International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

A REVIEW OF DIFFERENT TECHNIQUES FOR HEART DISEASE PREDICTION

Dr. Sharmila Gaikwad[1], Kanishka Patel[2]

[1]*Assistant Professor, Dept. of Computer Engineering, Rajiv Gandhi Institute of Technology, Mumbai, Maharashtra, India.*
[2]*Student, Dept. of Computer Engineering, Rajiv Gandhi Institute of Technology, Mumbai, Maharashtra, India.*

Abstract - *Currently, In the Healthcare Industry the biggest reason for ill health and impermanence is heart disease. The Risk of Cardiovascular disease is increasing rapidly, demanding the need for various heart disease prediction systems and efficient algorithms. In the medical field, computer Aided Decision Support System plays a major role. It sometimes becomes impractical for patients to undergo costly tests. In such a case, having an automated system for heart disease prediction system would reduce the costs. Moreover, it would improve medical diagnosis. With the advancement in Technological domains, the rampant increase in the research of heart disease prediction is observed. Therefore, it is vital to categorize the various techniques for heart disease prediction and provide an overview of existing systems. In the paper, different techniques and algorithms for heart disease prediction are summarized and compared to analyze the most practical and reliable algorithm.*

Key Words: Cardiovascular, impermanence, rampant, heart disease prediction, algorithms

1. INTRODUCTION

There is no scarcity of records of patients suffering from heart diseases remaining to be the major cause of deaths in the past and the huge amounts of data generated are too complex and voluminous to be processed and analysed by traditional methods [5]. However, the advent of technology and the use of various machine learning and data mining techniques for the development of prediction software as acted as a support for the doctors in diagnosing heart disease beforehand. Presently, the common method used for various diagnoses is most probably based on the doctor's intuition and experience rather than the clinical data available causing unintentional errors and reducing the quality of service provided to patients. In view of such problems caused, the Healthcare industry has started adapting information management systems for elevating efficiency and using various Machine Learning and Data Mining Techniques to extract hidden and abstract information for prediction. Thus, it is of at most importance to find a method that is best for heart disease prediction and globalize it in order to achieve uniform results. In this paper, various methods are studied and compared to make the prediction results uniform and accurate.

2. LITERATURE REVIEW

2.1 Dataset

S. no.	Name of the Attributes	Description
1.	Age	Age(years)
2.	Sex	Man=1, women=0
3.	Cp	Chest pain type
4.	Rbp	Resting Blood pressure upon hospital admission
5.	Chol	Serum Cholesterol in mg/dl
6.	Fbs	blood sugar during fasting >120 mg/dl true=1 and false=0
7.	Resting ECG	Resting electrocardiographic Results
8.	Thalach	Maximum Heart Rate
9.	Induced Angina	Does the patient experience angina as a result of exercise (value 1: yes, value 0: no)
10.	Old Peak	ST elevation during rest
11.	Slope	Heart rate slope
12.	Thal	Value 3: Normal ,value 6:fixed defect, value 7: reversible defect
13.	CA	Count of major vessels (value 0-3)
14.	Num	Heart disease Diagnosis (0 = healthy; 1 = low; 2 =medium 3 = high; 4= very high)

Fig -1: Dataset Attributes

All three methods have used the same dataset for prediction purposes. The Cleveland dataset part of the University of California Irvine (UCI) [6] machine learning repository data set was used [1][2][3]. The dataset has answered in the form of num values, absence means zero and presence means any value according to the chance of cardiovascular diseases as low, high, medium, very high. The dataset has various features or attributes. The main objective for the Review

paper is how can we turn data into useful information that can enable healthcare practitioners to make effective clinical decisions. [4]

2.1 Proposed Algorithms

A. Logistic Regression Algorithm

The system is designed to read data and classify it using Logistic regression. Data are classified according to the features into whether the patient has heart disease or not. The data is used to create a model that would predict the disease. To increase the efficiency different machine learning algorithms are used.

1) Sklearn Logistic regression

It is a sigmoid function used to represent data in the form of graphs for easy evaluation and analysis. It can be implemented using various libraries in python. In this algorithm, the data is imported and trained to provide high accuracy. The graphs are used to compare the various attributes in order to estimate the best and approximate coefficient to represent it. By using Sklearn Logistic regression a score is calculated.

2) Naive Bayes

It is a supervised classification algorithm used for the classification of large datasets. It uses conditional probability theorem for determining the class of the vector. Conditional probability values of vectors are computed, and new vectors class is formed. It is performed to achieve accurate results for the prediction and removal of correlated data.

3) Comparing and confusion matrices

It is used for summarization and analysis of the results which are classified based on attributes. The predictions whether true or false are marked with the help of count values. A confusion matrix is used to calculate various performance measures. It explains the performance of the characterization model and is represented in the form of a table. The analysis of the dataset and results get is performed easily by using Comparing and confusion matrices. Thus, an accurate prediction is made.[2]

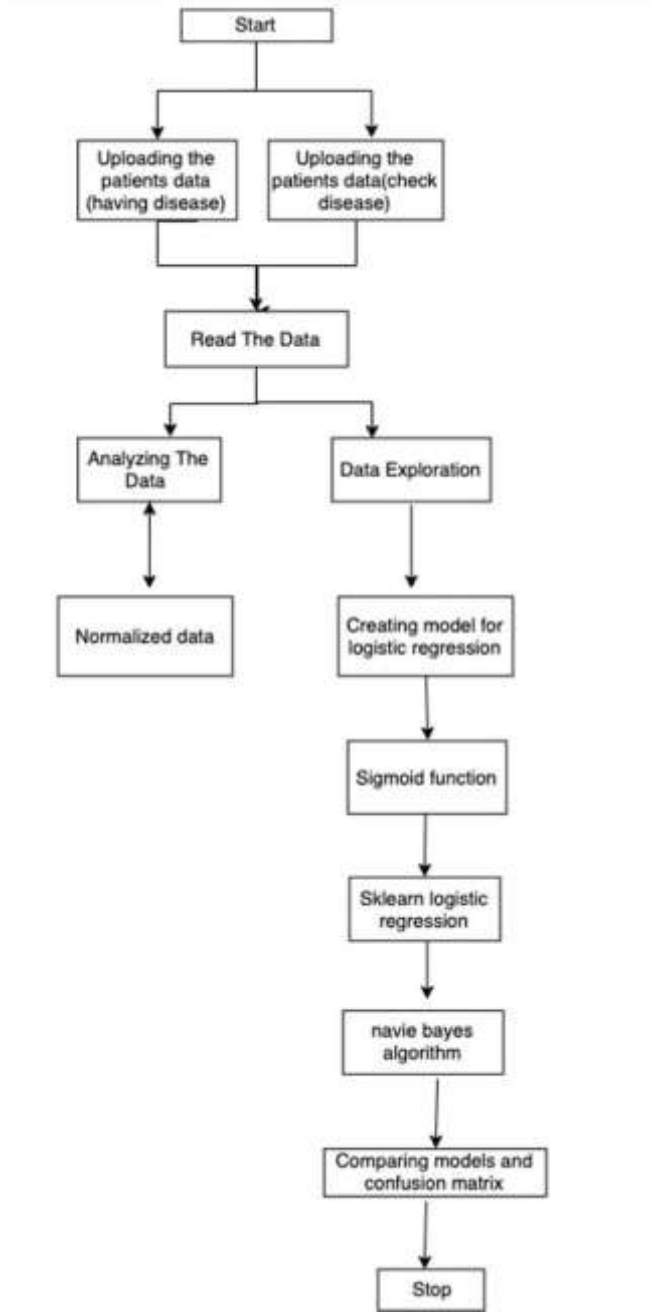

Fig -2: Implementation of Logistic Regression Algorithm [2]

B. Naïve Bayesian

This method utilizes Naive Bayesian - data mining classification technique for heart disease diagnosis. The supervision of different medical factors and the post-operational period is very important. The patients' records/data are encrypted using AES and are saved in a database. The results predict the risk level that is associated with heart diseases.

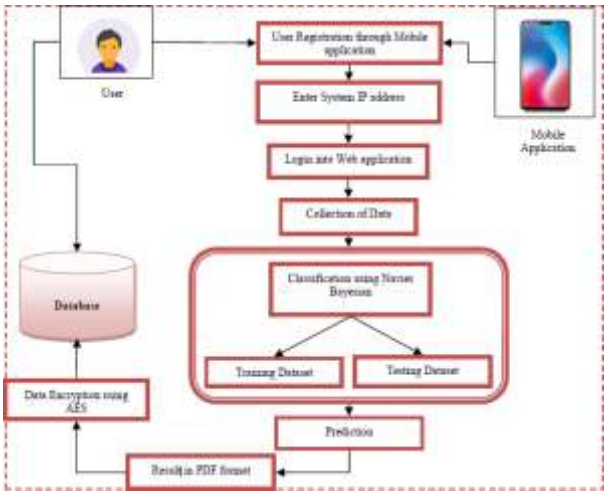

Fig -3: Proposed Architecture [1]

1) User Registration and Login:

The first step of the process is user registration wherein through a mobile application the user fills up the registration form. After successful completion of user registration, using the system IP (internet protocol) address the user can log in anytime by using his/her own username and password. Every registered user's credentials are saved in the database. After this, the complete symptoms list is given including the affected clinical features like age, sex, cholesterol, sugar, ECG, chest pain, Rest blood pressure, etc [1]

2) Navies Bayesian-based Classification

A Naive Bayesian (NB) classifier, also called the "independent feature model" is based upon the Bayesian theorem. The NB classifier presumes that the existence/absence of a specific class feature is independent of the existence of the other class feature.

3) Algorithm

- Step 1: Say D represents the training set and each record denoted by n-dimensional attribute vector, this means X=(x1,x2…, xn), predicting measurements from n attributes (say A1 to An.)
- Step 2: Consider m no: of classes for prediction (say C1, C2……. Cm)

  By Bayes' theorem: P (Ci | X) =

  P(X | Ci)∗ P(Ci)

  P(X)
- Step 3: Since P(X) being a constant for every class, hence P (X|Ci)*

  P(Ci) must be maximized.
- Step 4: Thereafter class conditional independence is presumed.

  Thus,

  P(X| Ci) = P(x1| Ci)* P(x2| Ci) P(xm| Ci)
- Step 5: For predicting class of X, P(X|Ci)P(Ci) is computed for every class Ci.Naive Bayes classifier predicts that class label of X = Ci class if

  P(X|Ci)P(Ci) > P(X|Cj)P(Cj)

  for 1 ≤ j ≤ m, j ≠ i [1]

4) Prediction

The existing method for the prediction of various medical diseases used includes the experience and knowledge of experts and doctors. But, with the advent of technology medical diagnosis systems were invented that brought a huge shift in the medical domain. Such systems analyze all the factors that may cause heart disease like patients test records, personal data, medical screening data along with the experience and knowledge of doctors that help give an accurate diagnosis. It has reduced the time consumption as well as the cost, resulting in a significant system.

5) Security for AES

AES(Advanced Encryption Standard) is a very strong symmetric encryption algorithm. It is the most used algorithm as it uses bytes for performing various operations. It works in two phases Encryption and Decryption. The encryption phase involves Byte substitution, Shiftrows, MixColumns, and Addroundkey while the decryption phase involves the same steps in reverse order. It is a very secure algorithm no attacks have been reported against it. Thus, sensitive data like medical data or patient details should be encrypted with the AES algorithm. It is very flexible it can generate results in PDF format also.

C. MapReduce

1) Meta-heuristic with prepared RFNN (Recurrent fuzzy neural network):

The recurrent fuzzy neural network was used that had 13 input layers, 7 hidden layers, and 1 output layer. The genetic algorithm-based neural network was used to train the dataset, 64 bits genes were used as weights and parameters for the genetic algorithm were the probability 0.05 of mutation, 0.25 of multipoint crossover and the population size was 100. Thus, a genetic algorithm-based neural network was created.

2) Evaluation criteria

The invention of the Hadoop distributed computing platform has made the parallel processing of all the machine learning algorithms easy. Thus, machine learning algorithms can be transformed into the Mapreduce paradigm easily using HDFS. (Hadoop Distributed File System). Here, it shows the estimation analysis of the genetic algorithm along with the trained neural network and MapReduce algorithm. The algorithms which were proved efficient like the Heuristic approach with prepared RFNN approach and ANN-

Fuzzy_AHP were incorporated with Mapreduce in order to achieve the highest accuracy in prediction. Mapreduce is being used for several reasons.

- It increases the processing time speed by increasing the count of nodes in the cluster.
- Its model can run across multiple nodes that save a lot of time without affecting the accuracy.
- It is efficient and performs well with large datasets.

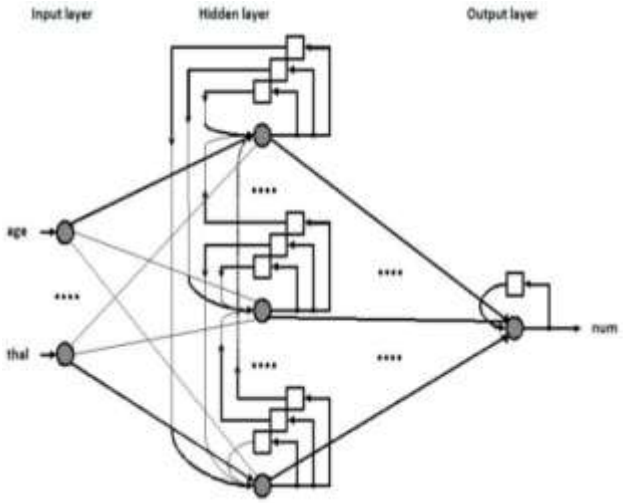

Fig -4: The structure of RFNN [3]

The genetic and Distribution algorithm was used to process and optimize the training data along with a trained neural network approach in order to work in parallel using Mapreduce eventually reducing the training time. The best approach for classification and regression is the Meta-heuristic approach with a prepared neural network. Some characteristics of the Meta-heuristic approach with prepared neural network approach based on the MapReduce algorithm.[3]

- Parallelization can be achieved by distributing, processing and optimizing the training data and directing them to participate nodes.
- The parallel approach reduces the training time and helps achieve scalability and performance requirements for large scale data mining.
- There is an increase in the computing and storage requirement of genetic algorithm with trained neural network in proportion to a number of training vectors addressed.

3. INFERENCE

Table -1: Accuracy and Features of the Algorithms

Algorithm used	Accuracy	Features
Logistic Regression Algorithm	86.89	The main advantage of this building software using this algorithm is that any nonmedical employee can also use it and predict heart disease. The data is cleaned and mined using the traditional method to create a dataset for Logistic regression –Machine Learning Algorithm can be used for predicting if the patient has heart disease or not.[2]
Navies Bayesian (NB)	89.77	In this method, a Smart Heart Disease Prediction is established using the Naïve Bayesian algorithm along with the AES algorithm for resolving the security issue in Heart disease Prediction. Although, less number of attributes were taken into consideration the accuracy of the system was not compromised.[1]
MapReduce Algorithm	98.12	This method yields the highest accuracy in comparison to the other two. The accuracy is achieved due to dynamic schema and linear scaling. It mainly implements the MapReduce algorithm by comparing the meta-heuristic approach along with a trained persistent fuzzy neural network for heart disease prediction.[3]

The Each algorithm has their own pros and cons and it is difficult to determine which among them works best. The most common formula for calculating accuracy can be described as follows.

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

where,
True Positive (TP) = the condition where the number of records classified as true are actually true.

False Positive (FP) = the condition where the number of records classified as true are actually false.
False Negative (FN) = the condition where the number of records classified as false are actually true.
True Negative (TN) = the condition where the number of records classified as false are actually false.[3]

4. CONCLUSIONS

To conclude this research venture, we tabulate the various features and the accuracy of each method discussed above.

Each method has been tested using a common dataset. The Logistic Regression Algorithm uses the Machine Learning concept for prediction of Heart Disease and mainly focuses on the application of this algorithm for classification. The Naive Bayesian algorithm is used along with the AES encryption algorithm to provide higher security of data used for prediction as compared to PHEA. And in the Mapreduce Algorithm, it uses a metaheuristic approach and fuzzy neural network, here Hbase is used to store the resultant data which can further be used to get more accurate results in comparison to other mining techniques as the batch processing size is reduced.

Future works include applying the techniques discussed above to gain higher accuracy and merge these techniques in order to get the best possible results. Combining these techniques will help establish a more efficient way for heart disease prediction. Some other potentially powerful techniques can also be reviewed for other researchers to benefit from. Also, including the various machine learning and artificial intelligence-related techniques would help find new, quick and efficient prediction techniques

REFERENCES

[1] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin "Design And Implementing Heart Disease Prediction Using Naives Bayesian" Proceedings of the Third International Conference on Trends in Electronics and Informatics , ICOEI 2019

[2] Reddy Prasad,Pidaparthi Anjali, S.Adil, N.Deepa "Heart Disease Prediction using Logistic Regression Algorithm using Machine Learning" International Journal of Engineering and Advanced Technology , IJEAT ,Volume-8, Issue-3S, February 2019

[3] T.Nagamani, S.Logeswari, B.Gomathy "Heart Disease Prediction using Data Mining with Mapreduce Algorithm" International Journal of Innovative Technology and Exploring Engineering , IJITEE ,Volume-8 , Issue-3, January 2019

[4] Purushottam , Prof. (Dr.) Kanak Saxena, Richa Sharma "Efficient Heart Disease Prediction System using Decision Tree" International Conference on Computing, Communication and Automation (ICCCA2015)

[5] Beant Kaur, Williamjeet Singh "Review on Heart Disease Prediction System using Data Mining Techniques" International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 2 , Issue: 10

[6] Aditi Gavhane, Gouthami Kokkula, Isha Pandya, Prof. Kailas Devadkar (PhD) "Prediction of Heart Disease Using Machine Learning" Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology, ICECA 2018