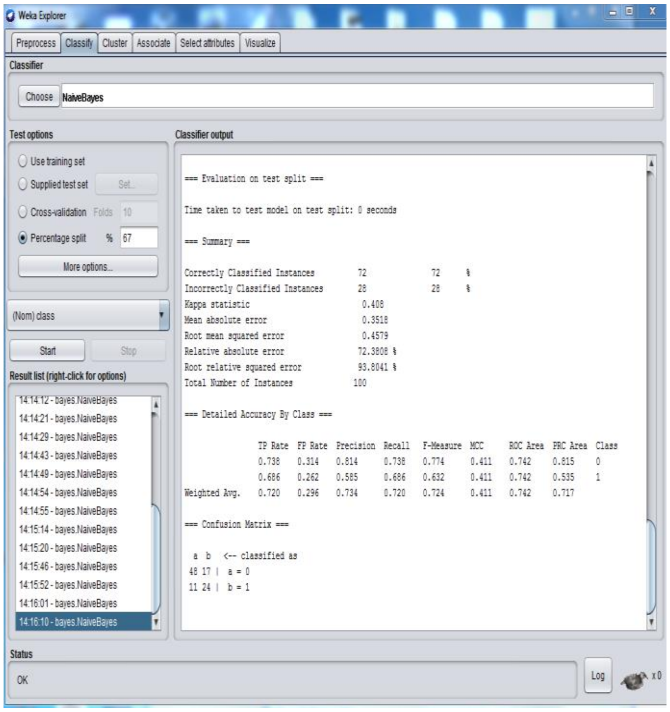


Figure 5: Accuracy rate and Error rate



Classifier: NaiveBayes

● Percentage split % 67

(Nom) class

```
14:14:12 - bayes.NaiveBayes
14:14:21 - bayes.NaiveBayes
14:14:29 - bayes.NaiveBayes
14:14:43 - bayes.NaiveBayes
14:14:49 - bayes.NaiveBayes
14:14:54 - bayes.NaiveBayes
14:14:55 - bayes.NaiveBayes
14:15:14 - bayes.NaiveBayes
14:15:20 - bayes.NaiveBayes
14:15:46 - bayes.NaiveBayes
14:15:52 - bayes.NaiveBayes
14:16:01 - bayes.NaiveBayes
14:16:10 - bayes.NaiveBayes
```

```
=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances          72               72      %
Incorrectly Classified Instances        28               28      %
Kappa statistic                          0.408
Mean absolute error                      0.3518
Root mean squared error                  0.4579
Relative absolute error                 72.3808 %
Root relative squared error             93.8041 %
Total Number of Instances              100

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.738    0.314    0.814      0.738    0.774      0.411    0.742     0.815     0
                 0.686    0.262    0.585      0.686    0.632      0.411    0.742     0.535     1
Weighted Avg.    0.720    0.296    0.734      0.720    0.724      0.411    0.742     0.717

=== Confusion Matrix ===

  a  b   <-- classified as
 48 17 |  a = 0
 11 24 |  b = 1
```

Status: OK

Figure 6: Accuracy rate and Confusion matrix

5. Conclusion

We have presented a new framework, MFES-NB, to address the issue of heart disease diagnosis. The model introduces a number of experiments to evaluate its performance. This system can help medical practitioner in efficient decision making based on the given parameter. We have train and test the system using a stratified 10-folds cross validation and obtained an accuracy score of 72%. This model demonstrates promising result and gives the patient to have early detection of heart disease presence.

References

[1] A. L. Bui, T. B. Horwich, and G. C. Fonarow, “Epidemiology and risk profile of heart failure,” *Nature Reviews Cardiology*, vol. 8, no. 1, pp. 30–41, 2011.

[2] J. Mourão-Miranda, A. L. W. Bokde, C. Born, H. Hampel, and M. Stetter, “Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data,” *NeuroImage*, vol. 28, no. 4, pp. 980–995, 2005.

[3] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, “Innovative artificial neural networks-based decision support system for heart diseases diagnosis,” *Journal of Intelligent Learning Systems and Applications*, vol. 5, no. 3, pp. 176–183, 2013.

[4] Q. K. Al-Shayea, “Artificial neural networks in medical diagnosis,” *International Journal of Computer Science Issues*, vol. 8, no. 2, pp. 150–154, 2011.

[5] R. Kavitha, & E. Kannan, “An efficient framework for heart disease classification using feature extraction and feature selection technique in data mining”. *International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS),* Pudukkottai, pp. 1-5, 2016.

[6] A. K. Paul, P. C. Shill, M. R. I. Rabin, and M. A. H. Akhand, “Genetic algorithm based fuzzy decision support system for the diagnosis of heart disease”. *In 5th International Conference on Informatics, Electronics and Vision (ICIEV),* pp. 145-150. IEEE, 2016.

[7] A. Dey, J. Singh, N. Singh, “Analysis of Supervised Machine Learning Algorithms for Heart Disease Prediction with Reduced Number of Attributes using Principal Component Analysis”. *Analytics.* 140(2), 27-31, 2016.

[8] S. A. Mostafa, A. Mustapha, M. A. Mohammed, M. S. Ahmad, & M. A. Mahmoud, “A fuzzy logic control in adjustable autonomy of a multi-agent system for an automated elderly movement monitoring application”. *International Journal of Medical Informatics,* 112, 173–184, 2018.

[9] A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor, & D. Keim, “Combining automated analysis and visualization techniques for effective exploration of high-dimensional data”. *In Visual Analytics Science and Technology,* VAST IEEE Symposium on (pp. 59–66), 2009.

[10] A. Ozcift, & A. Gulten, “Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms”. *Computer Methods and Programs in Biomedicine,* 104(3), 443–451, 2011.

[11] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms," *Mobile Information Systems*, vol. 2018, 2018.

[12] X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu, Q. Wang, *et al.*, "A hybrid classification system for heart disease diagnosis based on the RFRS method," *Computational and mathematical methods in medicine,* vol. 2017, 2017.

[13] S. U. Amin, K. Agarwal, and R. Beg, "Genetic neural network based data mining in prediction of heart disease using risk factors," in *2013 IEEE Conference on Information & Communication Technologies*, pp. 1227-1231, 2013.

[14] M. Anbarasi, E. Anupriya, and N. Iyengar, "Enhanced prediction of heart disease with feature subset selection using genetic algorithm," *International Journal of Engineering Science and Technology,* vol. 2, pp. 5370-5376, 2010.

[15] N. Bhatla and K. Jyoti, "An analysis of heart disease prediction using different data mining techniques,"

International Journal of Engineering, vol. 1, pp. 1-4, 2012.

[17] A. Shahi, M. N. Suleiman, M. N., N. Mustapha and T. Perumal, “Naïve Bayesian decision model for the interoperability of heterogeneous systems in an intelligent building environment”. *Automation in construction* 54, pp. 83-92, 2015.

Author -

Abba Babakura* obtained the B.Eng. degree in computer engineering from University of Maiduguri, Nigeria. He also received the MSc degree in intelligent systems from Universiti Putra Malaysia, Malaysia. He is currently doing his PhD degree in computer science from Usmanu Danfodiyo University Sokoto, Nigeria. His primary research interests include artificial intelligence, machine learning and intelligent building.

Mahmud Yusuf obtained the BSc degree in computer science from Bayero University Kano, Nigeria. He also received the MSc degree in computer science from Universiti Putra Malaysia, Malaysia. He is currently doing his PhD degree in computer science from School of Computing and Engineering, University of Huddersfield, United Kingdom. His primary research interests include artificial intelligence, machine learning and data mining.

Baba Yachilla obtained the B.Eng. degree in electrical and electronics engineering from University of Maiduguri, Nigeria. She also received the MSc degree in Electrical and Electronics Engineering from Nile University of Nigeria, Nigeria. She is currently doing her PhD degree in electrical and electronics engineering from Nile University of Nigeria, Nigeria. Her primary research interests include artificial intelligence, control system and machine learning.