International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Heart Disease Identification using Machine Learning

Rohini M[1], Keerthika A[2], Pavithra N[2], Sandhiya S[2], Vedha Smirtha S[2]
[1]Assistant Professor, Department of Computer Engineering, Coimbatore Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India
[2]UG Scholar, Department of Computer Engineering, Coimbatore Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India

Article Info
Volume 8 Issue 2
Page Number: 166-170

Publication Issue :
March-April-2021

Article History
Accepted : 15 March 2021
Published : 20 March 2021

ABSTRACT

Heart disease is one of the complex diseases and globally many people suffered from this disease. On time and efficient identification of heart disease plays a key role in healthcare, particularly in the field of cardiology. we proposed an efficient and accurate system to diagnosis heart disease and the system is based on machine learning techniques. The system is developed based on classification algorithms includes Naïve Bayes, decision tree, Random forest Algorithm. After classification, performance criteria including accuracy, precision, F-Score, recall, support is to be calculated. The comparison measure expose that Random Forest is the best classifier for the diagnosis of heart disease. Our experimental results show that accuracy improved over traditional classification techniques. This system is feasible and faster and more accurate for diagnosis of heart disease.

Keywords : Heart diagnosis, Machine Learning, Classifications, Random Forest.

I. INTRODUCTION

Human heart can be described as a compound body organ contains muscles together with biological nerves. Human heart pumps nearly 5l of blood in the body providing the human body with renewed materials. Medical research literature shows that there is much interest from the scientific researchers in implementing the Machine Learning, including classification algorithms in medical devices. The human heart operation is composite and any failure is risky to human lives. Hence, heart diagnose systems have been a main concern to the scientific researchers in the last decades. It is not possible to use linear systems to perform the heart diagnosis. Advanced heart diagnosis equipment is not always available in every medical center, especially in the rural areas where less support and care. Moreover, it is not possible for many people to travel to regional medical centers where high quality hospital services are affordable. Physician intuition and experience are not always enough to attain high quality medical results. Consequently, medical errors and undesirable results are reasons for a need for a state-of-the-art computer-based diagnosis system, which in turns reduce medical fatal errors, increase patient safety and save lives. so, to develop a non-invasive diagnosis system based on classifiers of machine learning (ML) to resolve these issues. In this research work, causes of heart diseases, the complications and the remedies for the diseases have been considered. An intelligent

system which can diagnose heart diseases has been implemented. This system will prevent misdiagnosis which is the major error that may occur by medical doctors. The dataset of heart disease has been used to carry out this experiment. The dataset comprises attributes of patients diagnosed for heart diseases. The diagnosis was used to confirm whether heart disease is present or absent in the patient. The datasets were obtained from the UCI Machine Learning. This dataset was divided into training, validation set and testing set, to be fed into the network.

II. LITERATURE SURVEY

In this survey various machine learning based diagnosis techniques have been proposed to diagnosis HD. This study presents some existing machine learning based diagnosis techniques. HD classification by using machine learning classification techniques and the performance of the system was 77% in term of accuracy developed by Detrano et al [5]. A diagnosis system using multi-layer perception and support vector machine (SVM) algorithm of HD classification and achieved accuracy 80.41% developed by Guadadhe et al [6]. HD classification system by utilizing a neural network with the integration of fuzzy logic. The classification system is achieved 87.4% accuracy developed by H. Kahramanli and N. Allahverdi et al [7]. HD prediction method by using hybrid machine learning techniques. It is also a new method for significant feature selection from the data for effective training and testing of machine learning classifier. the classification accuracy is 88.07% developed by S Mohan et al [8]. Diagnosis system for identification of heart disease based on (FCMIM) Fast Conditional mutual information -SVM classification technique. It gives better accuracy based on classification techniques [9]. HD prediction is calculating the precisions, recalls, support, f-score is compared with the existing system technique and decision tree classification technique is gives the better result [10]. Relief-based feature selection methods (RBAs) are reviewed in detailed context. RBAs can detect interactions without examining pairwise combinations. Iterative RBAs have been developed to scale them up to very large feature spaces. Research focused on core algorithms, iterative scaling, and data type flexibility. This increases the difficulty and computational burden placed on modelling methods. Feature selection could generically be defined as the process of identifying relevant features and discarding irrelevant ones [11].

III. III.PROPOSED SYSTEM

We will propose in this paper we used classification is done using Naïve Bayes, Decision Tree, Random Forest Algorithm will be applied on the data sets. Then analysis the classification report and finally compare the accuracy. This system is act as an efficient and accurate to diagnosis heart disease it is based on machine learning techniques. This system is feasible, faster and more accurate.

1. SYSTEM ARCHITECTURE

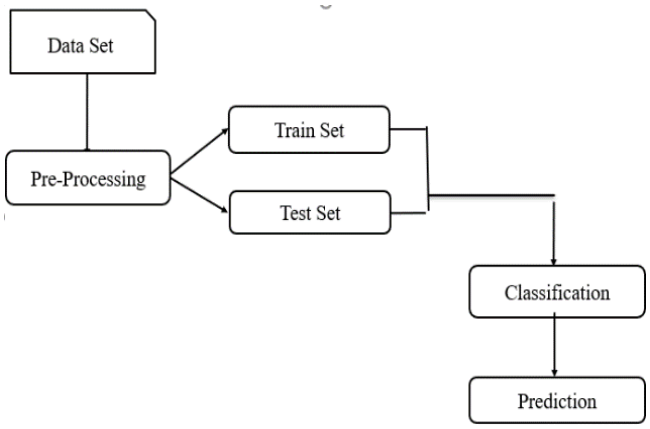

fig. 1. Proposed System Architecture

2. METHODS

A. DATASET

The data selection is the process of selecting the data for detecting the disease. In this project, the heart disease dataset is used for detecting the disease. The dataset contains 303 instances and 14 attributes.

B. DATA PREPROCESSING

Data pre-processing is the process of removing the unwanted data from the dataset. Missing data removal - This process is the null values such as missing values and Nan values are replaced by 0. Encoding Categorical data - That categorical data is defined as variables with a finite set of label values. That most machine learning algorithms require numerical input and output variables.

C. SPLITTING DATASET INTO TRAIN AND TEST DATA

Data splitting is the act of partitioning available data into two portions, usually for cross-validator purposes. One Portion of the data is used to develop a predictive model and the other to evaluate the model's performance. When you separate a data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing.

D. CLASSIFICATION ALGORITHMS

The Supervised classification algorithm such as Naïve Bayes, Decision Tree, Random Forest.

a. Naïve Bayes

The naïve bayes algorithm which is used for classification concerned problem. The Training dataset used by naïve bayes to compute the value of the conditional probability of vector for a given class. It supports continuous valued features and models each as conforming to a Gaussian(normal) distribution. An approach to create a simple model is to assume that the data is described to Gaussian distribution with no co-variance (independent dimensions) between dimensions.

b. Random Forest

This algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.

c. Decision Tree

Simple to understand and to interpret. Trees can be visualised. Requires little data preparation. Other techniques often require data normalisation, dummy variables need to be created and blank values to be removed. Note however that this module does not support missing values.

D. PREDICTION

Predictive analytics algorithms try to achieve the lowest error possible by either using "boosting" or "bagging". **Accuracy** -Accuracy of classifier refers to the ability of classifier. It predicts the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.

E. RESULT GENERATION

The Final Result will get generated based on the overall classification and prediction. The performance of this proposed approach is evaluated using some measures like, Precision, Recall, F-score, Support and Accuracy.

a. Precision

It is defined as the number of true positives divided by the number of true positives plus the number of false positives.

$$\text{Precision}=\frac{TP}{TP+FP}$$

b. Recall

It is the number of correct results divided by the number of results that should have been returned. In binary classification recall called sensitivity. It can be viewed as the probability that a relevant document is retrieved by the query.

$$\text{Recall}=\frac{TP}{TP+FN}$$

c. F- score

It is also called as F-measure and F1-Score. It is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test.

$$\text{F- score}=\frac{2TP}{2TP+FP+FN}$$

d. Accuracy

The accuracy of classifier refers to the ability of classifier. It predicts the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN}$$

IV. RESULT AND DISCUSSIONS

We compared the proposed heart disease identification accuracy on number of samples and show the result graphically. Let see the following graph and table shows the

heart disease prediction accuracy result based on naïve bayes, decision tree and Random forest classification technique. The result shows that classification reports include precision, recall, F1-score, support.

```
Naive Bayes
-------Confusion Matrix-------
[[21  4]
 [ 6 30]]
------Classification Report------
              precision    recall  f1-score   support

           0       0.78      0.84      0.81        25
           1       0.88      0.83      0.86        36

   micro avg       0.84      0.84      0.84        61
   macro avg       0.83      0.84      0.83        61
weighted avg       0.84      0.84      0.84        61

------Accuracy------
The accuracy score achieved using Naive Bayes is: 83.61 %
```

fig2. Naïve bayes report

```
Decision Tree
-------Confusion Matrix-------
[[22  6]
 [ 5 28]]
()
------Classification Report------
              precision    recall  f1-score   support

           0       0.81      0.79      0.80        28
           1       0.82      0.85      0.84        33

   micro avg       0.82      0.82      0.82        61
   macro avg       0.82      0.82      0.82        61
weighted avg       0.82      0.82      0.82        61

------Accuracy------
The accuracy score achieved using Decision Tree is: 81.97 %
```

fig3. Decision Tree report

```
Random Forest
-------Confusion Matrix-------
[[24  3]
 [ 3 31]]
()
------Classification Report------
              precision    recall  f1-score   support

           0       0.89      0.89      0.89        27
           1       0.91      0.91      0.91        34

   micro avg       0.90      0.90      0.90        61
   macro avg       0.90      0.90      0.90        61
weighted avg       0.90      0.90      0.90        61

------Accuracy------
The accuracy score achieved using Random Forest Tree is: 90.16 %
```

fig4. Random forest report

Compare the accuracy between the Classification techniques. It is represented in the bar chart in fig5. It shows the result is Random Forest classification techniques gives the efficient and high accuracy for heart disease identification system.

Table1

Algorithm	Accuracy
Naïve bayes	83.61%
Decision Tree	81.97%
Random Forest	90.16%

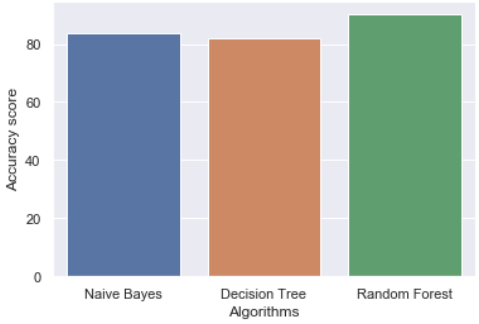

fig5. Accuracy comparison

V. V. CONCLUSION

In this paper, we have proposed an efficient machine learning based diagnosis system has been developed for the diagnosis of heart disease. Machine learning classifiers include Naive Bayes, Decision Tree and Random Forest are used in the designing of the system. The system is tested on heart disease dataset. Thus, the experimental results show that accuracy of Random Forest classification technique is gives very good as compared with other classification techniques show in Table1. Small amount improvement in prediction accuracy has great influence in diagnosis of diseases. The diagnosis of heart disease as we think that developing a decision support system through machine learning algorithms it will be more suitable. This research will contribute reliable and faster automatic heart disease diagnosis system, where easy diagnosis of heart disease will save lives. In future, it is possible to provide extensions or modifications to the proposed clustering and classification algorithms using intelligent agents to achieve further increased performance. Apart from the experimented combination of data mining techniques, further

combinations such as artificial intelligence, soft computing and other clustering algorithms can be used to improve the accuracy.

VI. REFERENCES

[1]. A. L. Bui, T. B. Horwich, and G. C. Fonarow, "Epidemiology and risk profile of heart failure," Nature Rev. Cardiol., vol. 8, no. 1, p. 30, 2011.

[2]. M. Durairaj and N. Ramasamy, "A comparison of the perceptive approachesfor preprocessing the data set for predicting Theory Appl., vol. 9, no. 27, pp. 255–260, 2016.

[3]. L. A. Allen, L. W. Stevenson, K. L. Grady, N. E. Goldstein, D. D. Matlock, R. M. Arnold, N. R. Cook, G. M. Felker, G. S. Francis, P. J. Hauptman, E. P. Havranek, H. M. Krumholz, D. Mancini, B. Riegel, and J. A. Spertus, "Decision making in advanced heart failure: A scientific statement from the American heart association," Circulation, vol. 125, no. 15, pp. 1928–1952, 2012.

[4]. S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, "Innovative artificial neural networks-based decision support system for heart diseases diagnosis," J. Intell. Learn. Syst. Appl., vol. 5, no. 3, 2013, Art. no. 35396.

[5]. R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, "International application of a new probability algorithm for the diagnosis of coronary artery disease," Amer. J. Cardiol., vol. 64, no. 5, pp. 304–310, Aug. 1989.

[6]. M. Gudadhe, K. Wankhade, and S. Dongre, "Decision support system for heart disease based on support vector machine and artificial neural network," in Proc. Int. Conf. Comput. Commun. Technol. (ICCCT), Sep. 2010, pp. 741–745.

[7]. H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases," Expert Syst. Appl., vol. 35, nos. 1–2, pp. 82–89, Jul. 2008.

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

[9]. Jian Ping Li, Amin Ul Haq, Salah Ud Din, Jalaluddin Khan, Asif Khan, And Abdus Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," IEEE Access, vol. 8, 2020.

[10]. Riddhi Kasabe, " Heart Disease prediction

[11]. J. Biomed. Informat, "Relief-based August-2020. feature selection: Introduction and Journal Of Engineering Research & Technology(IJERT), vol. 9 Issue 08, using machine learning", in International review," in 2008.

[12]. Senthilkumar Mohan, Chandrasegar Thirumalai, And Gautam Srivastava" Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" IEEE ACCESS 2019.

[13]. Zeinab Arabasadi , Roohallah Alizadehsani , MohamadRoshanzamir, Hossein Moosaei , Ali Asghar Yarifard " Computer aided decision making for heart disease detection using hybrid neural networkGenetic algorithm" Science Direct 2017.

[14]. A. U. Haq, J. Li, J. Khan, M. H. Memon, S. Parveen, M. F. Raji, W. Akbar, T. Ahmad, S. Ullah, L. Shoista, and H. N. Monday, "Identifying the predictive capability of machine learning classifiers for designing heart disease detection system," in Proc. 16th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process., Dec. 2019, pp. 130–138.

Cite this article as :

Rohini M, Keerthika A, Pavithra N, Sandhiya S, Vedha Smirtha S, "Heart Disease Identification using Machine Learning", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 8 Issue 2, pp. 166-170, March-April 2021. Journal URL : https://ijsrset.com/IJSRSET218247