Heart Disease Prediction System with Nearest Clinic Suggestion

Prof. Chitra Bhole[1], Mayuresh Pujari[2], Jainam Jatakia[3] and Ulkesh Chendwankar[4]

[1, 3]Computer Engineering, KJSIEIT, Mumbai, India
Email: {cbhole, jainam.jatakia}@somaiya.edu
[2, 4]Computer Engineering, KJSIEIT, Mumbai, India
Email: {pujmay, ulkeshchendwankar1}@gmail.com

Abstract—**Heart disease is one of the most critical disease to cure in health care sector. People above age 50 and above have possibility of facing so many diseases that might be linked with heart. But some facts have claimed that recent lifestyle has made people in their adolescence also stand a chance of getting heart disease. Being said heart disease is critical disease also the data generated from patients are in terabytes, this data can be helpful in making radical decisions. We can gain good results by using various machine learning techniques. In this paper we will study about a Heart Disease Prediction Software (HDPS) which is developed using Hybrid Algorithm including Linear Logistic and KNN for analyzing the presences of heart disease. The system uses 15 medical parameters such as age, blood pressure, sex, cholesterol, etc. for prediction. The HDPS will predict likelihood of patients having heart disease. The obtained results will be illustrating that the designed diagnostic system can effectively predict the heart disease. If the user is found out of having the disease then it suggests the nearest Cardiologist or Physician's Clinic for Treatment using Google API.**

Index Terms— **Logistic Regression, K-Nearest Neighbor, Google API.**

I. INTRODUCTION

There are 79 million unit calculable presently associate with diabetes in Indian sub-continent whereas another 200 million unit are suffering from hypertensive. It will perhaps increase dramatically with growing population. Interruptions in diagnosing, treatment, and regular check-ups for these patients will additional translate into the next disorder burden. disorder symptoms is also totally different for men and womens. as an example, men more likely to have chest pain; womens are likely to have alternative symptoms at the side of chest discomfort, like shortness of breath, nausea and extreme fatigue The patient won't be diagnosed with cardiovascular disease till you have got a heart failure, angina, stroke or cardiopathy. It's important to look at for cardiovascular symptoms and discuss considerations with the doctor. Cardiovascular disease will generally be found early with regular evaluations [7].

Today various doctor's facilities stores reasonably data frameworks to manage their social welfare or patient data. These knowledge frameworks ordinarily manufacture heaps of data which may be in distinctive organization like numbers, content, diagrams however unfortunately, this information that contains knowledge, which is pure data used by clinics, can be used for choice making.[8] Nearly millions of people are affected due

to the changes in their lifestyle. It occurs due to blockage of cholesterol in arteries and it will lead to heart stroke. When the person is getting stroke the pumping of blood over the body will not be in an efficient way. Another reason of heart disease can occur due to hypertension which is one of the main reasons. Heart disease is one of the prevalent diseases where many peoples are getting affected.

In this system we are implementing an effective heart disease prediction system using the KNN and logistic regression algorithm. We can give the input as in a CSV file or manual entry to the system. After taking input the algorithms apply on that input that is KNN and logistic regression. After accessing the data set the operation is performed and effective heart disease prediction is produced. Also it will recommend the nearest medical facility that the patient can visit. Helping him to reach the medical facility as quickly as possible, thus helping to save the life of the patient. Also recommending the nearest Cardiologist or Physician and thus helping the patient seek required medical attention as quickly as possible and thus reducing a possible heart attack.

II. LITERATURE SURVEY

Various machine learning based diagnosis techniques have been proposed by various researchers to diagnosis Heart disease. This analysis study presents a number of the prevailing machine learning based mostly identification techniques to clarify the vital of the planned work.

- Detrano et al. [1] developed HD classification system by using machine learning classification techniques and the performance of the system was 77% in terms of accuracy. Cleveland dataset was utilized with the method of global evolutionary and with features selection method.
- In another study, Gudadhe et al. [2] the system that was made of multi-layer Perceptron and SVM algorithms for High Definition classification and achieved accuracy of 80.41%.
- Palaniappan et al. [3] proposed an expert medical diagnosis system for HD identification. In development of the system the predictive model of machine learning Decision Tree (DT), and Artificial Neural Network were used. The DT classifier achieved 80.4% accuracy.
- Jayamin Patel et. al. [4] this paper instructed J48 provides 56.76% that is healthier than LMT algorithmic rule of accuracy 55.75% [4]
- Ashok Kumar Dwivedi et. al. [6] in step with this paper KNN logistical Regression gave 80% and 85% respective accuracy severally.
- During this following paper they suggested prime 5 specialist cherish the expected unwellness and filters choosen by the user victimization coreNLP.[5]

KNN and logistic regression gives the classification accuracy much better than the above approaches and the reduction of attributes gives more accuracy in predicting the results and also that it will recommend nearby clinics for the user using google API with the help of longitudes and latitude and type of place to search that's clinic.

III. Existing System

Very few systems uses the clinical available data for prediction purpose and even if they do, they are restricted by large no of association rules that apply. Diagnosis of the condition solely depends upon doctor's intuition and patient records. Existing system had KNN as prime algorithm it did not give satisfactory accuracy

IV. PROPOSED SYSTEM

In this system we are implementing effective heart attack prediction system using KNN algorithm and Logestic Regression. We can provide the input as in comma-separated values (CSV) file or manual entry to the system. After taking input the algorithms apply on that input that is KNN and logistic regression. After accessing data set the operation is performed and prediction result is produced. The heart attack prediction system designed to help the identify presence of disease and also giving the prescription details with related to the predicted result with suggestion of the nearest clinic for further treatments.

V. METHODOLOGY

A. Dataset Collection

A data set (or dataset) is a assortment of information. is a case of tabular knowledge, a knowledge set corresponds to at least one or a lot of information tables, wherever each column of a table represents a selected

variable, and every row corresponds to a given record of the info set in question[8]. Kaggle is one of the most popular online community websites for data science and machine learning algorithms.

TABLE I. DATASET PARAMETERS INFORMATION

Sr. No.	Attributes	Description
1	Age	Age (years)
2	Sex	Male or Female
3	Cp	Chest pain type
4	Thestbps	Resting Blood Pressure
5	Chol	Serum Cholesterol
6	Restecg	Resting electrographic results.
7	Fbs	Fasting Blood Sugar.
8	Thalach	Maximum Heart rate achieved
9	Exang	Exercise induced angina
10	Oldpeak	ST depression induced by exercise relative to rest.
11	Slope	Slope op the peak exercise ST segment
12	Ca	No. of major vessels colored
13	Thal	Defect type

B. KNN (K-Nearest Neighbour Algorithm)

K-Nearest Neighbour is used for both classification and regression technique. The basic idea of this algorithm is they use various datapoints as inputs and with these datapoints they derive the output that is full of assumption[9].

C. Graphical Representation

This classifier looks for the classes of K nearest neighbours of a given data point and based on the majority class, it assigns a class to this data point. However, the number of neighbours can be varied. Neighbours can be varied them from 1 to 20 neighbours and calculated the test score in each case.
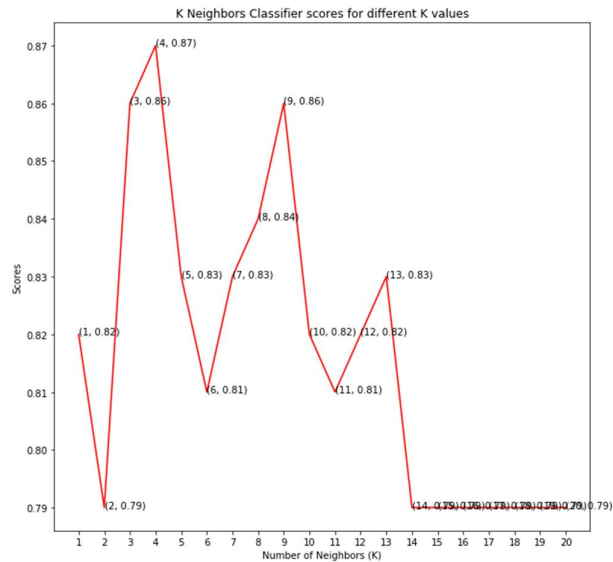

Fig.1: KNN (K-Nearest Neighbour Algorithm) Accuracy

Figure 1. Data points and class Mainly the KNN uses the supervised data i.e. the target data for predicting the output. When k=3, it gave accuracy of 87% and when k=2, it gave accuracy of 86% . The main usage of loss

351

function in KNN there is no training phase and during testing k neighbours are predicted with minimum distance for both regression, classification. It is one of the easy and simple machine learning algorithm. The computation cost is high when the data's are large.

D. Logistic Regression

Logistic Regression is not like a regression model instead it is like a classification model. This algorithm gives the output in the form of binary values i.e. like 0's and 1's. The figure 3 shows the final output of the algorithm logistic regression. The figure 2(a) explains the relationship between precision and recall which is important in logistic regression model. The second graph, figure 2(b) explains the relationship between true positive rate and also false positive rate. Both precision, recall, true positive rate, false positive rate always lies between the values from 0 to 1.
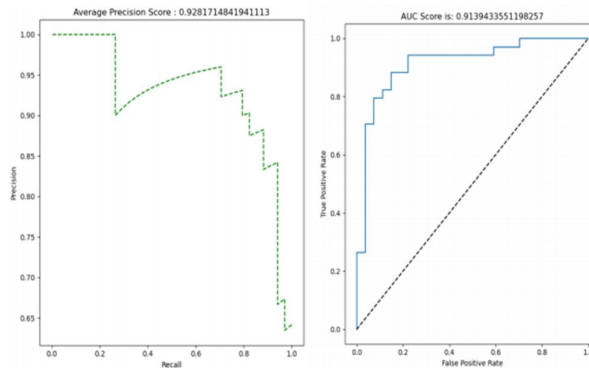

Fig.2(a) and 2(b): relation between precision and recall, true positive rate and also false positive rate

There are two category which is true positive and true negative similarly false positive and false negative. If my output gives that the person having the heart disease is positive if my output gives that the person having heart disease instead it gives in opposite way it is negative. True positive rate means if the person is having the heart disease. False positive rate means if the person is having heart disease but my output gives that he is not having the disease means it is nothing but false positive rate. The AUC score gives either positive value or negative value i.e. the deducted score should be in terms of numerical value so that the users can understand clearly and the score was literally high. The graph is drawn mainly for predicting the disease. Another graph is about precision and recall. Precision means percentage of the data which is relevant and recall means percentage of total relevant result that are correctly classified by the model.
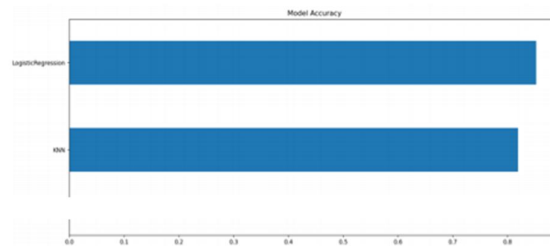

Fig.3: Accuracies of KNN and Logistic Regression

So, both the approaches will be we be used will running the system and if any of the both approaches produce the output as disease detected will produced the output of whole system as disease detected with users input.
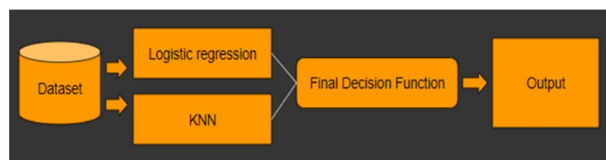

Fig.4: flow diagram of proposed system

E. Suggestion Module

In Suggestion module, location using the current location, this could be achieved using Python and Google Maps API. The system will use Google Maps API to find the nearest hospital's locations using Python. It will initialise Google Places constructor and then call the function google places. Nearby search with parameters as latitude, longitude, radius and type will be clinic and store the output in a variable. User will offer his own latitude and longitude within the parameters. The attributes value will be written like longitude, latitude.

VI. CONCLUSIONS

The proposed system will have more accuracy from previously implemented system with was solely based on KNN algorithm; but current system will process both logistic regression and KNN algorithm if disease found from any of the algorithm the patient will be acknowledged. This system can be help in reducing check-up cost and treatment costs by providing initial diagnostics in time. Physicians can also utilize this as a tool for initial diagnosis of cardio-patient. As this system can be still be explored to improve the scalability and accuracy.

REFERENCES

[1] 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.

[2] 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

[3] S. Palaniappan and R. Awang, ''Intelligent heart disease prediction system using data mining techniques,'' in Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl., Mar. 2008, pp. 108–115

[4] Patel, Jaymin & Tejalupadhyay, Samir & Patel, Samir. (2016). Heart Disease prediction using Machine learning and Data Mining Technique. 10.090592/IJCSC.2016.018.

[5] Dhanashri Gujar, Rashmi Biyani, Tejaswini Bramhane, Snehal Bhosale, Tejaswita Vaidya, "Disease Prediction and Doctor Recommendation System",International Research Journal Of Engineering and Technology (IRJET), Pune, India, Maar-2018

[6] Dwivedi, Ashok. (2016). Performance evaluation of different machine learning techniques for prediction of heart disease. Neural Computing and Applications. 29. 10.1007/s0052.

[7] https://www.biospectrumindia.com/news/79/17285/heart-disease-marginally-increases-compared-to-pre-covid.html

[8] Purushottama,c* , Prof. (Dr.) Kanak Saxenab , Richa Sharma, (2016)"Efficient Heart Disease Prediction System",ELSEVIER.

[9] https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761