Heart Disease Prognosis System with Nearest Clinic Recommendation.

Chitra Bhole [1]

Professor, Department of Computer Engineering, K. J. Somaiya Institute of Engineering and Information Technology, Mumbai, India
cbhole@somaiya.edu

Ulkesh Chendwankar [1], Jainam Jatakia[1], Mayuresh Pujari [1]

Undergraduate Students, Department of Computer Engineering,
K. J. Somaiya Institute of Engineering and Information Technology, Mumbai, India
ulkesh.c@somaiya.edu, jainam.jatakia@somaiya.edu, mayuresh.pujari@somaiya.edu

***Abstract*- Now-a-days, any age group of people can suffer through heart attack or any other heart disease so there's need for people to regularly do check-ups. Those check-ups can be costly if they are done directly at the cardiologist or physiologist for their particular test reports generated. So, what we intended to build a system which general clinic doctors can use for their patients for checking chances of getting heart disease by either making them undergo the test in their clinics or from already generated test reports for testing labs. So, we thought to use machine learning algorithms KNN and Logistic regression for computing the entries filled by general doctors of user to be tested. This system finds the final output with help of Cleveland Heart Disease Dataset from UCI machine learning repository and current patient data entries. After the generating final output, it suggests the location of the nearest Clinic or Hospitals in patient's vicinity. This suggestion feature is mostly for proper educated users who are knowledgeable about their reports parameters and who track their health regularly.**

Keywords— Logistic Regression, K-Nearest Neighbour. Cleveland Heart Disease, UCI machine learning repository

I. INTRODUCTION

According to Centres of Disease Control Prevention (CDC), Heart Disease is the leading cause of death in the United States and among 4 deaths every 1 is due to the result of heart disease. That sum the amount of almost 610,000 deaths from heath disease conditions each year [1] and the statics are predicting that 45% of the people in America will have at least one issues related to the disease by 2035 from American Heart Association (AHA). AHA also predicts that costs related to disease might get doubled till 2035 which could bankrupt the nation's economy and health care system; since the complication from heart disease are spreading faster than their original thoughts [2]. Even thought other diseases like cancer, diabetes is gaining more attention, but still cardiovascular disease remains the world most costly killer and most of time, people may not realize the risk or audacity of heart disease until they know someone who suffer from it or they themselves are suffering from it.

Previous decade had witnessed huge advances in the quantity of data that is usually generated and collected. Various Private and public sector industries generally generate tons of data that can be stored and analysed for the improvement of their services. Rather that increasing the profits and cutting down overheads; analysing the past data from Health care industry to predict epidemic, pandemic, curing diseases, increasing quality of life. These decisions can be made due to the changes driven by the events happening in the current or past to predict what can happen in the mere future. So, the focus can be now solely depending on patients records for detecting any signs of diseases early and making the treatment quicker and less painful. Nearly millions are affected by heart disease due to changes in their lifestyle. Mainly it is caused by blockage of cholesterol in arteries and which can lead to heart stroke and can be caused by hypertension.

In this system, we have implemented a heart disease prediction application which works on K Nearest Neighbours and Logistic regression algorithm. Initially, we input the values in the application as per the fields provided in system which is then analysed with respected to the dataset that we provided to the system containing 14 fields including a target field which helps in predicting when given processed inputs what output can be predicted from current inputs. This processing and analysing is done by K Nearest Neighbours and logistic algorithm. After that our system will predict the output and display it in window whether there is presence of heart disease and if there is presence of heart disease it will suggest the clinics with certain distances.

II. LITERATURE SURVEY

Before building the actual system, we did some research on the various ML based diagnosis techniques that have been already proposed in research papers. This analysis study presents a number of the usual machine learning based mostly identification techniques to clarify the vitality of the proposed system.

- Detrano et al. [3] developed they developed and Heart disease classification which gave accuracy of 77% in terms of accuracy. They used Cleveland dataset in this paper.
- Mohammed Jaeead Ali Junaid et al. [6] They proposed Naïve Bayes, ANN, SVM which gave accuracy of 82.97%, 85.30%, 86.12% respectively.
- Palaniappan et al. [5] proposed an Heart disease System using Decision Tree and ANN having accuracy of 80.41%
- Pahulpreet Singh Kohli, Shriya Arora, [4] used LR, DT, Adaboost with accuracies of 87.1%, 70.97%, 83.87% respectively.
- Khairina, D. M. (2017), Haversine formula is used to find the courier location of the nearest web-based JNE, the result of this research is a web that can be accessed by people in Samarinda, Indonesia [8].

- Dhanashri Gujar, Rashmi Biyani et al. [7] had recommended best 5 specialist by filtering the data and using Core NLP.

Logistic Regression and KNN has the classification accuracy better than above discussed methodsin predicting the results and the Suggestion of the nearest clinic will be done by obtain users gps coordinates and then using the Haversine Formula within location coordinates of Clinic in the database.

III. Existing Systems

Prior to any system, the patient had to visit the cardiologist with their report wherein cardiologist go through their report and tell the patient about their anomalies in health parameter and if there are serious report output, they would had got admitted to hospital and further so. So, this process is still used as priority and most trustful method. But still there are some efforts made to build system to automatize this analyses process. This system used one of the various algorithms of machine learning like Support vector method, Linear Regression, K Nearest Neighbour but this method did not provided sufficient accuracy.

IV. Proposed System

In this system we are implementing effective heart attack prediction system using KNN algorithm and Logistic Regression. We have to input manually in the input field according to the label given aside. After taking inputs we perform analysis on those inputs with K Nearest Neighbour and logistic regression algorithms. After accessing data set which is in Comma Separated Values format the operation is performed and prediction result is produced. The heart attack prediction system designed to help the identify presence of disease and with suggestion of the nearest clinic for further treatments.

V. System Architecture

Proposed System act as a whole providing the user with analysis on their inputs. This system gets the input from the user which with the help of dataset provided to it and KNN and logistic algorithm is processed on the Dataset as well as user inputs through which a final decision function generate the output which is activated if either of the algorithm provides the output as disease present as shown Figure 1.

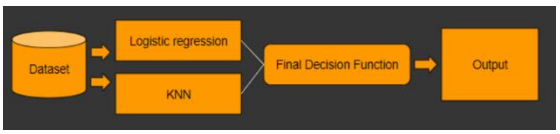

Figure 1: System Architecture of Prediction Module.

After generation of Output, if user has presences of heart disease, then there will be an suggest button which will take the user over the browser and ask for the access of location as you can see in figure 2, user's current GPS location is sent by using POST request. POST method is one of the formats for transmission of data in PHP for sending sensitive data. Furthermore, this location is processed with locations of the various other clinics present in the MySQL database using Haversine Formula and then generated response of all the clinics data within a particular range is sent in form of JSON (JavaScript Object Notation) to the User interface. We have used postman application to create an API for saving complex http request and for the creation of map we have used leaflet.js one of the JavaScript libraries used for creating interactive maps.

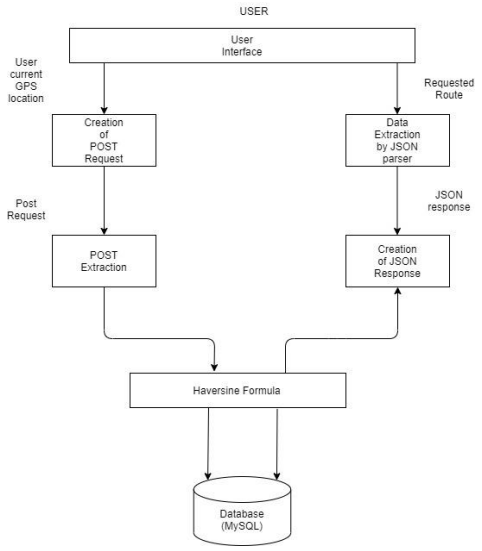

Figure 2: System Architecture of suggestion Module.

B. *Dataset Collection*

The dataset we used is Cleveland database and it contain 14 attributes as given below in Table 1.

S.N.	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
14	Target	Target value

Table I. Dataset Parameters Information

VI. Implementation

The implementation included two modules namely prediction module and suggestion module. These application within modules are built for optimized of easy

usage of use for the users and to get approximate location of nearest cardiologists clinic for user within the database.

A. Prediction Module

Prediction module consist of KNN and Logistic regression which are built using Python libraries like sklearn with an User Interface made of Tkinter which is also Python Library.

1. KNN (K-Nearest Neighbour Algorithm)

K- Nearest Neighbour Algorithm is used for classification of given input whether they belong any of the given class and in which class. There can be k classes. For, our system we used k=3 so the accuracy of our system is 87% as shown in figure 3 below.

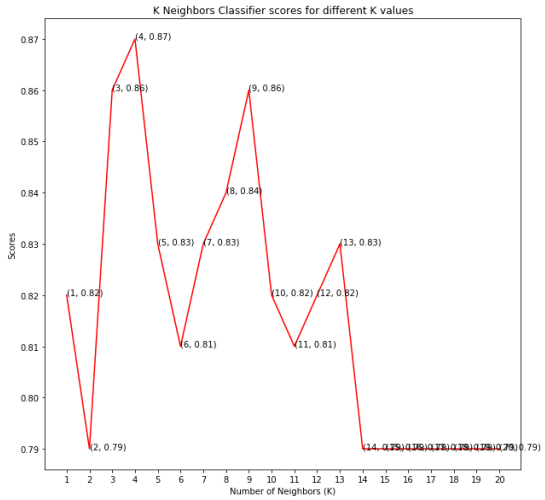

Figure.3: KNN (K-Nearest Neighbour Algorithm) Accuracy

2. Logistic Regression

Logistic Regression is an algorithm that gives the output in the form of binary values i.e., like 0's and 1's. Threshold is used for distinguishing this values. When we did accuracy check it gave us accuracy of 88.14% as shown in figure 4.

```
In [30]: from sklearn.linear_model import LogisticRegression
         logreg=LogisticRegression()
         logreg.fit(x_train,y_train)
         y_pred=logreg.predict(x_test)

In [31]: sklearn.metrics.accuracy_score(y_test,y_pred)
Out[31]: 0.881491344873502

In [35]: Accuracy of the model is 0.88
```

Figure 4: Logistic Regression Accuracy.

B. Suggestion Module

In Suggestion module, we take users current location using gaps based on user's current location nearby clinics are shown having distance of 5km. We have used Haversine formula for the calculation of distance of user and clinics. We have used postman application to create an API for saving complex http request. For the creation of map, we have used leaflet.js one of the JavaScript libraries used for creating interactive maps.

1. Haversine Formula:

Haversine Formula is the calculation of the distance from a point to another point on the surface of the earth. This calculation is affected due to certain degree of curvature [10]. This formula calculates the considered distance accurately unlike other methods like choice of distance calculation. So, Haversine formula is a method that exactly is calculation of the distance between the distance with two places longitude and latitude data.

Haversine formula: d =

$$2r\arcsin\sqrt{sin^2\left(\frac{\phi_2-\phi_1}{2}\right)+\cos(\phi_1).\cos(\phi_2).sin^2\left(\frac{\psi_2-\psi_1}{2}\right)}$$

Where d is the distances in Km, r is radius of earth (637 km), Φ is the latitude and Ψ is longitude. 1 degree is equal to 0.0174532925 radians.[9]

VII. RESULTS

Each Module of our System i.e., Prediction Module and Suggestion Module is depicted with following figures.

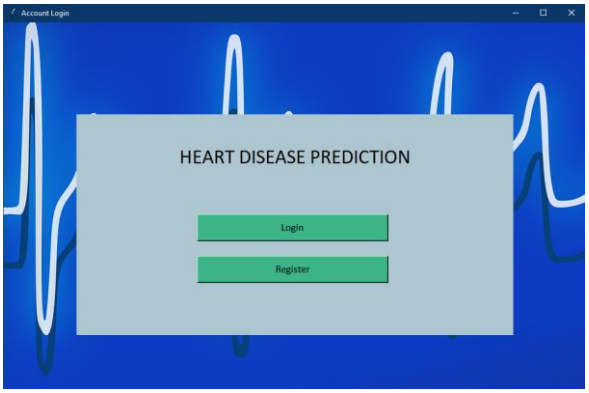

Figure 5: Login/Register Window of the System.

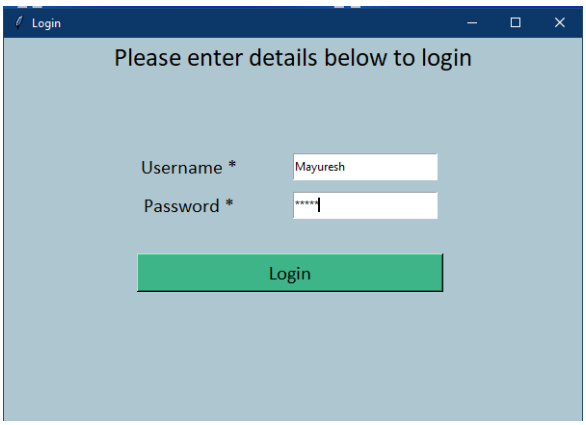

Figure 6: Login Window of the System.

Figure 7: Home Window of the System.

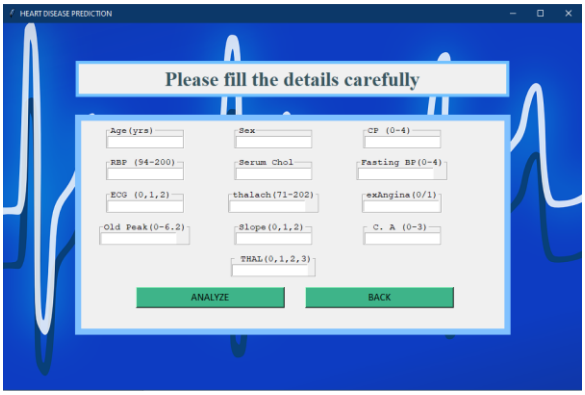

Figure 8: Analysis Window of the System

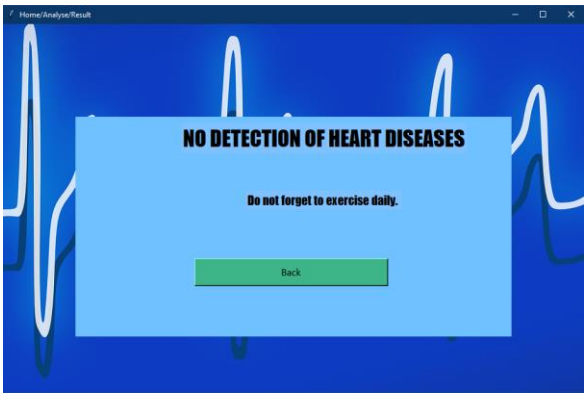

Figure 9: Result Page (Absence of Disease)

Figure 10: Result Page (Presence of Disease)

Figure 5 show the Window wherein user can login or register into the system.

Figure 6 shows the Login window where user has to enter his credentials.

Figure 7 shows the Home window wherein there are 3 option namely Analysis, about us and Log out.

Figure 8 shows Analysis window where the user has to enter all the input fields and press Analyse Button to start prediction on their inputs.

Figures 9 shows Window which is produced by the input which weren't leading to presences of heart disease.

Figure 10 shows Window which appears when user enters the input are leading to presence of heart disease. Here user can press the suggest button for suggesting a clinic nearest to their current location, this button will take user to browser opening the webpage.

Figure 11: Suggestion module Home Page

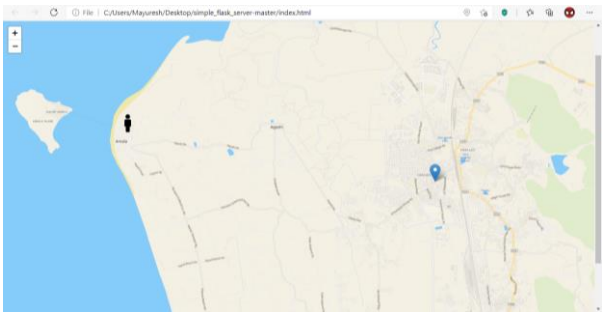

Figure 12: Suggesting a Clinic Page.

Figures 11 shows the webpage when user press the Suggest button so here when user press Nearby Clinics, they may be asked of get access to their current location as a pop-up where they have to enter Ok button.

Figure 12 shows the webpage where they are been suggested nearest clinic around their current location.

VIII. Conclusion

The Proposed system will be used by the general practitioner doctor or user with knowledge about the input parameters which will be enter from referring to the report of their tests and this it can be used by the medical students as a simulator. This system has improved accuracy since we are using both KNN and Logistic Regression, so this system will basically find if either of the algorithms produce output as presence of disease then it gives it has final output. This system can be help in reducing check-up cost and treatment costs by providing initial diagnostics in time. The User then will be suggested the clinic nearby with its name, the city in which it is located, the rating of the clinic on to the map with also their current location.

IX. Future Scope

In the future, our system can be improvised by using more accurate algorithm then current system uses and also the system can have single image input which will extract inputs from the image of the report and do analysis on those input so by this way user may not need doctor to input their values. Systems Suggestion system can be improvised by adding more features like tracking user's current location and showing the path towards the destinated nearest clinic. Mainly this system can be switched from desktop to mobile version so more users can use it.

References

[1] https://www.healthline.com/health/heart-disease

[2] https://www.healthline.com/health-news/why-is-heart-disease-on-the-rise

[3] 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.

[4] Pahulpreet Singh Kohli, Shriya Arora, "Application of Machine Learning in Dease Prediction", International Conference on Computing Communication and Automation (ICCCA),2018.

[5] 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

[6] Mohammed Jawwad Ali Junaid, Dr, Rajeev Kumar, "Data Science And Its Application In Heart Disease Prediction", International Conference on Intelligent Engineering and Management (ICIEM),2020.

[7] 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

[8] Khairina, D. M., Ramadhinata, F. W., & Hatta, H. R. (2017). Determinating of the Nearest Jalur Nugraha Ekakurir (JNE) Office Using Haversine Formula (Case Study in Samarinda). Prosiding SENIATI, 3(1), 10-1.

[9] M.Basyir, M.Nasir, Suryati, Widdha Mellyssa, "Determination of Nearest Emergency Service Office using Haversine Formula Based on Android Platform", EMITTER International Journal of Engineering Technology, Vol. 5, No. 2, December 2017, ISSN: 2443-1168.

[10] Putra, D., & Herwan, R. (2015). Implementation of Haversine method formula in the information system of geographical landscaping. Jurnal Sistem dan Teknologi Informasi (JustIN), 1(1).