International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 709-713 www.ijaem.net ISSN: 2395-5252

Prediction of Heart Disease with Advanced Machine Learning Techniques

K.Aruna[1], Dr D Suresh Babu[2], Dr.Amit Jain[3]

Scholar, Department of Computer Science & Engineering, SunRise University, Alwar, Rajasthan
Head, Department of Computer Science SR&BGNR Government College, Khammam, Telangana
Professor, Department of Computer Science & Engineering, SunRise University, Alwar, Rajasthan

---

Submitted: 05-07-2022 Revised: 15-07-2022 Accepted: 18-07-2022

ABSTRACT
Heart diseases are commonly caused and when neglected becomes life threatening. So, early detection of the disease is very important and for diagnosis to save lives. There can be many parameters that are to be considered to predict the heart disease. Some of them are like age, cholesterol, blood pressure levels. Etc., Here we are going to implement Machine Learning model to predict heart disease. Heart diseases are commonly caused and when neglected become life-threatening. So, early detection of the disease is very important for diagnosis to save lives. There can be many parameters that are to be considered to predict heart disease. Some of them are age, cholesterol, and blood pressure levels. Etc., Thus, all these parameters act as features. Based on these features model is developed. Machine Learning is a branch of Artificial Intelligence that is widely used to make predictions. Machine Learning Algorithms build a model based on the data that is provided as a training data. The model thus developed is capable of making predictions. Machine Learning has wide range of applications in the health care. The algorithms that are used for the implementation includes Naïve Bayes Classifier, decision tree algorithm, K- Nearest Neighbor and Random Forest algorithms

Keywords— Heart Disease Detection, Naïve Bayes, Decision Tree, Random Forest, K-Nearest Neighbour (KNN).

I. INTRODUCTION

We can see that from past ten years the major death cause for humans was occurring because of heart disease. Heart is the most important involuntary organ in a human. Heart attacks have become common in both older as well as young people due to improper diet, lack of exercise and stress. Heart diseases are life threatening and may sometimes cause to death. Early detection of diseases can predict or taking proper diagnosis. Prediction of heart disease symptoms might not be easy. So, a machine learning model can be used as a alternative to the diagnosis at the hospital. Heart attack can occur because of either narrow or blocked vessels. Some of the blood vessel diseases are life coronary artery disease and heart rhythm problems. Personal and professional practices, as well as genetic susceptibility are hazard factors for heart disease. To avert death, the well planned, precise and premature medical detection of heart disease plays a deciding part in proceeding benefits. Machine learning help in the exposure and diagnosis of distinct diseases. They are ML algorithms as follows K-Nearest Neighbor, Decision Tree, Naïve Bayes and Random Forest are correlated to find the most precise model. Supervised Learning concerns training on a labelled dataset using techniques to achieve exact awareness applying dependent and independent variable. In this project the algorithm is given with input variables and actual output obtained then algorithm compares between the actual and predicted output to identify errors and modifies the model precisely. The heart disease database is from the UCI repository. It consists of 303 instances and 14 different attributes which are the deciding factors of predicting a heart disease. The output we get in this project is the machine learning model which helps in predicting if a person has a chance of suffering from heart disease or not.

The knowledge-rich nature of the healthcare domain has made it an ideal surroundings, where knowledge on data mining should also have to be expanded further for the increasing need. However, the abstract nature of tacit healthcare knowledge has resulted in the under-utilization of such a fundamental component of the overall healthcare delivery system. There are

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 709-713 www.ijaem.net ISSN: 2395-5252

many algorithms for these problems, but they are not precise and perfect [3]. In this thesis, a new approach has been utilized named M-tree to predict heart disease. In addition to this a comparative assessment has been made with k-Mean clustering algorithm.

II. BACKGROUND

The heart disease patient's data is gathered from UCI repository. This data is availed to discover the disease using machine learning algorithms. The algorithms performance and accuracy outcomes are correlated. In this they mainly focused on Random Forest algorithm. This was proposed in the year 2020 [1]. The discovering patterns is done with ML algorithms such as NN, Genetic Algorithm and other algorithm for dataset. The outcomes are contrasted for execution and precision and the calculations. Electrocardiogram is used to analyze heart cycles with many beginning points [7]. For making prediction of heart diseases in a simple and efficient approach we have used advanced methods. With the help of machine learning and deep learning we can perform various experimentation methodologies which are used in this study [8]. Dimensionality reduction by using two methods Feature Extraction and Feature Selection is proposed by Ramalingam VV et al in the year 2020. Large number of features or attributes can direct to overfitting which results in poor output [4].

Karthick et al., show how classification approaches can be used to predict the existence of cardiovascular disease in those under the age of 50. According to a latest survey, ten percent of Americans suffer with cardiovascular disease, and is also the top cause of mortality [2]. Heart disease can strike anyone, including infants, toddlers, and teenagers. Machine learning and classification methods were utilized by Killana Sowjanya et al. to predict heart illnesses in their early stages [1].They concentrate on techniques such as the Apriori algorithm in conjunction with support vector machines.

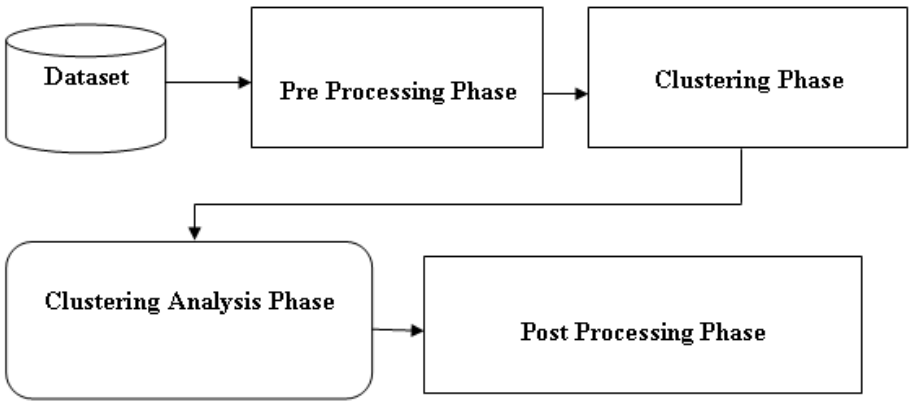
Azhar Hussein et al. used the particle swarm optimization technique in their paper. They employed a model that focuses on the pso algorithm, which were improved and compared based on accuracy [3].The Hidden naive bayes method was utilized by M. A. Jabbar et al. for heart disease prediction.The researchers collected data from the University of California at Irvine's heart disease databases and developed it using the WEKA tool [4].

III. IMPLEMENTATION MODULES

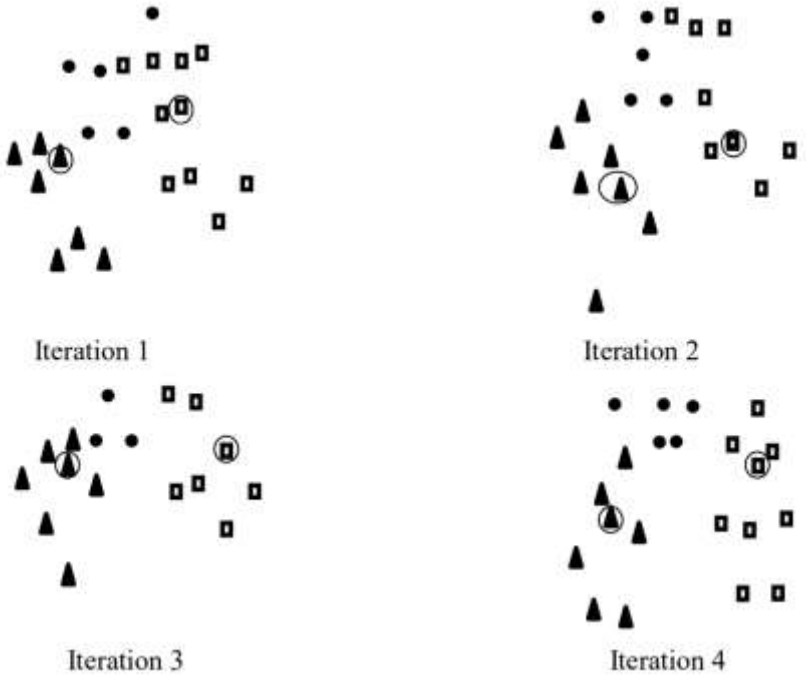
For disease prediction, different categorization algorithms are evaluated. There were two sorts of models built: single models and hybrid models, which were a combination of the two or even more models. In comparison, hybrid models are found to be more accurate [5].Different methodologies, as well as big data analysis, are used to build predictive analysis in the health care [6].

Implementation Modules:

Module- 1 :- Pre-processing
Module - 2 :- Training the DAE
Module - 3 :- DE noise Auto Encoder
Module - 4 :- Training the DAE
Module - 5 :- Plotting of Trained Data
Module - 6 :- Creating the Classifier
Module - 7 :- Training New Model
Module - 8 :- Evaluation of Classifier



International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 709-713 www.ijaem.net ISSN: 2395-5252



The centroid (mean) of the i cluster is defined by the following Equation:

$$c_i = \frac{1}{m_i}\sum_{x \in c_i} x$$

Step 3 and 4 in algorithm directly challenge to minimize the SSE. Step 3 forms group by conveying points to nearest centroid which minimized SSE for the given centroid. Step 4 recomputed the centroid so as the further minimize the SSE [42].

IV. EXPERIMENTATION AND RESULTS

In the survey research work, Analysis of algorithms is finished based on the Cleveland data sets. The algorithm used in the project for the final model is Random Forest. Now, we will show the result of M-tree algorithm of predictive value of heart disease data set the details off result shown below. The following tables show the analysis of the negative and positive test of the patients, unpredicted value of the instances and accuracy of the M-tree algorithm on PIMA Indian heart disease dataset. Further, we will k-means approach of heart disease dataset for prediction of positive and negative record set of patients, unpredicted value of the instances and accuracy of the M-tree algorithm on PIMA Indian heart disease dataset.

0 test negative	1 test positive	Test done by model
454	46	Negative
23	245	Positive test

Table 5.4 Analysis report of M-tree m

Model for heart disease data set

0 test negative	1 test positive	Test done by model
380	120	Negative
135	133	Positive test

Table 6.5 Analysis report of k-means model for heart disease data set

This table shows the final result of our analysis. It shows sum of errors/In corrected instances, Percentage of errors and accuracy.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 7 July 2022, pp: 709-713 www.ijaem.net ISSN: 2395-5252

Factors	Sum of errors/ In corrected instances	Percentage of errors	%(accuracy)
M-tree	69	8.999	92.125
k-means	255	33.023	66.977

Table 5.6 M-tree and k-means model detect for unpredicted value of instances and accuracy based on heart disease data set.

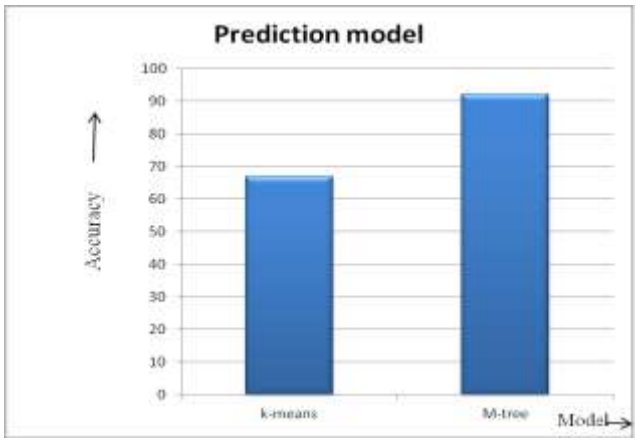


Figure: Graph of Prediction models

Algorithm	Accuracy
Naïve Bayes	82.45614035087719%
K-Nearest Neighbor	77.19298245614034%
Random Forest	78.94736842105263%
K-Means	68.891%
M-Tree	91.2867%

Table: Prediction Table

V. CONCLUSIONS

This is the final output where we give the input in the interface, which gives the output whether the person has heart disease or not. This is the final output where we give the input in the interface, which gives the output whether the person has heart disease or not. We have added several elements like label to display the description of features and textbox to enter the values of features, on clicking the predict button, the model will classify the data to either of the classes.

In the future where every field is getting automated, it is very important to automate health care sector as well. The proposed project work gives a model as the outcome that is used to predict the heart disease. An intelligent system may be built that may help a patient with heart disease choose the best treatment options. A lot of work has already gone into developing a model for predicting whether or not someone will develop heart disease. We can embed this application into a real time system which has sensors that measure certain attributes. This will help us get predictions and alerts in real time based on the body condition of the user.

REFERENCES

[1] C. Sowmya and P. Sumitra, (2017). Analytical study of heart disease diagnosis using classification techniques,'' in Proc. IEEE Int. Conf. In tell. Techno. Control, Optima. Signal Process. (INCOS), Mar. 2017, pp. 1–5

[2] Purusothaman, G. and Krishnakumar, P., "A survey of data mining techniques on risk prediction: Heart disease", Indian Journal of Science and Technology, Vol. 8, No. 12, (2015).

[3] Smys, S. "Survey on accuracy of predictive big data analytics in healthcare." Journal of Information Technology 1, no. 02 (2019): 77-86.

[4] Pavithra V. and Jayalakshmi V.

``Comparative study of machine learning classification techniques to predict the cardiovascular diseases using HRFLC.” International conference on intelligent computing and control systems (ICICCS 2021), ISBN. 978-0-7381-1326-5.

[5] Pavithra V. and Jayalakshmi V. “Hybrid selection technique for prediction of cardiovascular diseases” Materials today: proceedings, 2021.

[6] Bashir, S., Khan, Z.S., Khan, F.H., Anjum, A., Bashir, K.: Improving heart disease prediction using feature selection approaches. In: 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), pp. 619–623. IEEE (2019).

[7] Tarawneh, M., Embarak, O.: Hybrid approach for heart disease prediction using data mining techniques. In: International Conference on Emerging Internetworking, Data & Web Technologies, pp. 447–454. Springer, Cham (2019).

[8] Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." IEEE Access 7 (2019): 81542-81554.

[9] Savla, Anshal, et al. "Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture." 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). IEEE, 2015.

[10] S Satapathy and S Chattopadhya, Mining Important Predictors Of Heart Attack, International Conference On Advances In Recent Technologies In Communication And Computing 2011

[11] G.Subbalakshmi, K. Ramesh, M. Chinna Rao , Decision Support in Heart Disease Prediction System using Naive Bayes, ISSN : 0976-5166 Vol. 2 No. 2 Apr-May 2011.