Heart Disease Prediction System Using Machine Learning

SANDHYA WAGHERE[1], AISHWARYA PATIL[2], MADHUVANTI KULKARNI[3], BHAVESH GIDWANI[4], SAURABH MARATHE[5]

[1, 2, 3, 4, 5] *Department of Information Technology, Pimpri Chinchwad College of Engineering, Pune, India*

Abstract— In today's world most of the deaths are occur by heart disease and it is very difficult to diagnose this before and also people are ignorant about their about the kind off life style they live that cause heat diseases. This research paper focuses on the people who very likely to get a heart disease on various different parameters. Our aim is to find if a person is likely to get a heart disease in the near future by considering their past medical history also by some conditions which lead to heart disease. We used algorithms such as logistic regression and, Random Forest which will help us do this task more efficiently.

Indexed Terms-- Heart disease, Diagnose Logistic regression, random forest.

I. INTRODUCTION

It is very important to predict the heart disease at very earlier stage so that patient can take necessary action on the particular heart disease. According to world health organisation, people who are in their mid-thirties are most likely to get diagnosed with heart diseases because of the standard of living and also their food habits and also factors like alcohol, cigarettes, physical inactivity, stress, lot of consumption of caffeine, overweight. So, an early diagnoses of heart disease play a significant role in decreasing the number of deaths caused by it. There are a number of heart diseases know to us such as cardiac arrest, arrhythmia, coronary artery disease, blood pressure, peripheral artery disease and many more.

II. MACHINE LEARNING

Machine learning---Machine algorithm is basically used to train the data and then the trained data gives computer and trained the module and is used for predicting the new data .it is important component of growing field in data science. As machine learning technology advances, it has certainly made our lives easier.it solves our daily problems .it makes prediction for the database.

Machine learning falls into three main types:
--Supervised machine learning---it is way of training the model where we pass actual data as well as the data on which we want to predict on the basis of actual data which works as labelled or training data model will predict the output on our data which we are predicting algorithm used for supervised learning are regression and classification.

--Unsupervised machine learning---in unsupervised learning there is no training data or actual data for model.so the model is not trained using trained or actual data instead of learning from training data the model itself find the patterns from the respective data .it is exactly same like learning human brain the process is same like human brain does new things. Unsupervised learning is clustering and association.

Semi-supervised machine learning---Semi-supervised machine algorithm contains both supervised and unsupervised algorithm.it uses label data as well as self-train data to train the model.

III. WORK FLOW

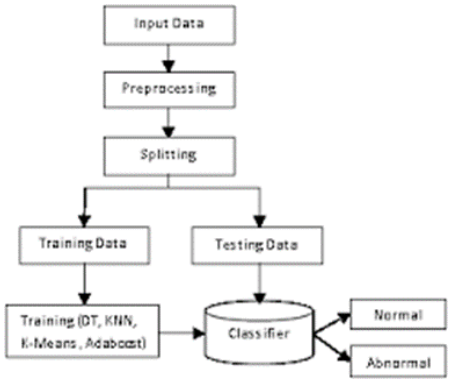


Fig. work flow diagram

IV. LITERATURE SURVEY

[1] K. Gomathi, (2016) presents multiple disease predictions with the use of techniques. With the advancement of medical science and machine learning, separate experiments and research have been done in recent years, which has led to the publication of some very important publications. Using WEKA software, helps in heart disease prediction. The precision of the algorithms is still unsatisfactory. They concluded that the Gaussian Nave Bayes and Random Forest provided a high accuracy of 91.2 percent in a study using the Cleveland heart rate database, which includes the emergence of 303 and hired 10-fold Cross Validation, testing 13 features and using 4 various functions.

[2] Ashwini Shetty A et al (2016), The study was conducted using four trained models and tested with great accuracy using the same database from Framingham, Massachusetts. 87 percent for K Neighbors Classifier, 83 percent for Support Vector Classifier, 79 percent for Decision Tree Classifier, and 84 percent for Random Forest Classifier.

[3] Heart Attack Prediction was done by Abhay Kishor uses Deep Learning (2018). This research done the prediction of heart related disease and also predicts infection regarding heart . Recurrent neural network is a state-of-the-art characterization computation that employs a deep learning technique. The paper discusses in depth the key modules of the framework as well as the associated hypotheses. The suggested algorithm uses deep learning and data mining to deliver accurate results with least errors.

[4] Bo Jin, Chao Che et al. (2018) studied, a model "Predicting the risk of heart fail with EHR data tracking" map out by initiating neural networks. In this paper used virtual health record (EHR) data from main datasets related to congestive heart disease to performs experiment and predict heart disease earlier. We use one-hot encryption with word vectors to model the incidence of diagnosis and predict the incidence of heart failure that falls victim to the essential principles of an extended memory network model. In results, we knew the status of respecting the sequential nature of medical records.

[5] Akash Chauhan et al. (2018) proposed "PREDICTION OF HEART DISEASE USING EVOLUTIONARY RULE LEARNING". This will remove the physical work which additionally helps in pull information or data directly from the virtual record. For output generate strong association rules, all of us have applied constant pattern growth association mining to patient datasets. This will help to minimize the capacity of services and generate the overwhelming quantity of regulations help with the first-rate prediction of coronary disease.

V. PROPOSED METHOD

In this research work we focused more on machine learning. thus, it is easier to divine the accuracy of result using various machine learning algorithms like we used k-nearest neighbor, random forest, logistic regression. Comparison between different algorithms is done based on accuracy, time complexity, advantages, and disadvantages of all these algorithms. Python language is use for implementation. We used postman for hitting API's for getting our output.

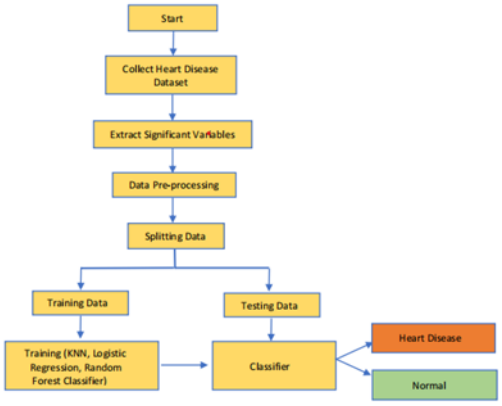


Fig. proposed method

So next part is about getting dataset. Here we used dataset in excel format. Dataset includes various parameters. The dataset consisting of the parameters including age, sex, chest pain type, serum cholesterol, resting blood pressure etc. Here, 70% of the dataset is

taken for training and the remaining is considered for testing. After collecting dataset data process into various parts. The data get splits into different way based on range provided to the parameters. Data get splits into two different ways. Some of the data goes for training and some goes for the testing. In training data irrelevant and unspecified data gets trained by machine learning algorithms. And thus, we get the results in specific format. Comparison of different algorithms is done. And the algorithm which gives high accuracy that is consider for better result. When we get the results, we got to know the person is having heart disease or not.

VI. IMPLEMENTATION

This research contains three different algorithms as random forest, logistic regression and k – nearest neighbour. As mentioned earlier this research contains 13 different attributes on which we compare the results. we take data from Kaggle as it contains various data. For experiment we used vs code and postman which is open tool. Postman is used for API testing.

VII. METHODOLOGY

ALGORITHMS USED:

1. K-nearest Neighbour

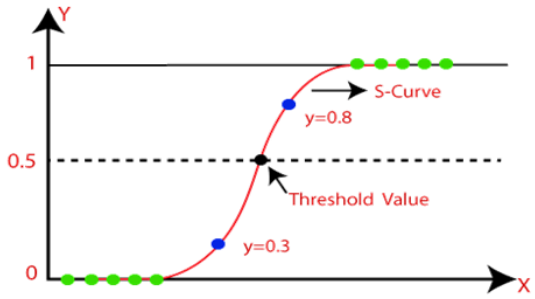
It is a supervised machine learning algorithm and it is non-parametric (which means it does not assume any underlying data).It is used for both classification and regression problems.It stores the data in the training phase and when it gets a new dataset it then classifies it into a new category that acts like a new dataset-nearest neighbor.

2. Random Forest

Random forest is a supervised machine learning algorithm. It can be used for classification problems like in prediction of heart disease. Random forest combines many decision trees and takes vote from each tree; hence it takes long time to process and is slower than decision tree. It has the same parameters as a decision tree. Random forest deals with problems like variance which is not solved in decision tree. The main advantage of random forest is that it can handle large data sets.

3. Logistic Regression

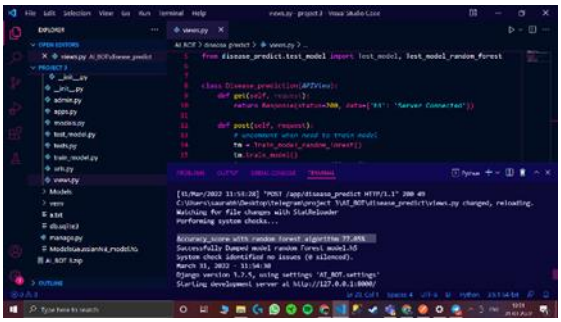
Logistic regression is similar to linear regression but it solves classification problems. It is a supervised machine learning algorithm used for predicting categorical dependent variable using a given dataset. As we see in the graph below we have a ‘S’ shaped logistic function instead of a regression line. Logistic regression is simple to implement and it give output in a Boolean form either True or False.

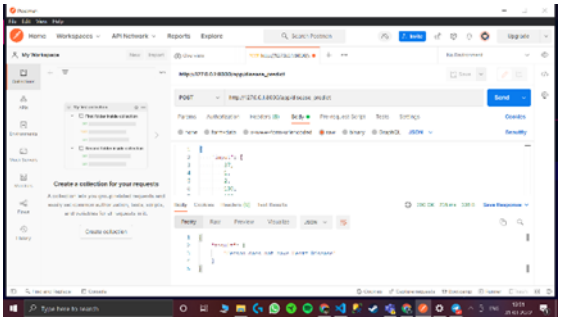


VIII. RESULTS AND FUTURE SCOPE

Here after implementing code and processing the data in different algorithms the result is ready with the three different algorithms and three different accuracies. From the results we get more accuracy from random forest algorithm.

Output-





FUTURE SCOPE

1. people could predict heart diseases remotely with increasing development of machine algorithms.
2. In future we can add AI BOT which will take direct input from user which is going to help people to use more friendly.
3. The future scope of this system aims at giving more sophisticated prediction models, risk calculation tools and feature extraction tools for other clinical risks.

IX. CONCLUSION

The whole purpose is to describe the various machine learning strategies that are useful in the effective prediction of heart diseases. We plan to do effective and right prediction with a small number of parameters and tests. In this study, we will consider only 14 important parameters. We have used 3 machine learning strategies, k-nearest neighbour, random forest, logistic regression, we can further expand this research with data mining techniques such as vector support, and the Naïve bayes algorithm. Considering the limitations of this study, there is a need to use complexity and combination of models to achieve high accuracy in predicting early heart diseases.

REFERENCES

[1] Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine learning improve cardiovascular risk prediction using routine clinical data? PLoS ONE. 2017;12(4):e0174944.

[2] Ramalingam VV, Dandapath A, Raja MK. Heart disease prediction using machine learning techniques: a survey. Int J Eng Technol. 2018;7(2.8):684–7.

[3] Bouali H, Akaichi J. Comparative study of different classification techniques: heart disease use case. In: 2014 13th international conference on machine learning and applications. IEEE. p. 482–86.. 10, no. 4, pp. 277–284, Dec. 2012.

[4] Sanjay Kumar Sen 1, Dr. Sujata Dash 21Asst. Professor, Orissa Engineering College, Bhubaneswar, Odisha– India.

[5] DomingosP andPazzani M. "Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier", in Proceedings of the 13th Conference on Machine Learning, Bari, Italy, pp 105-112,1996.

[6] Elkan C. "Naive Bayesian Learning, Technical Report CS97-557", Department of Computer Science and Engineering, University of California, San Diego, USA, 1997.

[7] B.L DeekshatuluaPriti Chandra "Reader, PG Dept. Of Computer Application North Orissa University, Baripada, Odisha – India. Empirical Evaluation of Classifiers Performance Using Data Mining Algorithm". 10, no. 4, pp. 277–284, Dec. 2012.

[8] Sanjay Kumar Sen 1, Dr. Sujata Dash 21Asst. Professor, Orissa Engineering College, Bhubaneswar, Odisha– India. Paper Title:- HEART DISEASE PREDICTION SYSTEM USING MACHINE LEARNING ISSN:-2349-3585 |www.ijrdt.org

[9] DomingosP andPazzani M. "Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier", in Proceedings of the 13th Conference on Machine Learning, Bari, Italy, pp 105-112,1996.

[10] Elkan C. "Naive Bayesian Learning, Technical Report CS97-557", Department of Computer Science and Engineering, University of California, San Diego, USA, 1997.

[11] B.L DeekshatuluaPriti Chandra "Reader, PG Dept. Of Computer Application North Orissa University, Baripada, Odisha – India. Empirical Evaluation of Classifiers Performance Using Data Mining Algorithm"

[12] Brown N, Young T, Gray D, Skene A M & Hampton J R (1997). Inpatient deaths from acute myocardial infarction,1982-92: analysis of data in the Nottingham heart attack register. BMJ, 315(7101), 159-64.

[13] Folsom A R, Prineas R J, Kaye S A & Soler J T (1989). Body fat distribution and self-reported prevalence of hypertension, heart attack, and other heart disease in older women. International journal of epidemiologyy, 18(2), 361-7.

[14] Chen A H, Huang S Y, Hong P S, Cheng C H & Lin E J (2011, September). HDPS: Heart disease prediction system. In 2011 Computing in Cardiology (pp. 557-60). IEEE.

[15] Parthiban, Latha and R Subramanian. "Intelligent heart disease prediction system using CANFIS and genetic algorithm." International Journal of Biological, Biomedical and Medical Sciences 3.3 (2008).

[16] Wolgast G, Ehrenborg C, Israelsson A, Helander J, Johansson E &Manefjord H (2016). Wireless body area network for heart attack detection [Education Corner]. IEEE antennas and propagation magazine, 58(5), 84-92.

[17] Patel S & Chauhan Y (2014). Heart attack detection and medical attention using motion sensing device -kinect. International Journal of Scientific and Research Publications, 4(1), 1-4.

[18] Jee S H, Jang Y, Oh D J, Oh B H, Lee S H, Park S W & Yun Y D (2014). A coronary heart disease prediction model: the Korean Heart Study. BMJ open, 4(5), e005025.

[19] Ganna A, Magnusson P K, Pedersen N L, de Faire U, Reilly M, Ärnlöv J &Ingelsson E (2013). Multilocus genetic risk scores for coronary heart disease prediction. Arteriosclerosis, thrombosis, and vascular biology, 33(9), 2267-72.

[20] Jabbar M A, Deekshatulu B L & Chandra P (2013, March). Heart disease prediction using lazy associative classification. In 2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s) (pp. 40- 6). IEEE.

[21] Dangare Chaitrali S and Sulabha S Apte. "Improved study of heart disease prediction system using data mining classification techniques." International Journal of Computer Applications 47.10 (2012).

[22] Soni Jyoti. "Predictive data mining for medical diagnosis: An overview of heart disease prediction." International Journal of Computer Applications 17.8 (2011):43-8.

[23] Chen A H, Huang S Y, Hong P S, Cheng C H & Lin E J (2011, September). HDPS: Heart disease prediction system. In 2011 Computing in Cardiology (pp. 557-60). IEEE.