Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018)
IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1-5386-0965-1

Prediction of Heart Disease Using Machine Learning

Aditi Gavhane
Department of Information Technology
Sardar Patel Institute of Technology
Mumbai, India
aditipg96@gmail.com

Isha Pandya
Department of Information Technology
Sardar Patel Institute of Technology
Mumbai, India
ishapandya126@gmail.com

Gouthami Kokkula
Department of Information Technology
Sardar Patel Institute of Technology
Mumbai, India
gouthami.kokkula@gmail.com

Prof. Kailas Devadkar (PhD)
Department of Information Technology
Sardar Patel Institute of Technology
Mumbai, India
kailas_devadkar@spit.ac.in

Abstract—**with the rampant increase in the heart stroke rates at juvenile ages, we need to put a system in place to be able to detect the symptoms of a heart stroke at an early stage and thus prevent it. It is impractical for a common man to frequently undergo costly tests like the ECG and thus there needs to be a system in place which is handy and at the same time reliable, in predicting the chances of a heart disease. Thus we propose to develop an application which can predict the vulnerability of a heart disease given basic symptoms like age, sex, pulse rate etc. The machine learning algorithm neural networks has proven to be the most accurate and reliable algorithm and hence used in the proposed system.**

Keywords—**Prediction;** ***heart disease; symptoms; machine learning.***

I. INTRODUCTION

There is no dearth of records regarding medical symptoms of patients suffering heart strokes. However the potential they have- to help us foretell similar possibilities in seemingly healthy adults are going unnoticed. For instance: As per the Indian Heart Association, 50% of heart strokes occur under 50 years of age and 25% of all heart strokes occur under 40 years of age in Indians. Urban population is thrice as vulnerable to heart attacks as rural population. [11]

We thus propose to collect relevant data pertaining all elements related to our field of study, train the data as per the proposed algorithm of machine learning and predict how strong is there a possibility for a patient to contract a heart disease. For the purpose of patients entering data, we suggest to make use of the easily available sensors in watches and cell phones to measure the simple factors.

In the section II, we discuss what all sensors are prevalent in the market and what all symptoms do they measure. Our main aim behind developing the system is to make it user-friendly so that regular monitoring of the patient is made possible. Thus it is of utmost importance that the factors required in the input are most accurate and easily available.

The section III discusses our system design and expounds its workflow and structure. The algorithm we have used is explained in greater detail.

Next we put forth the results of our experiment on sample dataset using the proposed algorithm. Also the statistics using different methods are explained. Finally the paper is described briefly with the conclusion at the end.

II. RESEARCH WORK

To initiate with the work we have started collecting data in each and every aspect towards the goal of the system. In the first place, the research was in the direction of the main causes or the factors which have strong influence on the heart health. Some factors are unmodifiable like age, sex and family background but there are some parameters like blood pressure, heart rate etc. which can be kept in control by following certain measures [4]. Many doctors suggest healthy diet and regular exercise to keep the heart healthy. Following are the parameters which are considered for the study in designing the system which have major risk percentage with respect to CAD [5]

1. Age
2. Sex
3. Blood Pressure
4. Heart Rate
5. Diabetes
6. Hyper cholesterol

7. Body Mass Index (obesity)

The next step was to collect dataset. For this we have used Cleveland dataset from UCI library. The dataset contains as many as 76 parameters describing the complete health status of heart. These parameters are obtained by expensive clinical tests like ECG, CT scan etc. Out of these, the traditional heart disease prediction system uses 13 major parameters [6][7]. Since these parameters require expensive lab tests to find ECG, chest pain type, ST depression etc. To avoid these and to make system less complex we selected above mentioned parameters which can be easily measured using different sensors available in the market. The following research work briefly explains the latest sensors in the market used to measure different parameters.

A. AliveKor

It comes as a touchpad connected to your cell phone (through wireless network) or as a wristband. The touchpad simulates an ECG of the patient on his cell phone via Bluetooth. Thus all the necessary parameters like pulse rate, blood pressure are easily available. The wristband on the other hand, uses finger touch to display the pulse function on the dial. It can also notify an atrial fibrillation.[12]

B. MyHeart

In this system, a number of on-body sensors are used to collect physiological data that are sent wirelessly to a PDA. The information is analyzed and health recommendations are given to the user based on the analysis. [1]

C. HealthGear

HealthGear is an application to track most common indexes like lab tests and physical parameters.[4]
Fields include:
- [Physical] Indexes like Height, Weight, BMI
- Blood Pressure, Hemoglobin, WBC, RBC, Platelets
- [Lipids]: Cholesterol, HDL, LDL, VLDL, triglycerides,
- [Sugar]: Fasting Glucose, after meals, HbA1C. [4]

D. Fitbit

This sensor is used to keep the track of health which has features of measuring pulse rate, BP, calories burned.
After this study, we have concluded with using Fitbit to collect the data which is easily available and less expensive and HealthGear for all the other parameters.

III. PROPOSED MECHANISM

The neural networks have proven to be the most popular and evolving branch of machine learning in recent studies. Now in the proposed system we used the neural network algorithm multi-layer perceptron (MLP) to train and test the dataset.

Multi-layer perceptron algorithm is a supervised neural network algorithm in which there will be one layer for the input, second for the output and one or more for hidden layers between these two layers. Each node in the input layer is connected to output nodes through the hidden layers. The connections between any two nodes are assigned weights to it and the resultant input is calculated using the following formula

$$Y_{in}=\sum_{i=0}^{n} wixi \quad (1)$$

Where x_i is the i^{th} input and w_i is the corresponding weight.

There is another identity input with weight b called as **Bias,** which is added to the node to balance the perceptron. The connection between the nodes can be either feedforward or feedback based on the requirement. Figure 3.1 describes the input output signal flow graph.

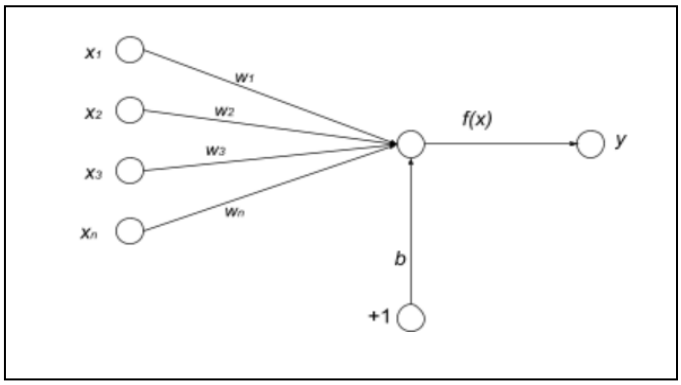

Figure 3.1: Signal flow graph of a perceptron

The activation function is applied on the weighted input to get the output. The role of hidden layer is to connect the input and output layer and to process the data internally. The fig. 4.2 depicts the logical structure of MLP

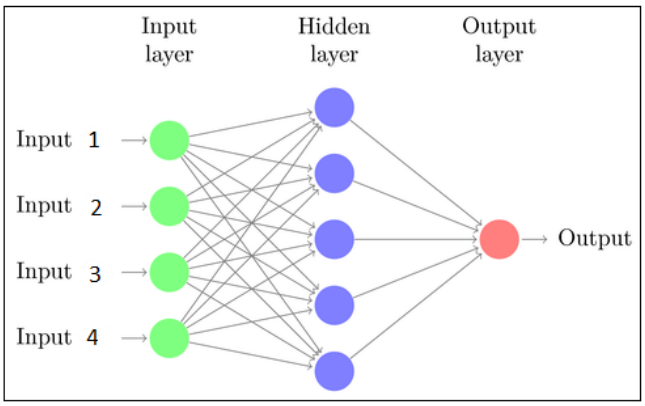

Figure 3.2: Logical Structure of MLP

The system is developed using python code using PyCharm IDE. With the help of python library sci-kit learn, the system is implemented. The sample parameter values taken for MLP function is as shown

```
MLPClassifier(hidden_layer_sizes=(100,),
activation='relu', solver='adam', alpha=0.0001,
batch_size='auto',learning_rate='constant',learning_rate
_init=0.001, power_t=0.5, max_iter=200, shuffle=True,
random_state=None, tol=0.0001,
verbose=False,warm_start=False, momentum=0.9,
nesterovs_momentum=True, early_stopping=False,
validation_fraction=0.1,beta_1=0.9, beta_2=0.999,
epsilon=1e-08)
```

For further details about the parameters visit the sci-kit learn library [6]. Different python libraries such as SciPy, NumPy and panda are used to support the algorithm.

IV. RESULTS

The output of the system will give a prediction result if the person has a heart disease, in terms of Yes or No. The system gives an idea about the heart status leading to CAD beforehand. If the person is prone to have heart disease then the result obtained will be Yes and vice versa. In case of an positive output, he needs to consult a cardiologist for further diagnosis. The statistics of the results obtained during the testing of the dataset is shown in the following table.

Class Label	Precision	Recall	Support
Yes	0.92	0.9	93
No	0.91	0.89	72
Average	0.91	0.89	165

Table 1. Calculations of classification report

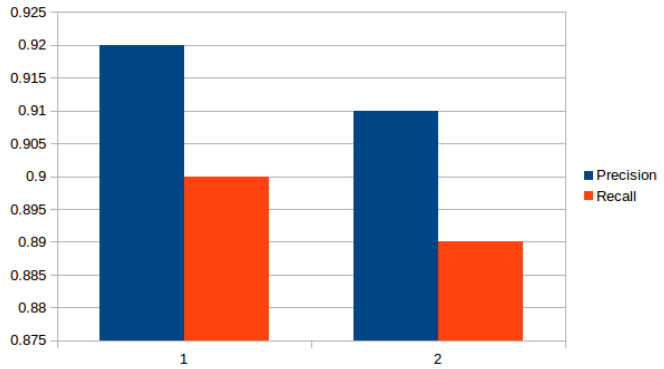

Fig3. Graphical Representation of the results

V. CONCLUSION

The Heart Disease Prediction System using Machine learning algorithm,viz. MLP provides its users with a prediction result that gives the state of a user leading to CAD. Due to the recent advancements in technology, the machine learning algorithms are evolved a lot and hence we use Multi Layered Perceptron (MLP) in the proposed system because of its efficiency and accuracy. Also, the algorithm gives the nearby reliable output based on the input provided by the users. If the number of people using the system increases, then the awareness about their current heart status will be known and the rate of people dying due to heart diseases will reduce eventually.

VI. FUTURE WORK

The similar prediction systems can be built for various other chronic or fatal diseases like Cancer, Diabetes, etc with the help of recent technologies like machine learning, fuzzy logics, image processing and many others. Also, new algorithms can be proposed to achieve more accuracy and reliability. The Big Data Technology like Hadoop can be used to store huge chunks of data of all the users worldwide and to manage the data or reports of the user; technologies like Cloud Computing can be made use of.

REFERENCES

[1] Machine learning based decision support systems (DSS) for heart disease diagnosis: a review. Online: 25 March 2017 DOI: 10.1007/s10462-01

[2]]Prerana T H M, Shivaprakash N C et al "Prediction of Heart Disease Using Machine Learning Algorithms- Naïve Bayes,Introduction to PAC Algorithm, Comparison of Algorithms and HDPS",Vol 3, PP: 90-99 ©IJSE, 2015

[3] Salam Ismaeel, Ali Miri et al "Using the Extreme Learning Machine (ELM) technique for heart disease diagnosis",IEEE Canada International Humanitarian TechnologyConference,DOI:10.1109/IHTC.2015.7238043, 03 September 2015

[4] F BrainBoudi,'Risk Factors for Coronary Artery Disease',2016.[Online]Available:https://emedicine.medscape.com/article/164163-overview.

[5] National Health Council,'Heart Health Screenings',2017.[Online]Available:http://www.heart.org/HEARTORG/Conditions/Heart-Health-Screenings_UCM_428687_Article.jsp#.WnsOAeeYPIV

[6] ScikitLearn,'MLPCLassifier',[Online]
Available:http://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html

[7] Prediction System for heart disease using Naïve Bayes *Shadab Adam Pattekari and Asma Parveen Department of Computer Science and Engineering Khaja Banda Nawaz College of Engineering

[8] Comak E, Arslan A (2012) A biomedical decision support system using LS-SVM classifier with an efficient and new parameter regularization procedure for diagnosis of heart valve diseases. J Med Syst 36:549–556

[9] Ahmed Fawzi Otoom , Emad E. Abdallah , Yousef Kilani , Ahmed Kefaye and Mohammad Ashour(2015)Effective Diagnosis and Monitoring of Heart Disease ISSN: 1738-9984 IJSEIA

[10] N. Oliver and F. F. Mangas, "HealthGear: a real-time wearable system for monitoring and analyzing physiological signals", IEEE International Workshop on Wearable and Implantable Body Sensor Networks, Cambridge, USA, (2006).

[11] r.Amit Mittal,'Increasing heart attacks in young Indians',2017,[Online]Available:https://timesofindia.indiatimes.com/life-style/health-fitness/every-heart-counts/increasing-heart-attacks-in-young-indians/articleshow/56295257.cms

[12] AliveKor,[Online]Available:https://www.alivecor.com/how-it-works.