International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. 05 Special Issue 01 | 2020

Heart Attack Risk Prediction Using Neural Network

[1]Mr. Shobhit K. Khandare, [2]Mr. Udayan Patankar, [3]Mr. Mayuresh Navare

[1,2,3] *Department of Electronics and Telecommunications, Vivekanand Education Society's Institute of Technology*, Mumbai, India

ABSTRACT

Heart attack is one of the most alarming problem faced by individuals in today's world. Millions of people die each year due to heart attack. This can be attributed to growing work pressure, mental stress, unhealthy food habits and ignorance towards one's health. For this, one is often advised to consult a doctor regarding possible risks his/her heart may have and the appropriate measures to be taken to mitigate the risk. Several tests are required to obtain information regarding the health of our heart such as ECG. Then the patient's report has to be carefully scrutinized by the doctor to find out any possible risk. However, this careful study of the patient's report is time consuming and requires a lot of effort. Our project aims to reduce this effort and time by developing a model to predict the risk an individual may have of suffering from a heart attack. An implementation of such a model is discussed in the upcoming sections. Also this model is accompanied by a user interface making it easy to use for the end user. The implementation of this model uses the dataset which includes data collected from various patients about all the relevant parameters required for the prediction of heart attack risk. The data is classified using neural network to predict the level of risk the individual has of suffering from a heart attack.

Keywords—Heart attack, Risk, Prediction, Machine learning, Neural network.

1. INTRODUCTION

A heart attack or in other words, myocardial infarction is a medical condition wherein the blood supply to the heart is blocked [1]. Heart attack occurs when atherosclerotic plaque on an artery supplying blood to a heart muscle ruptures. This promotes the formation of a blood clot in the artery. This can lead to the death of the tissue which is supplied blood by that artery. In order to indicate the patients that they have a risk of getting a heart attack based on their current health parameters, we have developed a system which will predict the risk of heart attack associated with them. This system will take specific health parameters of the patient as an input and on their basis will classify the patients into five categories (Very low risk, Low risk, Medium risk, High risk and Very high risk).In order to make the system easy to use we have developed a user interface. Through this interface the patient will enter his/her data and on clicking submit will be able to see his/her risk level.

2. LITERATURE SURVEY

In [2] the author proposed the use of a binary classification model to predict the risk factor an individual faced of having a heart attack. This dataset used consisted of 14 attributes 13 of which were predictor attributes and one attribute was the variable for indicating binary response whether the individual suffered a heart attack or not. Naïve Bayes classification was used to classify the data. The model was supplemented with a web based user interface enhancing the convenience for the end user.

Authors in [3] developed an android application which would help users predict the risk they have of suffering from a heart attack. After analyzing several attributes of the patients, chi square correlation and Fisher's exact test were employed to generate a risk score. Based on this score the users were classified as having low, medium or high risk. In case the person is classified as having low risk, additional information is taken from the user and the end result is shown to be medium risk to avoid underestimation of the risk.

In [4] the individual were placed in one of four categories based on their age. Then the values of each of the attributes were split to form boundaries for categorization into various risk levels. The prediction was made by assigning different weights to different attributes based on the amount of risk involved in each and every one of them.

In another attempt [5] the authors combined the concept of soft set theory and fuzzy rule with SVM (Support Vector Machine) to precisely identify the risk of having a heart attack among individuals. The entire architecture was divided into four modules namely the soft set creator, fuzzy rule generator, SVM classifier and decision maker. The variation ranges generated by fuzzy rule generator were fed to SVM which classified the individual as moderate risk or high-risk individual.

Ref. [6] provided a survey of various techniques of data mining that are extensively used in medical research. The various techniques used were Naïve Bayes classifier, k-NN technique, Decision list algorithm, Decision tree classifier and an artificial neural network.

Ref. [7] employed a feature selection technique for cardiovascular disease identification using feed forward neural networks. Out of the 17 features taken initially 12 features were selected after all combinations of features were given as input to the neural network and the accuracy obtained for that combination was observed.

In [8] the authors made an attempt to group a set of data objects into clusters which share similar properties. Fuzzy C-means clustering technique was used to partition the data objects from a dataset containing various parameters of heart attack patients. The data objects were classified into three clusters namely those predicted as heart attack, those predicted as no heart attack and wrongly predicted objects.

Here [9] the authors have used the Association rule mining technique for predicting heart attack. This paper uses a different approach for Association mining technique using sequence number and clustering of the transactional data base. Further the authors have displayed the effectiveness of the algorithm.

In this paper [10] the authors have discussed the use of Big Data analytics in heart attack prediction as well as the challenges and future trends. The authors have accessed 31 studies that used Big Data analytics for prediction of heart attack. They have proved how Big Data analytics can be an effective method in the health care field especially for heart attack prediction.

3. METHODOLOGY

3.1 Data Interpretation

The implementation of this system uses the Framingham Heart Study dataset which consists of 4240 entries of men and woman aged between 30 and 62 from the town of Framingham, Massachusetts. However some of the entries in the dataset contained missing values. The missing values were indicated by 'NA'. Therefore, the first step in the process was to clean the dataset. The entries containing an instance of 'NA' were dropped. The dataset also contained columns specifying the education and glucose levels of the individual. Since, these parameters did not contribute significantly in predicting the heart attack risk, these columns were also dropped. Thus, the final dataset which was considered contained 3658 entries with the following parameters:

1) *Gender:* It has been observed that men are more likely to suffer from a heart attack than women. However, the reason for this is not known. But, the higher rates of unhealthy habits among men compared to women seems one of the reason.
2) *Age:* People having age above 65 have higher chance of suffering from a heart attack. This is because as one ages their heart doesn't beat as fast during physical activity compared to when they were young.
3) *Does the person smoke? :* Due to smoking, the walls of the arteries become sticky leading to fatty materials sticking to them. This hampers the capacity of the artery to carry blood to the heart.
4) *Number of cigarettes smoked by a person in a day:* Higher the number of cigarettes smoked by a person in a day, higher is the chance of him/her suffering from a heart attack.
5) *Is the person on blood pressure medication? :* It has been observed that certain medication used to treat high blood pressure caused an increase in risk of heart attack among individuals.
6) *Has the person suffered from a stroke recently? :* It has been observed that people have much higher risk of dangerous heart problems soon after their first stroke compared to people who didn't have a stroke.
7) *Does the person suffer from hypertension? :* Hypertension is a case when systolic blood pressure rises above 130 mmHg and diastolic blood pressure rises above 80 mmHg.
8) *Does the person have diabetes? :* Adults with diabetes are twice a more likely to die of a heart attack than people who don't have diabetes.
9) *Total cholesterol:* Excess cholesterol builds up in walls inside arteries leading to a process called atherosclerosis. It narrows the arteries obstructing the blood flow.
10) *Systolic blood pressure:* Normal level of systolic blood pressure is said to be under 120. In the range of 120 – 129 it is said to be elevated.
11) *Diastolic blood pressure:* Normal level of systolic blood pressure is said to be under 80.
12) *BMI (Body Mass Index):* BMI is given by weight in kg divided by the square of the height in meters.
13) *Heart Rate:* Normally heart beats at around 60 – 100 beats per minute. But, it has been observed that a heart rate above 76 causes higher risk of heart attack.

3.2 Data Pre-processing

The scale of input and output variables are an important factor while training the model. Without scaling the learning process of the model becomes slow or unstable. In case of large input values, the model learns large weight values which leads to the model being unstable and give a poor performance. Scaling is done in order to bring the input and output variables in the range of 0 – 1. We have used the normalization technique of scaling data which is given by

$$y = \frac{x - \min(x)}{\max(x) - \min(x)}$$

3.3 Classification

We first started with classifying the person into two categories namely heart attack and no heart attack. For this we implemented a feed forward neural network. The neural network had thirteen inputs for each of the thirteen parameters considered and two outputs for classifying the patient into two categories namely heart attack and no heart attack. After training the neural network on the processed data we obtained an accuracy of 86%. For our neural network we used the 'Adam' optimiser. We used rectified linear unit and softmax activation functions in the hidden layers and the output layer respectively.

1) *Adam:* Adam is an optimization algorithm which can replace the classical stochastic gradient descent algorithm to update the weights in the network. Adam uses a single learning rate which does not change during the training of the neural network. But the learning rate used for different parameters is different.
2) *Rectified Linear Units:* This activation function will output the input as it is if the input is positive. Else, it will output zero. It is one of the preferred activation function in neural networks as it makes it easier to train the neural network and gives better performance. It is defined as
$g(z) = \max\{0, z\}$.

3) *Softmax:* It is a form of logistic regression that is used to normalize an input value into a vector of values which follows a probability distribution. The values sum up to 1.It is given by
$softmax\,(z)\,i = \frac{\exp(zi)}{\sum_j \exp(zj)}$

3.4 Predicting Risk Score

Only the "heart attack" output was considered and the "no heart attack" output was ignored. Due to the softmax function the output obtained was in terms of probability. This probability was then divided into five classes as shown in Table 1.

TABLE I Risk classification according to probability output

Probability	**Risk**
0 – 0.55	Very Low Risk
0.55 – 0.65	Low Risk
0.65 – 0.75	Medium Risk
0.75 – 0.85	High Risk
0.85+	Very High Risk

3.5 User Interface

We have developed an interactive user interface using Tkinter, a GUI library in Python. The data entered by the user is stored in the form of an array on clicking the 'submit' button. This array is then fed as an input to the neural network. Based on this data, the neural network makes a prediction by classifying the user into the five categories (Very low risk, Low risk, Medium risk, High risk and Very high risk). The user is indicated in which category he/she lies in through the user interface. Fig. 1 and Fig. 2 show an instance of the data entered by the user and the category in which he/she lies is displayed.

Fig. 1. User Interface

Since the above person is a non-smoker and has no other afflictions his risk of suffering a heart attack is medium.

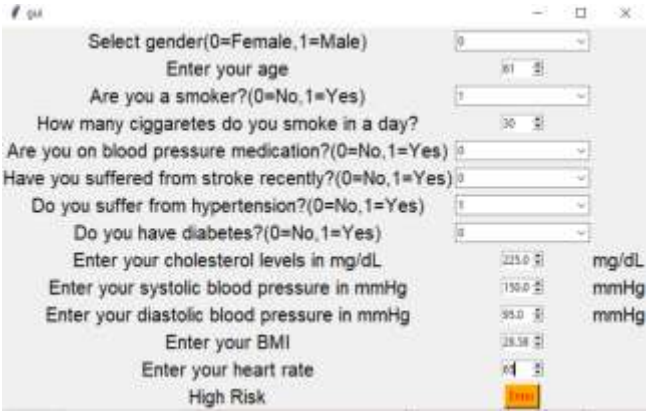

Fig. 2. User Interface

The above person is a smoker and smokes a lot of cigarettes every day. Also he suffers from hypertension and has high cholesterol level. Hence his risk of suffering a heart attack is high. Fig. 3 shows a basic flow diagram of system implementation.

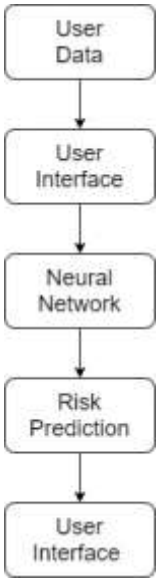

Fig.3. Flow diagram

4. RESULT

Consider the classification report shown in Fig. 3 It gives a deeper idea of the classification behaviour of the model. Precision is the ratio of true positives to the sum of true and false positives. Here, precision is 0.86 for class 0 (no heart attack) and 1.00 for class 1 (heart attack). It indicates the ability of the classifier to not label an instance positive while it is actually negative. Recall is the classifier's ability to find positive instances. It is the ratio of true positives to the sum of true positives and false negatives. Recall is 1.00 for class 0 and 0.04 for class 1. F1 score is the weighted harmonic mean of precision and recall where the best score is 1.00 and the worst score is 0.00. F1 score is 0.93 for class 0 and 0.07 for class 1. Support is the number of actual occurrences of the respective class in the dataset. In the testing dataset the occurrences of class 0 and 1 are 157 and 26 respectively.

	precision	recall	f1-score	support
0.0	0.86	1.00	0.93	157
1.0	1.00	0.04	0.07	26
micro avg	0.86	0.86	0.86	183
macro avg	0.93	0.52	0.50	183
weighted avg	0.88	0.86	0.81	183

Classification Report

5. CONCLUSION

The rate of heart attack among individuals is increasing at an alarming rate. According to statistics, 720,000 heart attacks occur in the US every year. The mortality rate of heart attack is 14%. Due to various reasons the risk of getting a heart attack is increasing. It is necessary to implement means to mitigate this risk so that necessary means can be taken beforehand. As it is said "Prevention is better than cure", it is better to know the risk one has of suffering from a heart attack before the risk escalates further so that, appropriate steps can be taken to reduce this risk. A neural network can be effectively employed to develop a system to predict the risk of heart attack among individuals. Our system provides an efficient way for people to check the risk of heart attack associated with them at the convenience of their homes. This coupled with an user interface makes the system easy to use.

6. REFERENCES

[1] https://en.wikipedia.org/wiki/Myocardial_infarction

[2] Manikandan, Sushmita. "Heart attack prediction system." In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), pp. 817-820. IEEE, 2017.

[3] Raihan, M., Saikat Mondal, Arun More, Md Omar Faruqe Sagor, Gopal Sikder, Mahbub Arab Majumder, Mohammad Abdullah Al Manjur, and Kushal Ghosh. "Smartphone based ischemic heart disease (heart attack) risk prediction using clinical data and data mining approaches, a prototype design." In *2016 19th International Conference on Computer and Information Technology (ICCIT)*, pp.299-303.IEEE,2016.

[4] Khurl, Sikander Singh, and Gurpreet Singh. "Ranking early signs of coronary heart disease among Indian patients." In *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 840-844. IEEE, 2015.

[5] Malode, CM Chethan, K. Bhargavi, B. G. Gunasheela, G. Kavana, and R. Sushmitha. "Soft set and Fuzzy Rules Enabled SVM Approach for Heart Attack Risk Classification Among Adolescents." In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pp. 1-6. IEEE, 2018.

[6] Soni, Jyoti, Ujma Ansari, Dipesh Sharma, and Sunita Soni. "Predictive data mining for medical diagnosis: An overview of heart disease prediction." *International Journal of Computer Applications* 17, no. 8 (2011): 43-48.

[7] Rajeswari, K., V. Vaithiyanathan, and T. R. Neelakantan. "Feature selection in ischemic heart disease identification using feed forward neural networks." *Procedia Engineering* 41 (2012): 1818-1823.

[8] Chitra, R., and V. Seenivasagam. "Heart attack prediction system using Fuzzy C Means classifier." *IOSR Journal of Computer Engineering* 14 (2013): 23-31.

[9] Jabbar, Ma. and D. P. Chandra. "CLUSTER BASED ASSOCIATION RULE MINING FOR HEART ATTACK PREDICTION." .

[10] Alex, Cheryl Ann et al. "Big Data Analytics in Heart Attack Prediction." *The journal of nursing care* 6 (2017): 1-9.