Tri-Hybrid Naive Bayes Classification Model for Early Cardiovascular Disease Detection

*Zulkiflu Umar, Danlami Gabi and Nurudeen M. Ibrahim
Department of Computer Science, Kebbi State University of Science and Technology, Aliero, Nigeria
{zikumar2015|gabsonley|nurex115}@gmail.com

Received: 03-JAN-2023; Reviewed: 14-JUN-2023; Accepted: 26-JUN-2023
http://doi.org/10.46792/fuoyejet.v8i2.980

ORIGINAL RESEARCH

Abstract- Cardiovascular disease is a major global health problem affecting people around the globe. Although modern medicine has contributed in providing significant data to help mitigate cardiovascular disease, however, there is still needs to provide an ideal solution that will help in detecting early cardiovascular diseases. As a result, existing researchers have proposed several hybrid cardiovascular disease detection techniques using both machine learning and deep learning approaches. However, most solution requires improvement especially in term of accuracy detection. In this paper, a tri-hybrid Naïve Bayes classification algorithm for accuracy prediction of early cardiovascular disease detection is developed. In the developed algorithm, the strength of the conventional Naïve Bayes classification algorithm is improved with that of decision tree approach to achieve better accuracy. To further enhance its accuracy, Neural Network (NN) procedure is later incorporated into the improved Naïve Bayes classification algorithm. Implementation of the developed tri-hybrid algorithm is carried out on Waikato Environment for Knowledge Analysis (WEKA) and evaluated using dataset from the kaggle website that contain instances of 1025 and 14 attributes. Experimental results based on confusion matrix for performance evaluation indicates the developed tri-hybrid Naïve Bayes algorithm has achieved high classification accuracy and recall rates of 98.54 and 99% respectively when compared with that of the benchmarked schemes. The performance of the proposed solution deployed can help cardiologist make better prediction in the diagnosis of the heart disease. Although several machine learning algorithms as seen in the literatures are trained on historical data, these algorithms may not be able to accurately predict the risk of heart disease in patients who are exposed to new or emerging risk factors. Further research is to experiments a wider range of algorithms with a build in model that can detect early cardiovascular diseases as well improve accuracy.

Keywords- Heart disease prediction, Data mining, Naïve Bayes, Neural network, Machine learning.

1 INTRODUCTION

The heart being the central part of the body's cardiovascular system is known as the most essential organs that pump in blood into the human body system (Nashif, 2018). It composed of blood vessels, such as arteries, veins, and capillaries. These formed the complex network of blood circulation all over the body (Yahaya et al., 2020). Hence, abnormality in normal blood circulation as well as flow arising from the heart could potentially result in severe complications of heart diseases commonly referred to as cardiovascular diseases (CVDs) (Kumar et al., 2020). Besides, the heart is prone to several diseases caused by smoking, high cholesterol, high blood pressure, physical inactivity, unhealthy diet, obesity, and poorly controlled diabetes (Tougui et al., 2020) (Gabi et ai., 2019). Although cardiovascular diseases can be prevented through life style changes and other related measures, however, the alarming concern is how World health organization (WHO) have predicted an estimated 17.9 million people died from CVDs in 2019, representing 32% of all global deaths. Of these deaths, 85% were due to heart attack and stroke (Adebimpe, 2019).

The existing conventional approaches used in the diagnosis of heart disease were via the collection of medical history, use of a stethoscope, Ultrasound, Electrocardiogram (ECG) and echocardiography examination reports. Consequently, these methods may not actually be reliable as poor diagnosis can possibly arise that may led to wrong prescriptions which could result in very complicated cases among patients or even lead to death (Repaka, 2019). In order to decrease mortality rate from the heart diseases, there should be a fast and effective detection method, most especially in developing countries where shortage of specialists exist and where wrong diagnosed cases are predominantly high. Although, prediction model is not alternative to clinical treatments, however, it can serve as first hand tool to be aware of any type of disease and be prepared for it.

Several techniques of data mining (e.g., Support Vector Machine (SVM), Random forest (RF), K-Nearest Neighbor (KNN), Artificial Neural Network (ANN)) were used for prediction and diagnosis of heart disease patients (Adebinpe, 2019; Yusuf, 2021). These techniques have shown promising results in prediction of heart disease. On the other hand, the conventional Naïve Bayes is one of the successful data mining techniques that can be used in the diagnosis of heart disease (Pujianto et al., 2019; Gabi et al., 2019). As a simple data mining classifier, the Naive Bayes classifier when deployed can be more promising than the existing techniques, especially when combined with any other data mining techniques. Similarly, it has also exhibited high accuracy and speed when applied to large database (Rubini et al., 2021).

*Corresponding Author

Section B- ELECTRICAL/COMPUTER ENGINEERING & RELATED SCIENCES
Can be cited as:
Umar Z., Gabi D., and Ibrahim N.M. (2023). Tri-Hybrid Naive Bayes Classification Model for Early Cardiovascular Disease Detection, FUOYE Journal of Engineering and Technology (FUOYEJET), 8(2), 147-155. http://doi.org/10.46792/fuoyejet.v8i2.980

In this paper, we proposed a tri-hybrid Naive Bayes classification algorithm for the prediction of cardiovascular diseases among patients. In the proposed algorithm, the strength of Naïve Bayes classification model is combined with a decision tree algorithm to achieve better performance. To further enhance its accuracy of prediction, Neural Network (NN) and Decision Tree is incorporated into improved Naïve Bayes classification algorithm. The rest of this paper is organized as follow: Section 2 discusses the related works. Discussion on the methodology used in this paper is reported in Section 3. Sections 4 provide an explanation on the results achieved by the proposed algorithm, while Section 5 concludes the paper.

2 Related Literature

Significant amount of research were carried out on cardiovascular (Heart) disease prediction using machine learning techniques (Shah et al., 2020). In one of the notable research conducted in(Sridhar and Kapardhi, 2018), the researchers use the Naïve Bayes (NB) and Decision Tree (DT) to predict hearts disease detection. In their simulation results, the researchers were able to show that the DT algorithm outperformed NB in term of accuracy. However, proper feature selection approach needs to be adopted in order to achieve more significant performance. On the other hand,(Hayatu et al., 2020) dwelt on the use of Support Vector Machine (SVM), Linear Regression (LR), and NB algorithms in their analysis to predict heart disease. The researchers explored datasets from the UCI machine learning repository to train the algorithms. Experimental results based on the three classification algorithms according to the researchers shows the SVM outperformed significantly well in term of accuracy them benchmarked schemes.

In another development, (Subhadra and Vikas, 2019) proposed Multi-Layer Preceptor Neural network (MLP-NN) to predicts heart disease. Their technique uses backward propagation in training the algorithm. To determine the performance of their proposed algorithm in term of accuracy, the researchers uses Cleveland data of the UCI machine learning repository containing 303 instances and 76 features to trained and test the accuracy of their proposed algorithm. According to the researchers experimental results based on simulation shows the MLP-NN is able to achieved an acceptable accuracy of 93.39% heart disease prediction. However, computational complexity in term of time can be minimized if further improvement to their proposed algorithm is carried out. Similarly,(Reddy et al., 2019) dwelt on the advantages of DT and SVM for heart disease prediction. The researchers used sample dataset characterizing of age, chest pain, blood pressure and cholesterol level for their prediction. According to the researchers, data preprocessing is used in removing inconsistencies and missing values. Their experimental results according to the researchers showed that the DT classifier outperformed that of the SVM classifier in term of achieving better accuracy.

In their part,(Annepu and Gowtham, 2019) put forward a heart disease prediction framework based on Random Forest (RF) algorithm. The researchers use a sample dataset that contains 303 instances and were splits into training set and testing sets. Experimental results according to the researchers shows their proposed RF classifier was able to achieve better performance in term of accuracy compared to the benchmarked scheme. A real time heart disease prediction model known as Predict-Risk is put forward in (Annepu and Gowtham, 2019. The researchers developed an android application by incorporating clinical data obtained from patients whom were admitted with chest pain in a cardiac hospital. Based on 917 instances of dataset collected, 281 were at health camps while 636 were from a cardiac hospital. To analyze the features of their dataset, the researcher uses the Predict-Risk which was incorporated in the android application to predict accuracy. They use statistical based approaches known as Chi-square test, Fisher's Exact Test, Probability, Percentage and Ratios to compute the risk score. Computation results according to the researchers show that accuracy of prediction depends on the features of the heart disease shown (Raihan et al., 2017).

In another development, put forward an SVM based approach with Framingham health parameters for risk prediction of cardiovascular diseases. According to the researchers the SVM is compared with LR and Framingham Risk equation (FRE) on 104 cardiovascular cases. The results of the experiment via simulation shows the SVM is able to achieved better accuracy compared to the benchmarked classifiers. In their part, (Prasad et al., 2019) took a comparison on several classifiers to predicting heart disease. In this study, the researchers uses 493 valid samples of cerebral vascular disease prevention and treatment program, and adopted three classification algorithms, decision tree, Bayesian classifier and back propagation neural network. Based on their results, the researchers' show the decision tree had returned better accuracy and sensitivity than other benchmarked classifiers. In another study, (Usman et al., 2019) uses an LR based approach for heart disease prediction. According to the researchers, experimental results via simulation shows an LR algorithm outperformed better by returning 86.89% accuracy compared to other benchmarked algorithms.

In another study, (Jahangiri, 2023) introduced modified Bayesian approach, where the conditional distributions of all features alongside their parameter estimation are used to improve the Naïve Bayes algorithm's performance in terms of different performance criteria. Four other data mining algorithms are employed to evaluate the proposed approach's efficacy. These algorithms include decision trees, Naïve Bayes, k-nearest neighbors (KNN). The results of this study show that the modified Bayesian approach performs the best among all other competing algorithms. Reference (Kavitha et al., 2021) proposed a web-based application for the heart disease prediction for predicting heart diseases using machine learning

http://doi.org/10.46792/fuoyejet.v8i2.980 http://journal.engineering.fuoye.edu.ng/

techniques. The algorithms used for classification were Random Forest, Decision Tree and Hybrid model (Hybrid of random forest and decision tree). Datasets from Cleveland machine learning repository were used. Experimental result shows Decision Tree achieves 79% accuracy, and Random forest achieves 81% accuracy, Hybrid model achieves 88% accuracy in predicting heart disease**.**

3 METHODOLOGY

3.1 DATASET DESCRIPTION

This paper used the dataset download from an open access dataset website (Kaggle, 2023). It consists of 1025 samples with 14 attributes as shown in Table 1. The dataset are split into two: 70% consists of the training dataset while 30% consists of the testing with one output attribute named Target. Target attributes contains 2 classes, indicating either healthy (tested negative) or not healthy (tested positive). Ten (10) fold cross validation were used for the data classification in four different iterations.

3.2 NAÏVE BAYESIAN CLASSIFICATION ALGORITHM

Naive Bayes is a probabilistic classifier that deals with classification problems. The algorithm make an assumption as it relies on the fact that occurrence of a certain feature is independent of the occurrence of other features (Repaka, 2019; Mutyala, 2018). The algorithm consists of a training set D with a class label as C_a and C_p. It also consists of an $n - dimensional$ attribute, vector $X = (X_1, X_2 \dots, X_n, X_{n-1})$, with n depicting measurements made on the tuples from n attributes. Assume there are m numbers of classes $C_1, C_2 \dots C_m$ for prediction. Given the value of X, a prediction is carried out by the classifier on the basis that X belongs to the class having the highest posterior probability as it conditioned on X according to Equation 1 (Usman et al., 2019).

$$P(C_i|X) > P(C_j|X) for 1 \le j \le m \; and \; j \ne i \tag{1}$$

Thus, we maximize $P(Ci|X)$ the class C_i for which $P(C_i|X)$ is maximized according to Equation 2.

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \tag{2}$$

$P(X)$, represent all classes and as such, only $P(X|Ci) * P(Ci)$ can be maximized. It is often assumed that the classes are equally likely as $P(C_1) = P(C_2) = \cdots P(C_{m-1}) = P(C_m)$, if the class prior probabilities are not known. Otherwise, $P(X|C_i)\,P(C_i)$ is maximized. On the other hand, with data sets having several attributes, the computation might be extremely expensive to compute $P(X|C_i)$. This might be reduced especially when naïve assumption of class conditional independence is made and as such, can be used in the evaluation of $P(X|C_i)$ according to Equation 3.

$$P(X/C_i) = \prod_{k=1}^{m} P(X_k/C_i) \tag{3}$$

$$= P(X_1/C_i) * P(X_2/C_i) * \dots * P(X_n/C_i) \tag{4}$$

In order to predict the class label of $X, P(X|Ci)P(Ci)$is evaluated for each class Ci. The classifier predicts that the class label of tuple X is the class Ci if and only if,

$$P(X|C_i)P(C_i) > P(X|Cj)P(Cj) \;\; for \; 1 \le j \, m, j \ne i \tag{5}$$

Table 1. Detailed description of the dataset

ATTRIBUTE	DESCRIPTION
Age	Displays the age of the individual.in years
Sex	Gender
Cp	Chest-pain type: displays the type of chest-pain experienced by the individual using the following format: 0 = typical angina 1 = atypical angina, 2 = non angina pain, 3 = asymptotic.
Trestbps	Resting Blood Pressure: displays the resting blood pressure value of an individual in mmHg (unit). Anything above 130-140 is typically cause for concern.
Chol	Serum cholesterol: displays the serum cholesterol in mg/dl (unit)
Fbs	Fasting Blood Sugar: compares the fasting blood sugar value of an individual with 120mg/dl. If fasting blood sugar > 120mg/dl then : 1 (true) else : 0 (false) '>126' mg/dl signals diabetes
Restecg	Resting ECG : displays resting electrocardiographic results 0 = normal, 1 = having STT wave abnormality, 2 = left ventricular hypertrophy.
Thalach	Max heart rate achieved: displays the max heart rate achieved by an individual.
Oldpeak	ST depression induced by exercise relative to rest: displays the value which is an integer or float.
Slope	Slope of the peak exercise ST segment: 0 = up sloping: better heart rate with exercise (uncommon) 1 = flat: minimal change (typical healthy heart), 2 = down sloping: signs of unhealthy heart.
Ca	Number of major vessels (0–3) colored by fluoroscopy: displays the value as integer or float.
Thal	Displays the thalassemia: 1, 3 = normal. 6 = fixed defect. 7 = reversible defect: no proper blood movement when exercising
Exang	Exercise induced angina : 1 = yes, 0 = no
Target	Displays whether the individual is suffering from heart disease or not : 1 = yes, 0 = no.

http://doi.org/10.46792/fuoyejet.v8i2.980 http://journal.engineering.fuoye.edu.ng/

Algorithm 1. Naive Bayes Classification Algorithm

Begin:
Input:
Initialize maximum number of iteration, Target set, Training Sets D, Testing sets, Class labels C_i& C_j.
Output: Best prediction values for accuracy, precision, recall & F-score

Randomly generate datasets $X = (X_1, X_2 \dots, X_n, X_{n-1})$ and splits into training and target sets)
Do {
Repeat
Initialize repetition counter $m \leftarrow 0$
//Execute the Naïve Bayes procedure
Compute the highest posterior hypothesis and store the results according to Equ.(1)
Maximize the maximum posteriori hypothesis according to Equ. (2) //reduce computational complexity
If no dependency among variables then,
{ Minimize complexity according to Equ.(4)
Predicts class label X according to Equ. (5)
If Predict_value= test_value
{ Match $m + 1 = 1$
Else
UnMatch $m + 1 = 1$
If
Class posterior probability is known,
{ Return the class maximum posterior probability value
Else
Maximize $P\ (X|Ci)\ P\ (Ci)$ according to Equation
Until stopping criteria is not exceeded;
}//*EndIf*
}//*EndIf*
}//*EndIf*
}//***EndDo***
While (predict_value ≠Test_value)
Output the best accuracy values.

End

3.3 Decision Tree (DT)

One of the most powerful tools for classification and prediction is the decision tree. It uses a classification model that allocates instances on an element and concludes about the objective value (Mutyala, 2018). They are trees that categorize attributes by organizing them according to their values in tree like fashion. Every node in the tree denotes an attribute that will be arranged, and each branch defines a value that the node can accept. Algorithm 2 lists a decision tree algorithm.

3.4 Neural Networks (NN)

Neural network implores networking concepts that applies techniques of how neurons communicate with each other along network path. It certifies how set of data relates to human brain, where, the network response to each input and then compares it with the target value. When there is a mismatch in the computed response with the target value, the network weights are adapted in accordance to the learning rule (Padmanabhan et al., 2019). Algorithm 3, show syntax of the Neural Network (NN) algorithm.

3.5 Proposed Tri-Hybrid Naive Bayes Classification Algorithm

In the proposed try-hybrid algorithm, the Naive Bayes is a probabilistic classifier that deals with classification problem. The conventional Naive Bayes can exhibit high computational complexity and may not converge on time which in turn can significantly affects accuracy of prediction. During the initialization phase, the Naive Bayes algorithm is used for the classification purposes which in turn splits data into training and testing phase. The highest posterior hypothesis is then computed where the results are stored in the archive. On the other hand, the maximum posterity is determined when the Decision Tree Classification Algorithm is incorporated through the uses of loop decision as shown in Algorithm 2. This is because, when the prediction value is equivalent to the testing value, the maximum posterity values are return. To determine the accuracy of the maximum posterity values, the Neural Network is incorporated into the hybrid Naive Bayes and decision tree algorithm which in turn help to predict better accuracy. The proposed algorithm as indicated in Algorithm 4.

http://doi.org/10.46792/fuoyejet.v8i2.980 http://journal.engineering.fuoye.edu.ng/

Algorithm 2. Decision Tree Classification Algorithm

Begin
Input:
X training set, T testing set, size, random state [0,1]
Output: predicted value, training test value

1. **For** i in range[0,1], length (T training)
2. {
3. Model= Decision Tree Classifier
4. Model= X training set & T testing set
5. **If** T training set=0
6. zeroes += 1
7. else:
8. ones += 1
9. match = 0
10. UnMatch = 0,
11. **For** i in range [0,1]:
12. {
13. **If** prediction value = testing value
14. match += 1
15. else:
16. UnMatch += 1
17. Output on accuracy
18. *}//EndIf*
19. *} Endfor*
20. *}//EndIf*
21. *}Endfor*

End

Algorithm 3. Neural Network Algorithm

Begin:
Input:
learning rate Y, bias, weights i, $\forall i = 1 \; to \; n$, perceptron, p, Training set $X = (X_1, X_2 \dots, X_n, X_{n-1})$,
Output:: $p = x * weights[0] + y * weights[1] + bias * weights[2]$

1. **If** $p > 0$
2. {
3. p = 1
4. else
5. $p = 0$
6. output the accuracy(x, "or", y, "is : ", p)
7. Repeat step 1…4
8. *}//EndIf*

End

4 Experimental Environment

Research was performed on window 10 of 64 bits operating systems with Intel Corei7 as well as an Installed Memory of 4.00GB. It consists of 697GB hard disk capacity. WEKA simulation environment is then installed on the operating system which provides the platform for the implementation of our proposed algorithm.

4.1 Weka

Stands for Waikato Environment for Knowledge Analysis. WEKA is a data mining tool developed at the University of Waikato, Newzeland. It uses GNU general public licenses and is freely available on following link: http://www.cs.waikato.ac.nz/~ml/weka. It is implemented in Java programming language and has Graphic user Interface (GUI) for loading data, running analysis and producing visualization of result. WEKA supports many data mining techniques and algorithms like classification, clustering, feature selection, data preprocessing, regression, visualization and clustering. The Graphic user Interface (GUI) of WEKA is shown in figure 1. The WEKA GUI chooser provides four interfaces.

WEKA Explorer: It is used for exploring the data with WEKA by providing access to all the facilities by the use of menus and forms.
WEKA Experimenter: It allows a user to create, analyze, modify and run large scale experiments.

WEKA Knowledge Flow: It has the same function as that of explorer. It supports incremental learning. It handles data on incremental basis. It uses incremental algorithms to process data.

WEKA Simple Command Line: Simple command line (CLI) stands for command line interface. It just provides all the functionality through command line interface.

4.2 Performance Evaluation Metrics

Table 1 indicates the metrics used in the evaluation of the proposed tri-hybrid classification algorithm. A correctly classified instance is one in which the classifier predicts the correct class of the test's instance.

Algorithm 4. Tri-hybrid Naive Bayes Classification Algorithm

Begin:

Input:

Initialize maximum number of iteration, Target set, Training Sets D, Testing sets, Class labels C_i& C_j, X training set, T testing set, size, random state [0,1], bias, weights i, $\forall i = 1 \; to \; n$, perceptron, p.

Output: Best prediction values for accuracy, precision, recall & F-score

1. Randomly generate datasets $X = (X_1, X_2 \dots, X_n, X_{n-1})$ and splits into training and target sets)
2. **Do**
3. {
4. **Repeat**
5. Initialize repetition counter $m \leftarrow 0$
6. *// Executes Naïve Bayes procedures*
7. Execute the Naïve Bayes procedure using **Algorithm 1**.
8. reduce computational complexity
9. {
10. **If** no dependency among variables then
11. reduce computational complexity by executing **Algorithm 2**,
12. {
13. **If**
14. Predict_value= test_value
15. Match $m + 1 = 1$
16. Elae
17. UnMatch $m + 1 = 1$
18. //Execute the Neural network procedure
19. **If**
20. {
21. Class posterior probability is known,
22. Execute **Algorithm 3**
23. Return the class maximum posterior probability value according to **Algorithm 1**
24. **Else**
25. Maximize $P\,(X|Ci)\;P\,(Ci)$ according to Equation
26. **Until** stopping criteria is not exceeded;
27. }//*EndIf*
28. }//*EndIf*
29. }//*EndIf*
30. }//*EndIf*
31. }//***EndDo***
32. **While** (predict_value ≠Test_value)
33. Output the best accuracy values.

End

http://doi.org/10.46792/fuoyejet.v8i2.980 http://journal.engineering.fuoye.edu.ng/

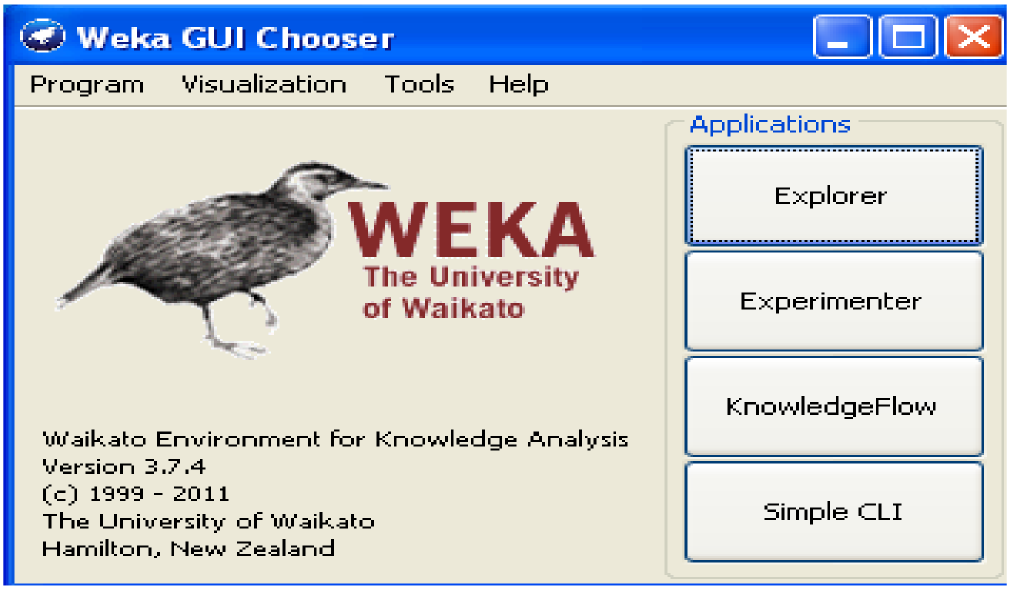

Fig. 1: GUI Interface of WEKA

Table 2. Performance evaluation metrics

Metric	Formula
Accuracy	$= \frac{\text{True positves } + \text{true negs}}{\text{total number of examples}}$
Precision	$= \frac{\text{True positves}}{\text{True positves } + \text{true negatives}}$
Recall	$= \frac{\text{True positves}}{\text{True positves } + \text{ false negatives}}$
F1 Score	$= \frac{2 * TP}{2 * TP + FN + FP}$

Table 3. Results of the three classifications algorithms

Model	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)	Execution time
Naïve Bayes (NB)	88.78	88.60	89.40	80.00	0.01s
Decision Tree(DT)	96.09	97.10	95.20	96.10	0.03s
Neural Network (NN)	96.58	98.00	95.20	96.60	0.09s
NB+ DT	98.04	99.00	97.00	98.10	0.06s
NB+DT+ NN	98.54	98.10	99.00	98.60	0.15s

5 Results Discussion

At the first instances on the results achieved in Table 3, the Neural Network (NN) has highest accuracy with 96.58%, and Precision of 98.00%, while Decision Tree (DT) and Neural Network (NN) have highest recall rate of 95.20% and NN has higher F-measure of 96.60%. At this instance, the Naïve Bayes classification algorithm could not return better accuracy compared to the conventional Decision Tree and Neural Network due to the fact that it has highest computational complexity. Figure 2, is a graphic indication on the performance of the three algorithms in terms of accuracy, precision, recall, and F-measures.

To deal with the complexity of the conventional Naïve Bayes algorithm, a hybrid approach is proposed based on DT algorithm. The incorporation of DT algorithm helps to reduce it computational complexity in detecting the cardiovascular diseases. The results in Table 3 also indicates the improved Naïve Bayes (NB+DT) has the highest accuracy, precision, recall, F-measure of 98.04%, 99.00, 97.00 and 98.10 respectively compared to the NB, DT and NN.

The results as indicated in Table 3 tri-hybrid algorithm. Here, time complexity of the prediction is reduced in term of accuracy when compared to hybrid approach. The results show it has achieved highest accuracy, precision, recall and F-measure of 98.54%, 98.10%, 99.0% and 98.06% respectively when compared to the hybrid approach. To further show performance of the tri-hybrid algorithms Figure 3 is used to illustrates trend of the algorithm.

http://doi.org/10.46792/fuoyejet.v8i2.980 http://journal.engineering.fuoye.edu.ng/

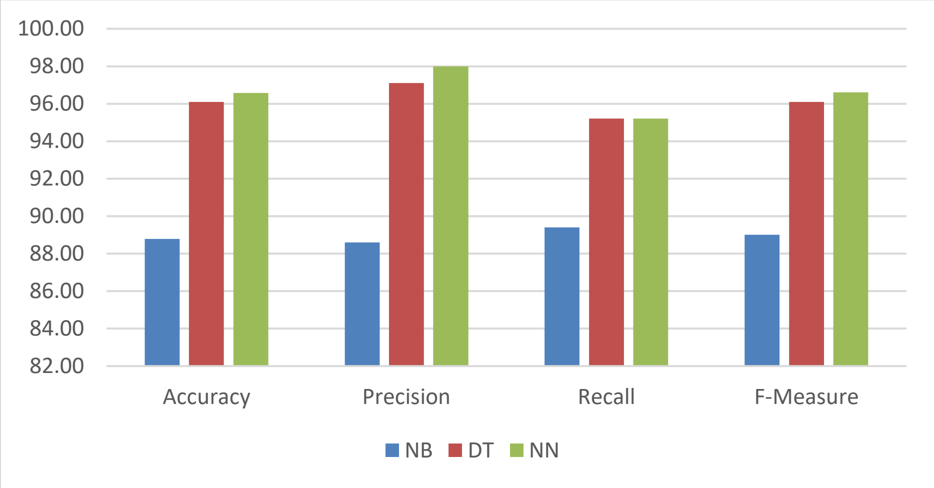

Fig 2: Comparative analysis of the three classifications algorithms.

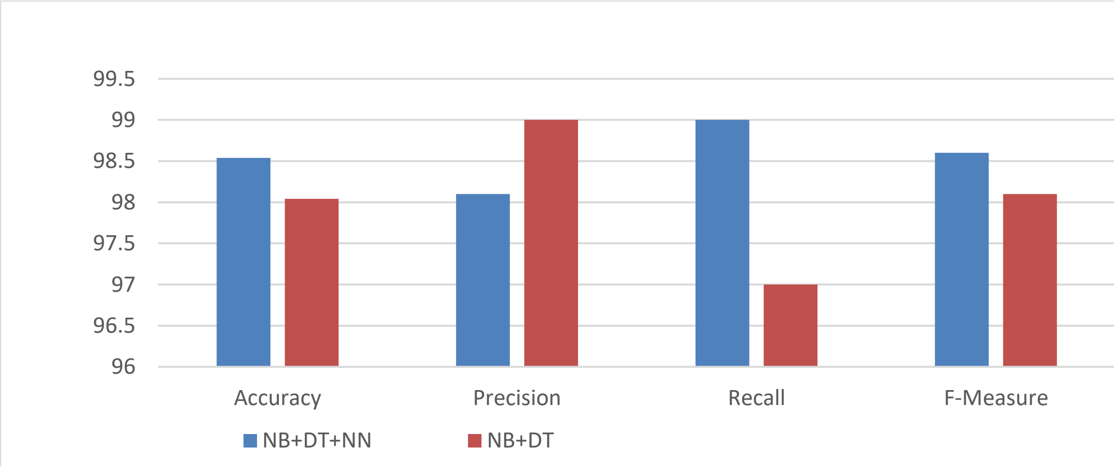

Fig. 3: Comparative analysis of the results of the try-hybrid and bi-hybrid classification algorithms

Although our experiment has achieved acceptable results through the developed try-hybrid algorithm, however, another important challenge is how to compare current study with other existing approaches. The majority of the existing approaches dwelt on uses of only the conventional algorithms which have less impact on accuracy prediction of the cardiovascular diseases while few are seen that dwelt on the use of hybrid approach. With all this among similar studies, the proposed try-hybrid Naive Bayes classification algorithm is more promising in the prediction of cardiovascular diseases compared to the existing algorithms A tri-hybrid approach combines the strengths of three different approaches, while a hybrid approach only combines the strengths of two. This means that a tri-hybrid approach is more robust than a hybrid approach. The tri-hybrid approach reduces time complexity in detection of cardiovascular disease as such; it has achieved 98.54% compared to other existing algorithms as indicated in Table 3.

6 Conclusion

In this paper, we proposed a try-hybrid Naive Bayes classification algorithm for cardiovascular disease detection. The algorithm predicts whether a patient have heart disease or not. Results show the try-hybrid algorithm achieves better accurate results than the conventional Naive Bayes in heart disease prediction. It has achieved the highest accuracy of 98.54% and 99% recall which outperformed that of Naïve Bayes algorithm. The performance of our proposed tri-hybrid algorithm indicates that well deployed in medical field, it can assist cardiologists to help make more consistent diagnosis of heart disease. Although, experiment conducted in this study was implemented with default parameters of the algorithms, further investigations can be perform with different parameter settings to enhance and expand the capabilities of the prediction models.

References

Adebimpe, M. M. (2019). Prediction Of Heart Disease Using Bayesian Network Model. MSc Thesis, June 2019. https://repository.aust.edu.ng/xmlui/bitstream/handle/123456789/4903/Muibideen%20Mistura.pdf?sequence=1&isAllowed=y

Annepu, D., and Gowtham, G. (2019). Cardiovascular disease prediction using machine learning techniques. International

Research Journal of Engineering and Technology, 6 (4): 3963–3971.

D. Gabi, D., Ismail, A. S., Anazida. Z. and Zalmiyah. Z. (2019). Quality of service task scheduling algorithm for time-cost trade off scheduling problem in cloud computing environment. International Journal of Intelligent Systems Technologies and Applications, 8(1): 469.

Hayatu, H. I., Mohammed, A., Barroon, A., Ali, Y. S., and Mohammed, U. S. (2020). Feature Relevance Analysis and Classification of Kaduna State Road Traffic Accident Data using Machine Learning Techniques. International Journal of Information Processing and Communication, 10(1):86–98.

Jahangiri, S. (2023). An Improved Naïve Bayes Approach to Diagnose Cardiovascular Disease : A Case Study. 0–36.

Kaggle.com (2023), Heart disease classification dataset. Available from https://www.kaggle.com/datasets/adhyanmaji31/heart-disease-classification, accessed January 2023

Kumar, N. K., Sindhu, G. S., Prashanthi, D. K., and Sulthana, A. S. (2020). Analysis and Prediction of Cardio Vascular Disease using Machine Learning Classifiers. Ml, 15–21.

Mutyala, N. K. (2018). Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 3(3):2456-3307.

Nashif, S. R. (2018). Heart disease detection by using machine learning algorithms. World Journal of Engineering and Technology, 1(6): 854–873.

Padmanabhan, M., Yuan, P., Chada, G., and Nguyen, H. V. (2019). Physician-friendly machine learning: a case study with cardiovascular disease risk prediction. Journal of Clinical Medicine. 8(7): 1050.

Prasad, R., Anjali, P., Adil, S. and Deepa, N. (2019). Heart disease prediction using logistic regression algorithm using machine learning. International Journal of Engineering and Advanced Technology, 8 (35): 659–662.

Pujianto, U., Luki, A., Ar, H., and Mohammad, A. M. (2019). Comparison of Naïve Bayes Algorithm and Decision Tree C4 . 5 for Hospital Readmission Diabetes Patients using HbA1c Measurement. 2(2): 58–71.

Raihan, M., Mondal, S., More, A., Boni, P. K., and Sagor, M. F. (2017). Smartphone based heart attack risk prediction system with statistical analysis and data mining approaches. Advances in Science, Technology and Engineering Systems Journal, 2 (3): 1815–1822.

Reddy, P. K., Reddy, T. S., Balakrishnan, S., Basha, S. M., & Poluru, R. K. (2019). Heart disease prediction using machine learning algorithm. International Journal of Innovative Technology and Exploring, 8 (10):2603–2606.

Repaka, A. N. (2019). Design And Implementing Heart Disease Prediction Using Naives Bayesian. 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), pp. 292–297.

Rubini, P. E., Subasini, C. A., Vanitha Katharine, A., Kumaresan, V., Gowdhamkumar, S., and Nithya, T. M. (2021). A cardiovascular disease prediction using machine learning algorithms. Annals of the Romanian Society for Cell Biology, 25(2): 904–912.

Shah, Devansh, S., Patel1, S., and Santosh, K. B. (2020). Heart Disease Prediction using Machine Learning Techniques, SN Computer Science (2020) 1:345.

Sridhar, A. and Kapardhi, A. (2018). Predicting heart disease using machine learning algorithm. International Research Journal of Engineering and Technology, 6 (4): 36-38.

Subhadra, K. and Vikas, B. (2019). Neural network based intelligent system for predicting heart disease. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(5): 484-487.

Tougui, I., Jilbab, A. and El Mhamdi, J. (2020). Heart disease classification using data mining tools and machine learning techniques. Health and Technology, 10(5):1137–1144.

Usman, I., Aliyu, S., Isma, B. and Yusuf, A. (2019). A Comparative Study of Base Classifiers in Predicting Students' Performance Based on Interaction with LMS Platform. FUDMA Journal of Sciences (FJS), 3(1):231–239.

Yahaya, L., David O. N. and Joshua G. E. (2020). A Comprehensive Review on Heart Disease Prediction Using Data Mining and Machine Learning Techniques. American Journal of Artificial Intelligence, 4(1): 20-29.

Yusuf, A. B. ., Dima, R. M. and Aina, S. K. (2021). Optimized Breast Cancer Classification using Feature Selection and Outliers Detection. Journal of the Nigerian Society of Physical Sciences, 3(4): 298–307.

http://doi.org/10.46792/fuoyejet.v8i2.980 http://journal.engineering.fuoye.edu.ng/